qPLEXdata

(Analysis of qPLEX-RIME data and Total proteome data)

Kamal Kishore

May 20, 2021

Contents

T	Description	1
2	$\mathbf{U}\mathbf{sage}$	1
3	Experiment 1: ER interactome in MCF7 cells	2
4	Experiment 2: Crosslinking comparison	2
5	Experiment 3: ER complex upon OHT treatment	3
6	Experiment 4: Total proteome upon OHT treatment	7
7	Experiment 5: ER interactome in PDX tumours	8
8	Experiment 6: ER interactome in human breast cancer tumours	9
9	Experiment 7: NCOA3 interactome in MCF7 cells	10
10	Experiment 8: CBP interactome in MCF7 cells	10
11	Experiment 9: POLR2A (RNA polymerase II) interactome in MCF7 cells	11
12	Session Information	12

1 Description

The report outlines the steps for the statistical analysis of the datasets generated from the application of the qPLEX-RIME approach in breast cancer cells and clinical tumour material (PDX and human breast cancer tumours). In addition, it also contains statistical analysis of total proteome data in breast cancer cells.

2 Usage

```
library(dplyr)
suppressWarnings(library(qPLEXanalyzer))
suppressWarnings(library(qPLEXdata))
data(human_anno)
```

3 Experiment 1: ER interactome in MCF7 cells

In this experiment we have used the qPLEX-RIME approach to identify ER-associated proteins. We performed replicate ER RIME pull-downs in five independent biological replicates and an equal number of matched IgG mock samples was included. As IgG samples represent the low background intensity, their intensity distribution profile is different from ER samples. Hence, normalizing the two together would have resulted in over-correction of the IgG intensity resulting in inaccurate computation of differences between the two groups. Therefore, the peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities. Thereafter, differential protein expression was performed using limma based analysis.

4 Experiment 2: Crosslinking comparison

An ER qPLEX-RIME experiment was performed to compare two different ways of cell crosslinking. MCF7 cells were crosslinked with DSG/formaldehyde (double) or with formaldehyde alone (single). Four biological replicates were obtained for each condition along with two IgG samples pooled from replicates of each group. One of the outlier sample was removed from the analysis. The peptide intensities were normalized using median scaling within the group, treating ER pull downs as one group and IgG pull downs as another. The normalized peptide intensities were aggregated (by summing) to protein intensities. Thereafter, differential protein expression was performed using limma based analysis.

5 Experiment 3: ER complex upon OHT treatment

Three ER qPLEX-RIME experiments were performed to investigate the dynamics of the ER complex assembly upon 4-hydroxytamoxifen (OHT) treatment at 2h, 6h and 24h or at 24h post-treatment with the vehicle alone (ethanol). Two biological replicates of each condition were included in each experiment to finally consider a total of six replicates per time point. Additionally, MCF7 cells were treated with OHT or ethanol and cross-linked at 24h post-treatment in each experiment to be used for mock IgG pull-downs and to enable discrimination of non-specific binding.

The peptide intensities in each 10plex experiment were aggregated (by summing) to protein intensities. Thereafter, protein intensity in each sample was divided by its average intensity across all the samples and log2 transformed. The scaled protein intensities from each 10plex experiment were then combined and only proteins identified in all the experiments were kept for further analysis. To filter non-specific proteins, a *limma* based differential analysis was performed comparing ER and IgG pull-downs. This step filtered out non-specific binding. Further, a subset of dataset was created excluding IgG pull-down samples.

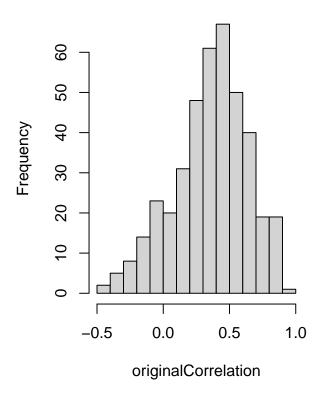
A linear regression method was applied on proteins (of this dataset) that were identified across all three replicate experiments to correct for the ER dependency. Finally, *limma* based statistical was applied on the regressed intensities for the identification of differentially bound proteins.

