CITRC\_Data\_Summary

Jordan Lo

2023-07-11

# readxl allows us to read in xlsx files  
library(readxl)  
# tidyverse gives us access to data processing tools  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

# dplyr gives us access to data processing tools  
library(dplyr)  
# ggplot2 lets us create plots and figures  
library(ggplot2)  
# skimr gives us dataframe summaries  
library(skimr)

# reads in an excel file as a dataframe  
master <- read\_excel("B:\\CITRC\\Plasma Biomarker August 2022\\master\_data\_sheets\\CITRC\_Data\_Master\_20230616.xlsx")  
# subsets the above file   
master <- master[, c(1,3:12)]  
colnames(master) <- c("Subject\_ID",  
 "Ang-1\_V1",  
 "Ang-2\_V1",  
 "Ang-1\_V2",  
 "Ang-2\_V2",  
 "sTREM1\_V1",  
 "sTREM1\_V2",  
 "TNFRI\_V1",  
 "TNFRI\_V2",  
 "AngPTL4\_V1",  
 "AngPTL4\_V2")  
  
# pivots the dataframe from wide to long and filters for just samples with concentrations  
long\_master <- pivot\_longer(master,  
 cols = c("Ang-1\_V1",  
 "Ang-1\_V2",  
 "Ang-2\_V1",  
 "Ang-2\_V2",  
 "sTREM1\_V1",  
 "sTREM1\_V2",  
 "TNFRI\_V1",  
 "TNFRI\_V2",  
 "AngPTL4\_V1",  
 "AngPTL4\_V2"),  
 names\_to = "Assay",  
 values\_to = "Concentration") %>%  
 drop\_na() %>%  
# filter(Concentration > 0) %>%  
 group\_by(Assay) %>%  
 add\_count()  
  
# pivots the dataframe from wide to long and filters for just the values without concentrations  
long\_master\_empty <- pivot\_longer(master,  
 cols = c("Ang-1\_V1",  
 "Ang-1\_V2",  
 "Ang-2\_V1",  
 "Ang-2\_V2",  
 "sTREM1\_V1",  
 "sTREM1\_V2",  
 "TNFRI\_V1",  
 "TNFRI\_V2",  
 "AngPTL4\_V1",  
 "AngPTL4\_V2"),  
 names\_to = "Assay",  
 values\_to = "Concentration") %>%  
 drop\_na() %>%  
 filter(Concentration < 0) %>%  
 group\_by(Assay) %>%  
 add\_count()  
  
master <- pivot\_wider(long\_master,  
 id\_cols = Subject\_ID,  
 names\_from = Assay,  
 values\_from = Concentration)

# subsets the long\_master file  
test <- long\_master[, c(2,4)]  
  
# takes all distinct values from the test dataframe  
table\_1 <- test %>%  
 distinct(Assay,  
 .keep\_all = TRUE)  
  
# subsets all the data from long\_master that is tagged as being below the LLOD  
llod <- long\_master %>%  
 filter(Concentration == -89) %>%  
 # groups by assay and adds a counter  
 group\_by(Assay) %>%  
 add\_count() %>%  
 distinct(Assay,  
 .keep\_all = TRUE)

## Storing counts in `nn`, as `n` already present in input  
## ℹ Use `name = "new\_name"` to pick a new name.

# further subsets the dataframe into the needed columns  
llod <- llod[c(2,5)]  
  
# merges the two tables and renames the column  
table\_1 <- merge(table\_1,  
 llod,  
 by = 'Assay',  
 all.x = TRUE,  
 all.y = TRUE)  
colnames(table\_1)[3] <- "n\_below\_LLOD"  
  
# does the same as above but for ULOD  
ulod <- long\_master %>%  
 filter(Concentration == -99) %>%  
 group\_by(Assay) %>%  
 add\_count() %>%  
 distinct(Assay,  
 .keep\_all = TRUE)

## Storing counts in `nn`, as `n` already present in input  
## ℹ Use `name = "new\_name"` to pick a new name.

ulod <- ulod[c(2,5)]  
  
table\_1 <- merge(table\_1,  
 ulod,  
 by = 'Assay',  
 all.x = TRUE,  
 all.y = TRUE)  
colnames(table\_1)[4] <- "n\_above\_ULOD"  
  
# replaces all NA with 0  
table\_1[is.na(table\_1)] <- 0  
  
# adds a column of all values within range  
within\_range <- long\_master[, c(1:3)] %>%  
 filter(Concentration > 0) %>%  
 group\_by(Assay) %>%  
 add\_count %>%  
 distinct(Assay,  
 .keep\_all = TRUE)  
within\_range <- within\_range[, c(2,4)]  
  
table\_1 <- merge(table\_1,  
 within\_range,  
 by = "Assay",  
 all.x = TRUE,  
 all.y = TRUE)  
colnames(table\_1)[5] <- "n\_within\_range"  
colnames(table\_1)[2] <- "n\_total"  
  
# separates assay and visit  
table\_1 <- table\_1 %>%  
 separate(Assay,  
 into = c("Assay",  
 "Visit"),  
 sep = "\_")  
  
write\_csv(table\_1,   
 file = "table\_1.csv")  
  
table\_1

## Assay Visit n\_total n\_below\_LLOD n\_above\_ULOD n\_within\_range  
## 1 Ang-1 V1 401 22 0 379  
## 2 Ang-1 V2 371 26 0 345  
## 3 Ang-2 V1 401 0 1 400  
## 4 Ang-2 V2 371 0 0 371  
## 5 AngPTL4 V1 404 0 0 404  
## 6 AngPTL4 V2 371 0 0 371  
## 7 sTREM1 V1 406 0 0 406  
## 8 sTREM1 V2 371 0 0 371  
## 9 TNFRI V1 406 0 5 401  
## 10 TNFRI V2 371 0 2 369