EMIT\_Enroll\_Condition\_for\_Negative\_Samples.R

jbueno

Thu Feb 7 19:50:32 2019

# Title: enroll condition for negative samples  
# Author: Jacob Bueno de Mesquita, with material from Jing Yan and Donald Milton  
  
# Summary: I'm moving this script to the git lab repository and also doing some minor cleaning to enabling the clean reproduction of this script.  
# Signed: Jacob Bueno de Mesquita  
# Date: January 8; February, 2019  
  
# Description: This script reads in the Clinical Database (EMIT\_samples.cc.RDS) and the subtyping file (EMIT\_subtypes.RDS) to create a summary file that contains a list of the sampling instances with culture data (passage or quantitative focus assay) that shows that there is positive virus, despite a negative first visit NP swab (first visit NP swab is the basis of the EMIT\_subtypes.RDS file).   
# This output is: "/Users/jbueno/Box Sync/EMIT/EMIT\_Data\_Analysis\_Jake/EMIT\_UMD\_Natural\_Infection/Curated Data/Cleaned Data/negative subtype sample with positive culture or focus.RDS"  
# This output can be used in subsequent analyses or exploration into the dataset as it pertains to counting individuals with evidence of viral infection by both culture and qRT-PCR methods.   
  
#########################  
# Name: enroll condition for negative samples  
# Authors: Jing Yan & Donald Milton  
# Date: February 3, 2016  
# # Purpose:   
# merge "EMIT\_subtypes" (df with list of type and subtype from NP swabs) with   
# "samples.cc" (df with a list of all samples, culture & some clinical data),   
# Then output the list of subjects with negative sample type and alos not enrolled,   
# and negative sample type with positive culture results  
#########################  
  
library(dplyr)  
library(tidyr)  
  
sessionInfo()

## R version 3.5.1 (2018-07-02)  
## Platform: x86\_64-apple-darwin15.6.0 (64-bit)  
## Running under: macOS 10.14.2  
##   
## Matrix products: default  
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib  
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib  
##   
## locale:  
## [1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] htmlTable\_1.12 rmarkdown\_1.10 markdown\_0.8   
## [4] bindrcpp\_0.2.2 lme4\_1.1-19 Matrix\_1.2-14   
## [7] arsenal\_1.5.0 lubridate\_1.7.4 data.table\_1.11.8  
## [10] knitr\_1.20 readxl\_1.1.0 RcppRoll\_0.3.0   
## [13] forcats\_0.3.0 stringr\_1.3.1 dplyr\_0.7.7   
## [16] purrr\_0.2.5 readr\_1.1.1 tidyr\_0.8.2   
## [19] tibble\_1.4.2 ggplot2\_3.1.0 tidyverse\_1.2.1   
##   
## loaded via a namespace (and not attached):  
## [1] tidyselect\_0.2.5 splines\_3.5.1 haven\_1.1.2 lattice\_0.20-38   
## [5] colorspace\_1.3-2 testthat\_2.0.1 htmltools\_0.3.6 yaml\_2.2.0   
## [9] utf8\_1.1.4 rlang\_0.3.0.1 pillar\_1.3.0 nloptr\_1.2.1   
## [13] glue\_1.3.0 withr\_2.1.2 modelr\_0.1.2 bindr\_0.1.1   
## [17] plyr\_1.8.4 munsell\_0.5.0 gtable\_0.2.0 cellranger\_1.1.0  
## [21] rvest\_0.3.2 htmlwidgets\_1.3 evaluate\_0.12 fansi\_0.4.0   
## [25] highr\_0.7 broom\_0.5.0 Rcpp\_0.12.19 checkmate\_1.8.5   
## [29] scales\_1.0.0 backports\_1.1.2 jsonlite\_1.5 digest\_0.6.18   
## [33] hms\_0.4.2 stringi\_1.2.4 rprojroot\_1.3-2 grid\_3.5.1   
## [37] cli\_1.0.1 tools\_3.5.1 magrittr\_1.5 lazyeval\_0.2.1   
## [41] crayon\_1.3.4 pkgconfig\_2.0.2 MASS\_7.3-51 xml2\_1.2.0   
## [45] assertthat\_0.2.0 minqa\_1.2.4 httr\_1.3.1 rstudioapi\_0.8   
## [49] R6\_2.3.0 nlme\_3.1-137 compiler\_3.5.1