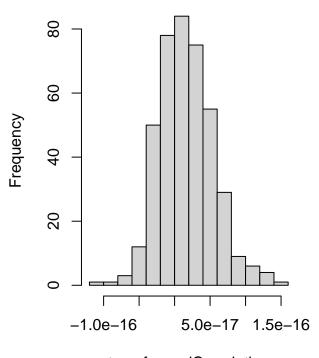
```
indExpData=c(7:16), Sequences=2, Accessions=6)
pData(MSnset_data3)$Run <- 3</pre>
## Summation of peptide to protein intensity
MSnset_P1 <- summarizeIntensities(MSnset_data1, sum, human_anno)
MSnset_P2 <- summarizeIntensities(MSnset_data2, sum, human_anno)
MSnset_P3 <- summarizeIntensities(MSnset_data3, sum, human_anno)
## Normalization
MSnset_P1 <- rowScaling(MSnset_P1,mean)</pre>
MSnset_P2 <- rowScaling(MSnset_P2,mean)</pre>
MSnset_P3 <- rowScaling(MSnset_P3,mean)</pre>
###### Compute common unique peptides
features1 <- fData(MSnset_data1)</pre>
features1 <- as.data.frame(features1[, c("Sequences",</pre>
                                           "Accessions")],
                              stringsAsFactors = FALSE)
features2 <- fData(MSnset_data2)</pre>
features2 <- as.data.frame(features2[, c("Sequences",</pre>
                                             "Accessions")],
                              stringsAsFactors = FALSE)
features3 <- fData(MSnset_data3)</pre>
features3 <- as.data.frame(features3[, c("Sequences",</pre>
                                             "Accessions")],
                              stringsAsFactors = FALSE)
features <- rbind(features1,features2,features3)</pre>
features <- unique(features)</pre>
features$Sequences <- as.character(features$Sequences)</pre>
features$Accessions <- as.character(features$Accessions)</pre>
counts <- features %>% count(Accessions)
colnames(counts)[2] <- "Count"</pre>
##### create combine MSnSet object
MSnset_P1 <- updateFvarLabels(MSnset_P1)</pre>
MSnset_P2 <- updateFvarLabels(MSnset_P2)</pre>
MSnset_P3 <- updateFvarLabels(MSnset_P3)</pre>
MSnset_P1 <- updateSampleNames(MSnset_P1)</pre>
MSnset_P2 <- updateSampleNames(MSnset_P2)</pre>
MSnset_P3 <- updateSampleNames(MSnset_P3)</pre>
suppressWarnings(MSnset_comb <- combine(MSnset_P1, MSnset_P2, MSnset_P3))</pre>
tokeep <- which(complete.cases(fData(MSnset_comb))==TRUE)</pre>
MSnset_comb <- MSnset_comb[tokeep,]</pre>
sampleNames(MSnset_comb) <- pData(MSnset_comb)$SampleName</pre>
pData(MSnset\_comb)$BioRep <- c(rep(1,4),rep(2,4),c(1,2),rep(3,4),rep(4,4),c(3,4),
```

```
rep(5,4), rep(6,4), c(5,6))
fData(MSnset_comb) <- fData(MSnset_comb)[,c(1:4)]</pre>
colnames(fData(MSnset_comb)) <- c("Accessions", "Gene", "Description",</pre>
                                                "GeneSymbol")
ind <- match(fData(MSnset_comb)$Accessions, counts$Accessions)</pre>
fData(MSnset_comb)$Count <- counts$Count[ind]</pre>
### create separate MSnSet for IgG comparision
pheno <- pData(MSnset_comb)</pre>
pheno$SampleGroup <- c(rep(c(rep("Exp",8),rep("IgG",2)),3))</pre>
pheno$SampleGroup <- factor(pheno$SampleGroup)</pre>
MSnset_IgG <- MSnset_comb</pre>
pData(MSnset_IgG) <- pheno</pre>
### Differential analysis to find ER specific interactors
contrasts <- c(
  Exp_vs_IgG = "Exp - IgG"
diffstats <- computeDiffStats(MSnSetObj=MSnset_IgG, contrasts=contrasts,</pre>
                                transform = FALSE)
results <- getContrastResults(diffstats=diffstats, contrast=contrasts,
                                transform = FALSE)
### create subset of protein filtering non-specific IgG
ind <- which(results$adj.P.Val < 0.01 & results$log2FC > 1)
diff_IgG <- results[ind,]</pre>
ind <- match(diff_IgG$Accessions, fData(MSnset_comb)$Accessions)</pre>
MSnset_subset <- MSnset_comb[ind]</pre>
IgG_ind <- which(pData(MSnset_subset)$SampleGroup == "IgG")</pre>
### perform regression analysis on dataset
MSnset_reg <- regressIntensity(MSnset_subset, controlInd=IgG_ind, ProteinId="P03372")</pre>
```

