2_dataPreprocess.R

Brandon

Tue Apr 24 13:46:05 2018

```
dataPreprocess <- function(){
  library(dplyr)
  library(splitstackshape)

  # variables from Python
  QuantType <- summarySettings["QuantType",]

  # Remove Contaminants, Reverse hits ####
  proteinGroups <- read.delim("~/example_txt/proteinGroups.txt", sep = "\t")
  PGs <- proteinGroups
  PGs <- PGs[PGs$Potential.contaminant != "+" & PGs$Reverse != "+",]

  # Make Quantitation file (choose LFQ/Intensities from python GUI**) ####
  uniprot.IDs <- data.frame(Protein.ID = PGs$Majority.protein.IDs, stringsAsFactors = F)
  uniprot.IDs$Protein.ID <- as.character(uniprot.IDs$Protein.ID)
  uniprot.IDs <- cSplit(indt = uniprot.IDs, splitCols = "Protein.ID", sep = ";")
  uniprot.IDs <- data.frame(uniprot.IDs[,1])
  colnames(uniprot.IDs) <- "Protein.ID"

  if (QuantType == "Intensity"){
    intensity.PG <- cbind(uniprot.IDs, PGs[grep(colnames(PGs), pattern = "Intensity.")])
    intensity.PG$Protein.ID <- as.character(intensity.PG$Protein.ID)
    # Make quant file with common name despite LFQ/Intensity (makes downstream tasks easier)
    ProteinQuant <- intensity.PG
  }

  if (QuantType == "LFQ"){
    LFQ.PG <- cbind(uniprot.IDs, PGs[grep(colnames(PGs), pattern = "LFQ.")])
    LFQ.PG$Protein.ID <- as.character(LFQ.PG$Protein.ID)
    # Make quant file with common name despite LFQ/Intensity (makes downstream tasks easier)
    ProteinQuant <- LFQ.PG
  }

  # Normalise data (usually for Intensity data, Choose in python GUI**) ####
  ProteinQuant.norm <- .


}
```