#\_\_\_\_\_\_\_\_\_\_\_\_\_  
# Setup all I/O  
# setwd('C:/Users/Jing/Box Sync/EMIT/EMIT\_Data\_Analysis')  
# setwd('/Users/dmilton/Box Sync/0\_DKM/Lab/Biodefense/EMIT/EMIT\_Data\_Analysis')  
# setwd('/Volumes/Internal RAID Set 1/Box Sync/0\_DKM/Lab/Biodefense/EMIT/EMIT\_Data\_Analysis')  
# setwd("/Users/jbueno/Box Sync/EMIT/EMIT\_Data\_Analysis\_Jake")  
  
# Comment out the working directory to facilitate markdown report compilation  
  
# Out.dir <- "R\_output/"  
# sink(file = "R\_output/enroll condition for negative samples.txt", split = TRUE)  
#\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
samples <- readRDS("/Users/jbueno/Box Sync/EMIT/EMIT\_Data\_Analysis\_Jake/EMIT\_UMD\_Natural\_Infection/Curated Data/Cleaned Data/EMIT\_samples.cc.RDS")  
flu.types <- readRDS("/Users/jbueno/Box Sync/EMIT/EMIT\_Data\_Analysis\_Jake/EMIT\_UMD\_Natural\_Infection/Curated Data/Cleaned Data/EMIT\_subtypes.RDS")  
flu.types <- select(flu.types, subject.id, type.inf)  
  
samples.types <- left\_join(samples, flu.types, by = "subject.id")  
  
# cat("\n Rows samples:", nrow(samples), " Rows flu.types:", nrow(flu.types), " Rows samples.types:", nrow(samples.types))  
  
subjects.types <- samples.types %>%  
 distinct(subject.id)  
  
# print(with(subjects.types,(ftable(addmargins(table(enrolled,type.inf,exclude = c()))))))  
  
neg.samples <- samples.types %>%  
 filter(type.inf == 'Negative' | type.inf == 'bad assay')  
  
# cat("\n With a negative sample type (or bad assay), the number of subjects have a screen visit 2 is ",  
print(nrow(filter(neg.samples, visit.num == 2)))

## [1] 0

# With a negative sample type (or bad assay), the number of subjects have a screen visit 2 is...  
nrow(filter(neg.samples, visit.num == 2))

## [1] 0

neg.unenroll <- neg.samples %>%   
 filter(g2.run == 0) %>%  
 distinct(date.visit, subject.id, .keep\_all = TRUE) %>%  
 select(subject.id, date.visit, sample.id, sample.type, g2.run,   
 visit.num, passpos, validp, focus.ct, enrolled)  
  
# cat("\n subjects with negative sample type and are also not enrolled. \n")  
print(neg.unenroll)

## subject.id date.visit sample.id sample.type g2.run visit.num  
## 1 10 2012-12-05 10\_1 Nasopharyngeal swab 0 1  
## 2 236 2013-02-12 236\_1 Nasopharyngeal swab 0 1  
## 3 239 2013-02-12 239\_1 Nasopharyngeal swab 0 1  
## 4 39 2012-12-21 39\_1 Nasopharyngeal swab 0 1  
## 5 9 2012-12-05 9\_1 Nasopharyngeal swab 0 1  
## passpos validp focus.ct enrolled  
## 1 NA NA NA FALSE  
## 2 NA NA NA FALSE  
## 3 NA NA NA FALSE  
## 4 NA NA NA FALSE  
## 5 NA NA NA FALSE

enroll.np <- neg.samples %>%   
 filter(g2.run != 0)  
neg.pcr.pos.passage <- neg.samples %>%   
 filter(passpos == TRUE & validp == TRUE)  
neg.pcr.pos.focus <- neg.samples %>%   
 filter(focus.ct > 0)  
  
pos.culture <- neg.pcr.pos.passage %>%  
 full\_join(neg.pcr.pos.focus, by = c("subject.id", "type.inf", "date.visit", "sample.id", "sample.type", "visit.num", "passpos", "g2.run", "validp", "focus.ct", "enrolled"))  
  
pos.culture <- pos.culture %>%  
 arrange(subject.id) %>%   
 select(subject.id, date.visit, sample.id, sample.type, g2.run, visit.num, passpos, validp, focus.ct, enrolled)  
  