Corr Raw data

Corr Regressed data





transformedCorrelation

```
### Differential analysis
contrasts <- c(</pre>
  tam.2h_vs_vehicle = "tam.2h - vehicle",
 tam.6h_vs_vehicle = "tam.6h - vehicle",
  tam.24h_vs_vehicle = "tam.24h - vehicle"
suppressWarnings(diffstats <- computeDiffStats(MSnSetObj=MSnset_reg, contrasts=contrasts,</pre>
                               transform = FALSE))
diffexp1 <- getContrastResults(diffstats=diffstats, contrast=contrasts[1],</pre>
                               transform = FALSE)
diffexp1 <- diffexp1[which(diffexp1$adj.P.Val < 0.05 & abs(diffexp1$log2FC) > 0.5),]
diffexp2 <- getContrastResults(diffstats=diffstats, contrast=contrasts[2],</pre>
                               transform = FALSE)
diffexp2 <- diffexp2[which(diffexp2$adj.P.Val < 0.05 & abs(diffexp2$log2FC) > 0.5),]
diffexp3 <- getContrastResults(diffstats=diffstats, contrast=contrasts[3],</pre>
                               transform = FALSE)
diffexp3 <- diffexp3[which(diffexp3$adj.P.Val < 0.05 & abs(diffexp3$log2FC) > 0.5),]
```

