Project goals:

1. Extend the HMP16SData Bioconductor package to HMP2 data study

(<https://www.biorxiv.org/content/early/2018/08/29/299115>).

1. Package format: similar to HMP16SData

See:

library(HMP16SData)

browseVignettes("HMP16SData")

1. We will need to extend HMP16SData format to multi-omics modalities. Namely, each -omics modality should be easily matched by subject id and not necessarily all subjects will have each of the different data types available
2. dbGAP data access available for HMP1 project only. Will need to work without metadata for a moment. Current steps: how dbGAP data is merged with the 16S data in HMP16SData package (see vignette Ch 3.3). Repository key is in ProjectNotes/HMP1/ folder
3. Simple analysis using phyloseq
4. See phyloseq tutorials <https://joey711.github.io/phyloseq/courses.html>
5. Data structure: <https://joey711.github.io/phyloseq/import-data.html>
6. otu\_table - Works on any numeric matrix. You must also specify if the species are rows or columns. This is 16S data counts matrix.
7. sample\_data - Works on any data.frame. The rownames must match the sample names in the otu\_table if you plan to combine them as a phyloseq-object. This is for importing metadata (i.e. covariates)
8. tax\_table - Works on any character matrix. The rownames must match the OTU names (taxa\_names) of the otu\_table if you plan to combine it with a phyloseq-object. Information about microbial species (called taxa) from greengenes.
9. MDS plots for different body sites
10. multi-table approaches

Major HMP2 studies available through HMP DAC portal <http://portal.hmpdacc.org/>:

1. Multi-Omic Microbiome Study-Pregnancy Initiative (MOMS-PI)  -- VCU study, Buck lab. Data contact Jennifer Fettweis jennifer.fettweis@vcuhealth.org
2. The Inflammatory Bowel Disease (IBD) Multi'omics Data (IBDMDB) – Harvard study, Huttenhower lab. Data contact Cesar Arze: carze@hsph.harvard.edu

<https://www.dropbox.com/s/nhloprbetszkda5/322196_1_merged_1536386292.pdf?dl=0>

1. Type 2 diabetes mellitus (T2D)

Multi-omics data available for each study through the portal:

16S, whole-genome sequencing, cytokine, transcriptome, metatranscriptome, lipidomic

Metadata:

1. very limited information (sample id, visit number, sample type) available through the portal.
2. Metadata will be available through controlled access dbGAP data
3. MOMS-PI Study Accession**:** phs001523.v1.p1 is assigned but data not available yet
4. Other studies data are not yet available either