# cat("\n subjects with negative sample type but positive culture results. \n")  
print(pos.culture)

## subject.id date.visit sample.id sample.type g2.run visit.num  
## 1 35 2012-12-19 35\_1 Nasopharyngeal swab 1 1  
## 2 35 2012-12-19 35\_5 Throat Swab 1 1  
## 3 50 2013-01-07 50\_1 Nasopharyngeal swab 1 1  
## 4 50 2013-01-07 50\_3 GII condensate NO mask 1 1  
## 5 59 2013-01-10 59\_5 Throat Swab 1 1  
## 6 105 2013-01-28 105\_5 Throat Swab 1 1  
## 7 105 2013-01-28 105\_1 Nasopharyngeal swab 1 1  
## 8 105 2013-01-28 105\_3 GII condensate NO mask 1 1  
## 9 210 2013-02-09 210\_1 Nasopharyngeal swab 1 1  
## 10 210 2013-02-09 210\_5 Throat Swab 1 1  
## 11 223 2013-02-11 223\_1 Nasopharyngeal swab 1 1  
## 12 223 2013-02-11 223\_3 GII condensate NO mask 1 1  
## 13 223 2013-02-11 223\_5 Throat Swab 1 1  
## 14 226 2013-02-11 226\_1 Nasopharyngeal swab 1 1  
## 15 231 2013-02-11 231\_5 Throat Swab 1 1  
## 16 231 2013-02-11 231\_1 Nasopharyngeal swab 1 1  
## 17 234 2013-02-12 234\_5 Throat Swab 1 1  
## 18 234 2013-02-12 234\_1 Nasopharyngeal swab 1 1  
## 19 234 2013-02-12 234\_3 GII condensate NO mask 1 1  
## 20 290 2013-02-19 290\_7 Nasopharyngeal swab 2 999  
## 21 306 2013-02-20 306\_3 GII condensate NO mask 1 1  
## 22 306 2013-02-20 306\_5 Throat Swab 1 1  
## 23 327 2013-02-26 327\_3 GII condensate NO mask 1 1  
## 24 327 2013-02-26 327\_5 Throat Swab 1 1  
## 25 356 2013-03-07 356\_3 GII condensate NO mask 1 1  
## passpos validp focus.ct enrolled  
## 1 TRUE TRUE 0.000000 TRUE  
## 2 TRUE TRUE 0.000000 TRUE  
## 3 FALSE TRUE 136.666667 TRUE  
## 4 NA FALSE 396.666667 TRUE  
## 5 TRUE TRUE 0.000000 TRUE  
## 6 TRUE TRUE 40.000000 TRUE  
## 7 FALSE TRUE 60.000000 TRUE  
## 8 NA FALSE 6.666667 TRUE  
## 9 TRUE TRUE 0.000000 TRUE  
## 10 TRUE TRUE 0.000000 TRUE  
## 11 TRUE TRUE 0.000000 TRUE  
## 12 TRUE TRUE 0.000000 TRUE  
## 13 TRUE TRUE 66.666667 TRUE  
## 14 FALSE TRUE 13.333333 TRUE  
## 15 TRUE TRUE 0.000000 TRUE  
## 16 FALSE TRUE 66.666667 TRUE  
## 17 TRUE TRUE 10.000000 TRUE  
## 18 FALSE TRUE 20.000000 TRUE  
## 19 FALSE TRUE 3.333333 TRUE  
## 20 TRUE TRUE 0.000000 TRUE  
## 21 FALSE TRUE 246.666667 TRUE  
## 22 FALSE TRUE 13.333333 TRUE  
## 23 TRUE TRUE 20.000000 TRUE  
## 24 TRUE TRUE 86.666667 TRUE  
## 25 TRUE TRUE 0.000000 TRUE

saveRDS(pos.culture, "/Users/jbueno/Box Sync/EMIT/EMIT\_Data\_Analysis\_Jake/EMIT\_UMD\_Natural\_Infection/Curated Data/Cleaned Data/negative subtype sample with positive culture or focus.RDS")  
  
# sink()  
# closeAllConnections()