6 Experiment 4: Total proteome upon OHT treatment

We performed two 10plex-TMT time-course experiments to study the effect of OHT on total protein levels. MCF7 cells were treated with OHT for 2h, 6h, 24h or for 24h with the vehicle alone (ethanol) and a total number of four biological replicates was obtained. The peptide intensities in each 10plex experiment were aggregated (by summing) to protein intensities and only proteins that were identified in both experiments were used for further analysis. The protein intensities were then normalized using median scaling to account for sample loading variability. Thereafter, differential protein expression was performed using *limma* based analysis.

```
## 1.0 a.d. d.a.t.a.
data(exp4_OHT_FP)
## create MSnSet object
MSnset_data1 <- convertToMSnset(exp4_OHT_FP$FP_1,</pre>
                                   metadata=exp4_OHT_FP$metadata_FP1,
                                   indExpData=c(7:14), Sequences=2, Accessions=6)
pData(MSnset_data1)$Run <- 1</pre>
MSnset_data2 <- convertToMSnset(exp4_OHT_FP$FP_2,</pre>
                                   metadata=exp4_OHT_FP$metadata_FP2,
                                   indExpData=c(7:14), Sequences=2, Accessions=6)
pData(MSnset_data2)$Run <- 2</pre>
## Summation of peptide to protein intensity
MSnset_P1 <- summarizeIntensities(MSnset_data1, sum, human_anno)
MSnset_P2 <- summarizeIntensities(MSnset_data2, sum, human_anno)
### Computing common unique peptides
features1 <- fData(MSnset_data1)</pre>
features1 <- as.data.frame(features1[, c("Sequences", "Accessions")],</pre>
                              stringsAsFactors = FALSE)
features2 <- fData(MSnset_data2)</pre>
features2 <- as.data.frame(features2[, c("Sequences", "Accessions")],</pre>
                              stringsAsFactors = FALSE)
features <- rbind(features1,features2)</pre>
features <- unique(features)</pre>
features$Sequences <- as.character(features$Sequences)</pre>
features$Accessions <- as.character(features$Accessions)</pre>
counts <- features %>% count(Accessions)
colnames(counts)[2] <- "Count"</pre>
##### create combine MSnSet object
MSnset_P1 <- updateFvarLabels(MSnset_P1)</pre>
MSnset_P2 <- updateFvarLabels(MSnset_P2)</pre>
MSnset_P1 <- updateSampleNames(MSnset_P1)</pre>
MSnset_P2 <- updateSampleNames(MSnset_P2)</pre>
suppressWarnings(MSnset_comb <- combine(MSnset_P1, MSnset_P2))</pre>
```

```
tokeep <- which(complete.cases(fData(MSnset_comb))==TRUE)</pre>
MSnset_comb <- MSnset_comb[tokeep,]</pre>
sampleNames(MSnset_comb) <- pData(MSnset_comb)$SampleName</pre>
fData(MSnset_comb) <- fData(MSnset_comb)[,c(1:4)]</pre>
colnames(fData(MSnset_comb)) <- c("Accessions", "Gene", "Description",</pre>
                                               "GeneSymbol")
ind <- match(fData(MSnset_comb)$Accessions,counts$Accessions)</pre>
fData(MSnset_comb)$Count <- counts$Count[ind]</pre>
## Normalization
MSnset_Pnorm <- normalizeScaling(MSnset_comb, median)</pre>
## Differential analysis
contrasts <- c(
  tam.2h_vs_vehicle = "tam.2h - vehicle",
  tam.6h_vs_vehicle = "tam.6h - vehicle",
  tam.24h_vs_vehicle = "tam.24h - vehicle"
)
batchEffect <- c("Run", "BioRep")</pre>
diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts,</pre>
                                batchEffect=batchEffect)
diffexp1 <- getContrastResults(diffstats=diffstats, contrast=contrasts[1])</pre>
diffexp1 <- diffexp1[which(diffexp1$adj.P.Val < 0.05 & abs(diffexp1$log2FC) > 0.5),]
diffexp2 <- getContrastResults(diffstats=diffstats, contrast=contrasts[2])</pre>
diffexp2 <- diffexp2[which(diffexp2$adj.P.Val < 0.05 & abs(diffexp2$log2FC) > 0.5),]
diffexp3 <- getContrastResults(diffstats=diffstats, contrast=contrasts[3])</pre>
diffexp3 <- diffexp3[which(diffexp3$adj.P.Val < 0.05 & abs(diffexp3$log2FC) > 0.5),]
```

7 Experiment 5: ER interactome in PDX tumours

An ER qPLEX-RIME experiment was performed using three independent ER+ human Patient Derived Xenograft (PDX) tumours. Cryosections of each tumour were double-crosslinked and each tumour was split in two parts that were used for ER and IgG RIME pull-down assays. One of the tumours was split in three different parts to be used as ER or IgG qPLEX-RIME in order to assess technical variability. The peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities. Thereafter, differential protein expression was performed using limma based analysis.

```
MSnset_data <- MSnset_data[,-c(7:10)]

## Normalization
MSnset_norm <- groupScaling(MSnset_data, median)

## Summation of peptide to protein intensity
MSnset_Pnorm <- summarizeIntensities(MSnset_norm, sum, human_anno)

## Differential analysis
contrasts <- c(PDX_vs_IgG = "PDX - IgG")

diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts)

## Warning in duplicateCorrelation(intensities, design = design, block = pData(MSnSetObj)$TechRep Blocks all of size 1: setting intrablock correlation to zero.

diffexp <- getContrastResults(diffstats=diffstats, contrast=contrasts)
diffexp <- diffexp[which(diffexp$adj.P.Val < 0.05 & diffexp$log2FC > 1),]
```

