|  |  |  |
| --- | --- | --- |
| Variable | Use | Description |
| SampleID | Structural | Unique ID number given to each clinical sample |
| study\_id | Structural | Participant ID number, repeated for each time point |
| day | Structural | Day of sample collection relative to the date of HCT |
| total\_metagenomic | Structural | Total number of metagenomic (bacteria, fungi, archaea) reads generated from sample |
| R1\_metag\_paired | Structural | Total number of paired metagenomic reads generated by sequencing |
| pct\_metagenomic | Structural | Proportion of reads that were assigned to microbes |
| age | Fixed covariate | Age (in years) at the time of enrollment |
| sex | Fixed covariatae | F=female  M=male |
| race | Fixed covariate | Participant/race ethnicity  1=Black or African-American  2=Non-Hispanic White  3=Middle Eastern or Arab-American  4=East Asian or Asian-American  5=South Asian or Indian-American  6=Latino or Hispanic-American  7=Native Hawaiian or Pacific Islander  8=Native American or Alaska Native |
| diagnosis | Fixed covariate | HSCT indication (category) |
| primary\_dx | Fixed covariate | HSCT indication (detail) |
| type\_hsct | Fixed covariate | Type of HSCT (allogeneic or autologous) |
| donor\_hsct | Fixed covariate | Related or unrelated donor |
| source\_hsct | Fixed covariate | Source of allogeneic stem cells  BM=bone marrow  UCB=umbilical cord blood |
| hsct\_prep\_cat | Fixed covariate | Preparative regimen category  Myelo=Myeloablative conditioning  Non-myelo=Non-myeloablative conditioning  RIC=Reduced intensity conditioning |