Research Use Statement (2,200 characters max).

The main objective of the proposed research project is to provide an R software-based framework for processing, storing and analysis of the integrative Human Microbiome Project (iHMP) data. These data are publicly available via the iHMP data portal (<http://portal.hmpdacc.org/>) and contains the collection of microbiome data with a total of 16,531 samples from 4 major HMP studies: Multi-Omic Microbiome Study: Pregnancy Initiative (MOMS-PI), Type 2 diabetes mellitus (T2D), and Inflammatory Bowel Disease Multi-Omics database (IBDMBD) (<https://www.hmpdacc.org/ihmp/> ). Currently, this is the largest collection of publicly available longitudinal multi-omics studies collected to understand the microbial biomarkers of preterm pregnancy, type 2 diabetes, and inflammatory bowel disease. While basic clinical annotation information, such as pregnancy or IBD outcome, is available through the data portal, the majority of participants data important for the analysis of disease outcomes, such as age, socioeconomic status, and medical history is not available through the portal. The goal of this project is to securely integrate these data into the R software framework enabling authorized users the in-depth analysis of microbiome data. This project extends our previous work on the integration of microbiome data with the controlled access dbGAP data (<https://www.biorxiv.org/content/early/2018/08/29/299115>).

**Non-Technical Summary**

The human microbiome contributes to human well-being, disease progression, pregnancy outcomes. Several consortia provide open-access microbiome data; however, software tools for the analysis of it remain undeveloped. This project will develop an R package enabling the integrative analysis of the microbiome data securely integrated with clinical annotations available for authorized users.

**Research objectives**

**Study design**

**Analysis plan**

**including the phenotypic characteristics that will be evaluated in association with genetic variants**

**Explanation of how the proposed research is consistent with the data use limitations for the requested dataset(s)**

**Brief description of any planned collaboration with researchers at other institutions, including the name of the collaborator(s) and their institution(s)**