8 Experiment 6: ER interactome in human breast cancer tumours

An ER qPLEX-RIME experiment was performed using five independent ER-positive human breast cancer tumours. Cryosections of each tumour were double-crosslinked and each tumour was split in two parts that were used for ER and IgG RIME pull-down assays. The peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities followed by differential protein expression using *limma* based analysis.

```
## load data
data(exp6_ER)
## create MSnSet object
MSnset_data <- convertToMSnset(exp6_ER$intensities, metadata=exp6_ER$metadata,
                                 indExpData=c(6:15), Sequences=2, Accessions=5,
                                rmMissing=FALSE)
exprs(MSnset_data)[is.na(exprs(MSnset_data))] <- 0</pre>
exprs(MSnset_data) <- exprs(MSnset_data)+0.01</pre>
## Normalization
MSnset_norm <- groupScaling(MSnset_data, median, groupingColumn="SampleGroup")
## Summation of peptide to protein intensity
MSnset_Pnorm <- summarizeIntensities(MSnset_norm, sum, human_anno)
## Differential analysis
contrasts <- c(ER_vs_IgG = "ER - IgG")</pre>
diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts)</pre>
diffexp <- getContrastResults(diffstats=diffstats, contrast=contrasts)</pre>
diffexp <- diffexp[which(diffexp$adj.P.Val < 0.01 & diffexp$log2FC >1),]
```

9 Experiment 7: NCOA3 interactome in MCF7 cells

In this experiment we have used the qPLEX-RIME method to identify and characterize NCOA3 (SRC-3) associated proteins. We performed NCOA3 RIME pull-downs in five independent biological replicates and in five matched IgG mock samples. The peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities followed by differential protein expression using *limma* based analysis.

```
## load data
data(exp7_NCOA3)
## create MSnSet object
MSnset_data <- convertToMSnset(exp7_NCOA3$intensities, metadata=exp7_NCOA3$metadata,
                                indExpData=c(7:16), Sequences=2, Accessions=6,
                                rmMissing=FALSE)
exprs(MSnset_data)[is.na(exprs(MSnset_data))] <- 0</pre>
exprs(MSnset_data) <- exprs(MSnset_data)+0.01</pre>
## Normalization
MSnset_norm <- groupScaling(MSnset_data, median, groupingColumn="SampleGroup")
## Summation of peptide to protein intensity
MSnset_Pnorm <- summarizeIntensities(MSnset_norm, sum, human_anno)
## Differential analysis
contrasts <- c(NCOA3_vs_IgG = "NCOA3 - IgG")</pre>
diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts)</pre>
diffexp <- getContrastResults(diffstats=diffstats, contrast=contrasts)</pre>
diffexp <- diffexp[which(diffexp$adj.P.Val < 0.01 & diffexp$log2FC >1),]
```

10 Experiment 8: CBP interactome in MCF7 cells

A qPLEX-RIME experiment was designed for the characterization of the CBP (CREB-binding protein) interactome. Five independent biological replicates of CBP RIME pull-downs and five IgG RIME pull-downs were prepared for this experiment. The peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities followed by differential protein expression using *limma* based analysis.

```
MSnset_norm <- groupScaling(MSnset_data, median, groupingColumn="SampleGroup")

## Summation of peptide to protein intensity

MSnset_Pnorm <- summarizeIntensities(MSnset_norm, sum, human_anno)

## Differential analysis

contrasts <- c(CREBBP_vs_IgG = "CREBBP - IgG")

diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts)

diffexp <- getContrastResults(diffstats=diffstats, contrast=contrasts)

diffexp <- diffexp[which(diffexp$adj.P.Val < 0.01 & diffexp$log2FC >1),]
```

11 Experiment 9: POLR2A (RNA polymerase II) interactome in MCF7 cells

The qPLEX-RIME method was applied for the characterization of the largest and catalytic component of RNA polymerase II (RPB1). Particularly, the phosphorylated form at Serine 5 in the C-terminal domain (CTD) was used as the bait protein. Five biological replicates of RNA polymerase II RIME pull-downs and five IgG pull-downs were included for the identification and characterization of RNA polymerase II-associated proteins. The peptide intensities were normalized by median scaling within each group separately. The normalized peptide intensities were aggregated (by summing) to protein intensities followed by differential protein expression using *limma* based analysis.

```
## load data
data(exp9_PolII)
## create MSnSet object
MSnset_data <- convertToMSnset(exp9_PolII$intensities, metadata=exp9_PolII$metadata,
                                indExpData=c(7:16), Sequences=2, Accessions=6,
                                rmMissing=FALSE)
exprs(MSnset_data)[is.na(exprs(MSnset_data))] <- 0</pre>
exprs(MSnset_data) <- exprs(MSnset_data)+0.01</pre>
## Normalization
MSnset_norm <- groupScaling(MSnset_data, median, groupingColumn="SampleGroup")
## Summation of peptide to protein intensity
MSnset_Pnorm <- summarizeIntensities(MSnset_norm, sum, human_anno)
## Differential analysis
contrasts <- c(POLR2A_vs_IgG = "POLR2A - IgG")</pre>
diffstats <- computeDiffStats(MSnset_Pnorm, contrasts=contrasts)</pre>
diffexp <- getContrastResults(diffstats=diffstats, contrast=contrasts)</pre>
diffexp <- diffexp[which(diffexp$adj.P.Val < 0.01 & diffexp$log2FC >1),]
```

