LINCs Population-Level Data and Metadata Merge

2015-07-06

Summary

This script merges population data and metadata for the LINCs QA MEMA data.

In the code, the variable ss determines which staining set (SS1, SS2 or SS3) to merge. All .txt data files in the ./Population/Raw Data folder will be merged with the well (xlsx) and log (XML) data from the ./Metadata folder.

This merging assumes the labels in the data from the A rows and B rows are swapped. That is, the input file names actual well A01 as B01, A02 as B02, etc. The data is renamed to match the actual, physical layout of the 8 well plate.

This merging also assumes that the actual, physical B row wells (B01-B04) have been printed upside-down. That is, rotated 180 degrees resulting in the spot 1,1 being in the lower right corner instead of the upper left corner. The metadata is matched to the actual printed orientation.

```r
library("limma")
library("MEMA")
library("data.table")

#Select a staining set
ss <- "SS1"
```

```r
#Read in the spot metadata from the gal file
smd <- readSpotMetadata(paste0("./",ss,"/Metadata/20150515_LI8X001_v1.gal"))
#Make a short name from the content names for labeling in plots
smd$ShortName <- gsub("_.*","",smd$Name)
smd$ShortName <- gsub("-","blank",smd$ShortName)

#Add the print order and deposition number to the metadata
  ldf <- readLogData(paste0("./",ss,"/Metadata/20150512-112336.xml"))
  spotMetadata <- merge(smd,ldf, all=TRUE)
  setkey(spotMetadata,Spot)
#Make a rotated version of the spot metadata
spotMetadata180 <- rotateMetadata(spotMetadata)
ARowMetadata <- data.table(spotMetadata,Well=rep(c("A01", "A02","A03","A04"),each=nrow(spotMetadata)))
BRowMetadata <- data.table(spotMetadata180,Well=rep(c("B01", "B02","B03","B04"),each=nrow(spotMetadata18
```

```r
#The next steps are to bring in the well metadata, the print order and the Tecan intensity
#data. The Tecan data includes the raw, background and net data values.
popDataFiles <- dir(paste0(ss,"/Population/Raw Data"),full.names = TRUE)

annDTList <- lapply(popDataFiles, function(popDataFile){

  spotsPerWell <- max(spotMetadata$Block) * max(spotMetadata$Row) * max(spotMetadata$Column)
  #Read data from one plate
  popData<-fread(popDataFile,stringsAsFactors=FALSE)
  #Change the name of the first column
  setnames(popData,1,"Index")
```

```r
  #Delete the statistical summary after the population data
  popData <- popData[1:spotsPerWell,]
  #Organize the data by well
  popData <- melt8Well(popData)
  setnames(popData,"Grid","Block")
  setkey(popData, Well, Block, Row, Column)

  #Merge in the well metadata
 #Read the well metadata from a multi-sheet Excel file
  wellMetadata <- data.table(readMetadata(paste0(ss,"/Metadata/",unique(popData$Barcode),".xlsx")), key=
  popData <- merge(popData, wellMetadata, by="Well")
  # Eliminate Variations in the Endpoint metadata
    endpointNames <- grep("End",colnames(popData), value=TRUE)
    endpointWL <- regmatches(endpointNames,regexpr("[[:digit:]]{3}",endpointNames))
    setnames(popData,endpointNames,paste0("Endpoint",endpointWL))
  return(popData)
})
annDT <- rbindlist(annDTList, use.names=TRUE, fill=TRUE)
#Merge the data with spot metadata that is well row specific
sDT <- merge(annDT,rbind(ARowMetadata,BRowMetadata), by = c("Well","Block","Row","Column"), all=TRUE)


#Add the median normalized loess values of each channel on a per well basis
sDT <- sDT[,Loess488 := loessModel(.SD,value="Net.488",span=.1), by="Barcode,Well"]
sDT <- sDT[,Loess555 := loessModel(.SD,value="Net.532",span=.1), by="Barcode,Well"]
sDT <- sDT[,Loess647 := loessModel(.SD,value="Net.635",span=.1), by="Barcode,Well"]


#Median normalize to the plate's control well for each channel's Net value
sDT <- sDT[,Net.488.MedNorm := normWellsWithinPlate(.SD, value="Net.488", baseECM = ".*",baseGF = "High
sDT <- sDT[,Net.532.MedNorm := normWellsWithinPlate(.SD, value="Net.532", baseECM = ".*",baseGF = "High
sDT <- sDT[,Net.635.MedNorm := normWellsWithinPlate(.SD, value="Net.635", baseECM = ".*",baseGF = "High


#Add randomized version of the normalized CellMask signal
sDT <- sDT[,Net.532.MedNorm.Random := randomizePositions(Net.532.MedNorm), by="Barcode,Well"]


write.table(sDT, paste0(ss,"/Population/Annotated Data/",unique(sDT$CellLine),"_",ss,"_","PopAnn.txt"),
```