Module 2: Project Statement Initial Draft

Jacob Anderson

Article #1, Gut microbiome metagenomics in clinical practice: bridging the gap between research and precision medicine

This article provides context for metagenomic data, reviewing how it is now being used to identify microbial signatures associated with specific diseases, stratify patients, and inform personalized treatments. It also reviews importance of standardization, metadata quality and reproducibility, while also highlighting confounding factors (diet, medication, geography) and population bias, which we'll need to consider in our project. By reviewing real-world clinical applications, this article was and is helpful in identifying what questions or topics we would like to explore in our project, and what parts of our dataset are and would be meaningful.

Article #2, Computational Metagenomics: State of the Art

This article gives a summary of current methods used in computational metagenomics, explaining common algorithms and analytic frameworks to extract value and provide insights from complex metagenomic datasets. Some of the earlier portions of the paper aren't or won't be as important as the latter parts as the R dataset we've chosen has most of the data cleaning already done, but it's important to understand how the data is cleaned and prepared before analyzing it. It also goes over some machine learning techniques, best practices, as well as challenges in metagenomics as it is inherently a high dimensional problem set.

Questions:

1. Some microbial patterns can be common or present for a given disease, can we detect an effect level for each for that disease, or link biological function of that pattern?
2. Check for robustness, because the dataset combines multiple studies, can we test whether observed microbial differences appear across multiple studies, aiding in detecting true biological patterns from artifacts from singular studies or data collection.
3. How does host metadata such as age, geography, diet, study source influence observed microbiome differences and how can we control for them?
4. Can statistical or machine-learning techniques improve phenotype prediction from metagenomic data and find interactions among microbes / pathways rather than abundance / population proportions or presence

5. Like question 1, but just for any single given microbial pattern, can we connect microbial signatures, clinical importance and their function? i.e. Can we link findings back to microbes' roles in the body rather than just match pattern of presence of microbes in the body to certain diseases.

Our (Therayess) project will explore how human microbiome composition and functions differ across health conditions using curated metagenomic data. By comparing microbial species and functional pathways across studies (the dataset collects data from multiple studies), we aim to identify consistent, biologically meaningful patterns while accounting for technical and demographic variation.

References

Tegegne, H. A., & Savidge, T. C. (2025). G**ut microbiome metagenomics in clinical practice: bridging the gap between research and precision medicine**. *Gut microbes*, *17*(1), 2569739. https://doi.org/10.1080/19490976.2025.2569739

Pita-Galeana, M. A., Ruhle, M., López-Vázquez, L., de Anda-Jáuregui, G., & Hernández-Lemus, E. (2025). Computational Metagenomics: State of the Art. *International journal of molecular sciences*, *26*(18), 9206. https://doi.org/10.3390/ijms26189206

Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D., Beghini, F., Malik, F., Ramos, M., Dowd, J., Huttenhower, C., Morgan, M., Segata, N., & Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nature Methods, 14*(11), 1023–1024. https://doi.org/10.1038/nmeth.4468