P202A Analysis

Sophie Shaw

03/12/2018

DEP Package - https://bioconductor.org/packages/devel/bioc/vignettes/DEP/inst/doc/DEP.html

Install - only once. Also needs dplyr from tidyverse Commented out for HTML

```
#install.packages("BiocManager")
#BiocManager::install("DEP")
#install.packages("tidyverse")
```

Library

```
library("DEP")
```

```
## Warning in fun(libname, pkgname): mzR has been built against a different Rcpp version (0.12.19)
## than is installed on your system (1.0.0). This might lead to errors
## when loading mzR. If you encounter such issues, please send a report,
## including the output of sessionInfo() to the Bioc support forum at
## https://support.bioconductor.org/. For details see also
## https://github.com/sneumann/mzR/wiki/mzR-Rcpp-compiler-linker-issue.
```

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Launch Shiny App - off for markdown

```
#run_app("LFQ")
```

Manual Analysis

Loading the data = 4311 proteins

```
data<-read.table("../AP17112_PBMC/Maxwell/Maxwell_proteinGroups_v2.txt", header=T, sep="\t")
```

Remove the Reverse hits - 4262 proteins remain

```
data <- filter(data, Reverse != "+")
```

Remove the contaminants - As contaminant marking wasn't run, not needed in this case

```
#data <- filter(data, Potential.contaminant != "+")
```

The next step resolves duplicated gene names First check if there is any - TRUE

```
data$Gene.names %>% duplicated() %>% any()
```

```
## [1] TRUE
```

Then make a table of the duplicated names - 56 in total!

```
data %>% group_by(Gene.names) %>% summarize(frequency = n()) %>%
  arrange(desc(frequency)) %>% filter(frequency > 1)
```

```
## # A tibble: 56 x 2
##    Gene.names frequency
##    <fct>          <int>
##  1 ""                15
##  2 HLA-B             14
##  3 HLA-A              9
##  4 HLA-C              5
##  5 HLA-DRB1           4
##  6 ITGB2              3
##  7 AAK1               2
##  8 ABLIM1             2
##  9 ACP1               2
## 10 ACTB               2
## # ... with 46 more rows
```

Edit the name of these genes to be a combination of gene name and protein name:

```
data_unique <- make_unique(data, "Gene.names", "Protein.IDs", delim = ";")
```

And then check that it's worked. Now false so it's worked!

```
data_unique$name %>% duplicated() %>% any()
```

```
## [1] FALSE
```

Need to upload the experimental data spreadsheet for creating the SummarizedExperiment object Also need to isolate just the LFQ columns from the data sheet

```
experimental_design<-read.table("../AP17112_PBMC/experimental_design.txt", header=T, sep="\t")
LFQ_columns<-grep("LFQ.",colnames(data_unique))
```

Make the SE

```
data_se<- make_se(data_unique, LFQ_columns, experimental_design)
```

```
## Error: words is not a character vector
```

Session Info

```
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: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
```

```
## other attached packages:
## [1] bindrcpp_0.2.2 dplyr_0.7.8    DEP_1.4.0
##
## loaded via a namespace (and not attached):
##   [1] ProtGenerics_1.14.0         bitops_1.0-6
##   [3] matrixStats_0.54.0          doParallel_1.0.14
##   [5] RColorBrewer_1.1-2          rprojroot_1.3-2
##   [7] GenomeInfoDb_1.18.1         MSnbase_2.8.2
##   [9] tools_3.5.1                 backports_1.1.2
##  [11] utf8_1.1.4                  R6_2.3.0
##  [13] DT_0.5                      affyio_1.52.0
##  [15] tmvtnorm_1.4-10             lazyeval_0.2.1
##  [17] BiocGenerics_0.28.0         colorspace_1.3-2
##  [19] GetoptLong_0.1.7            tidyselect_0.2.5
##  [21] compiler_3.5.1              preprocessCore_1.44.0
##  [23] cli_1.0.1                   Biobase_2.42.0
##  [25] DelayedArray_0.8.0          sandwich_2.5-0
##  [27] scales_1.0.0                mvtnorm_1.0-8
##  [29] readr_1.2.1                 affy_1.60.0
##  [31] stringr_1.3.1               digest_0.6.18
##  [33] rmarkdown_1.10              XVector_0.22.0
##  [35] pkgconfig_2.0.2             htmltools_0.3.6
##  [37] limma_3.38.3                htmlwidgets_1.3
##  [39] rlang_0.3.0.1               GlobalOptions_0.1.0
##  [41] rstudioapi_0.8              impute_1.56.0
##  [43] shiny_1.2.0                 shape_1.4.4
##  [45] bindr_0.1.1                 zoo_1.8-4
##  [47] mzID_1.20.0                 BiocParallel_1.16.2
##  [49] RCurl_1.95-4.11             magrittr_1.5
##  [51] GenomeInfoDbData_1.2.0      MALDIquant_1.18
##  [53] Matrix_1.2-15               fansi_0.4.0
##  [55] Rcpp_1.0.0                  munsell_0.5.0
##  [57] S4Vectors_0.20.1            imputeLCMD_2.0
##  [59] vsn_3.50.0                  stringi_1.2.4
##  [61] yaml_2.2.0                  MASS_7.3-51.1
##  [63] SummarizedExperiment_1.12.0 zlibbioc_1.28.0
##  [65] plyr_1.8.4                  grid_3.5.1
##  [67] parallel_3.5.1              promises_1.0.1
##  [69] shinydashboard_0.7.1        crayon_1.3.4
##  [71] lattice_0.20-38             hms_0.4.2
##  [73] circlize_0.4.5              mzR_2.16.0
##  [75] knitr_1.20                  ComplexHeatmap_1.20.0
##  [77] pillar_1.3.0                GenomicRanges_1.34.0
##  [79] rjson_0.2.20                codetools_0.2-15
##  [81] stats4_3.5.1                XML_3.98-1.16
##  [83] glue_1.3.0                  evaluate_0.12
##  [85] pcaMethods_1.74.0           BiocManager_1.30.4
##  [87] httpuv_1.4.5                foreach_1.4.4
##  [89] tidyr_0.8.2                 gtable_0.2.0
##  [91] purrr_0.2.5                 norm_1.0-9.5
##  [93] assertthat_0.2.0            ggplot2_3.1.0
##  [95] mime_0.6                    xtable_1.8-3
##  [97] later_0.7.5                 ncdf4_1.16
##  [99] tibble_1.4.2                iterators_1.0.10
```

```
## [101] gmm_1.6-2                  IRanges_2.16.0
## [103] cluster_2.0.7-1
```