12 Session Information

```
sessionInfo()
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.2 LTS
## Matrix products: default
## BLAS:
         /home/biocbuild/bbs-3.13-bioc/R/lib/libRblas.so
## LAPACK: /home/biocbuild/bbs-3.13-bioc/R/lib/libRlapack.so
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8
                                   LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8
                                   LC_COLLATE=C
## [5] LC_MONETARY=en_US.UTF-8
                                   LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8
                                   LC_NAME=C
## [9] LC_ADDRESS=C
                                   LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats4
                parallel stats
                                     graphics grDevices utils
## [7] datasets methods
##
## other attached packages:
## [1] qPLEXdata_1.10.0
                             qPLEXanalyzer_1.10.0 MSnbase_2.18.0
## [4] ProtGenerics_1.24.0 S4Vectors_0.30.0
                                                  mzR_2.26.0
## [7] Rcpp_1.0.6
                             Biobase_2.52.0
                                                  BiocGenerics_0.38.0
## [10] dplyr_1.0.6
##
## loaded via a namespace (and not attached):
## [1] ggdendro_0.1.22
                               vsn_3.60.0
## [3] tidyr_1.1.3
                               splines_4.1.0
## [5] foreach_1.5.1
                               assertthat_0.2.1
## [7] statmod_1.4.36
                               highr_0.9
## [9] BiocManager_1.30.15
                               affy_1.70.0
## [11] GenomeInfoDbData_1.2.6 impute_1.66.0
## [13] pillar_1.6.1
                               lattice_0.20-44
## [15] glue_1.4.2
                               limma_3.48.0
## [17] digest_0.6.27
                               RColorBrewer_1.1-2
## [19] XVector_0.32.0
                               colorspace_2.0-1
## [21] preprocessCore_1.54.0 plyr_1.8.6
## [23] MALDIquant_1.19.3
                               XML_3.99-0.6
## [25] pkgconfig_2.0.3
                               zlibbioc_1.38.0
## [27] purrr_0.3.4
                               scales_1.1.1
## [29] affyio_1.62.0
                               BiocParallel_1.26.0
## [31] tibble_3.1.2
                               generics_0.1.0
## [33] IRanges_2.26.0
                               ggplot2_3.3.3
## [35] ellipsis_0.3.2
                               magrittr_2.0.1
```

```
## [37] crayon_1.4.1
                                evaluate_0.14
## [39] ncdf4_1.17
                               fansi_0.4.2
## [41] doParallel_1.0.16
                               MASS_7.3-54
## [43] tools_4.1.0
                               hms_1.1.0
## [45] lifecycle_1.0.0
                                stringr_1.4.0
## [47] munsell_0.5.0
                                cluster_2.1.2
## [49] Biostrings_2.60.0
                               pcaMethods_1.84.0
## [51] compiler_4.1.0
                               GenomeInfoDb_1.28.0
## [53] mzID_1.30.0
                               rlang_0.4.11
## [55] grid_4.1.0
                               RCurl_1.98-1.3
## [57] iterators_1.0.13
                               MsCoreUtils_1.4.0
## [59] bitops_1.0-7
                               gtable_0.3.0
## [61] codetools_0.2-18
                               DBI_1.1.1
## [63] R6_2.5.0
                               knitr_1.33
## [65] utf8_1.2.1
                               clue_0.3-59
## [67] readr_1.4.0
                               stringi_1.6.2
## [69] vctrs_0.3.8
                               tidyselect_1.1.1
## [71] xfun_0.23
```