

Figure4_v2

Nikita Verheyden

2023-06-13

Setup

dir

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Figure4")
```

Packages

```
source("D:/Krueger_Lab/Publications/miR181_paper/Supporting_scripts/themes/theme_paper.R")
library(ggplot2)
library(rtracklayer)
```

```
## Loading required package: GenomicRanges
## Loading required package: stats4
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:stats':
## 
##     IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
## 
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following objects are masked from 'package:base':
## 
##     expand.grid, I, unname
## Loading required package: IRanges
##
## Attaching package: 'IRanges'
```

```

## The following object is masked from 'package:grDevices':
##
##      windows

## Loading required package: GenomeInfoDb
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:GenomicRanges':
##
##      intersect, setdiff, union

## The following object is masked from 'package:GenomeInfoDb':
##
##      intersect

## The following objects are masked from 'package:IRanges':
##
##      collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':
##
##      first, intersect, rename, setdiff, setequal, union

## The following objects are masked from 'package:BiocGenerics':
##
##      combine, intersect, setdiff, union

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

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

library(tibble)
library(Rtsne)
library(cliProfiler)
library(GenomicRanges)
library(cowplot)
library(eulerr)

```

Data

```

#Ribo profiling
RNA <- read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RNA_masterframe.csv")
RPF <- read.csv("D:/Krueger_Lab/Publications/miR181_paper/Figure3/RPF_masterframe.csv")

#load the gtf file to compare genes
gff23 <- import.gff3("D:/Krueger_Lab/Ribo_Profiling/run15112022M23/ref_genome/gencode.vM23.annotation.gff3")
gff23path <- file.path("D:/Krueger_Lab/Ribo_Profiling/run15112022M23/ref_genome", "gencode.vM23.annotation.gff3")

#targets
target <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure3/mir181_bs_with_seeds.rds")

```

```

largetframe <- as.data.frame(larget)
table(largetframe$set)

##
##          ago_bs_mir181_chi    ago_bs_mir181_chi&mir181_enriched
##                      5815                                1082
##          mir181_enriched
##                      3576

largetframe <- largetframe[!(largetframe$set == "ago_bs_mir181_chi"),]

# modify dataset to only include working targetsets

#only use miR181_enriched und both
# largetframe <- largetframe[largetframe$set %in% c("mir181_enriched", "ago_bs_mir181_chi&mir181_enriched")]

#MMSat4
repeat_masker <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure2/MMSat4/repeat_masker.rds")
MMSAT4 <- repeat_masker[repeat_masker$repName == "MMSAT4"]
MurSatRep1 <- repeat_masker[repeat_masker$repName == "MurSatRep1"]

```

colours

```

#colours
farbeneg <- "#b4b4b4"
farbe1 <- "#0073C2FF"
farbe2 <- "#EFC000FF"
farbe3 <- "#CD534CFF"
farbe4 <- "#7AA6DCFF"
farbe5 <- "#868686FF"
farbe6 <- "#003C67FF"
farbe7 <- "#8F7700FF"
farbe8 <- "#3B3B3BFF"
farbe9 <- "#A73030FF"
farbe10 <- "#4A6990FF"
farbe11 <- "#FF6FOOFF"
farbe12 <- "#C71000FF"
farbe13 <- "#008EA0FF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"

RNAPcol <- "#b56504"
RNAncol <- "#027d73"
RPFpcol <- "#c4c404"
RPFnncol <- "#8d0391"

```

axis limits

```

xlim <- 0.5
zoomfac <- 0.45

```

number of binding sites

```
bsnum <- as.data.frame(table(largetframe$geneName))
colnames(bsnum) <- c("gene_symbol", "BS_number")
```

wobble plot

ecdf

```
#RNA
```

```
RNA$wobble <- "Non-target"
RNA$wobble[RNA$gene_symbol %in% largetframe[is.na(largetframe$first_seed_200down.wobble) , "geneName"]] <- "Non-target"
RNA$wobble[RNA$gene_symbol %in% largetframe[largetframe$first_seed_200down.wobble == FALSE , "geneName"]] <- "Non-target"
RNA$wobble[RNA$gene_symbol %in% largetframe[largetframe$first_seed_200down.wobble == TRUE , "geneName"]] <- "Target"
RNA$wobble[RNA$gene_symbol %in% bsnum[bsnum$BS_number > 1, "gene_symbol"]]] <- "Multiple"
#RPF
RPF$wobble <- "Non-target"
RPF$wobble[RPF$gene_symbol %in% largetframe[is.na(largetframe$first_seed_200down.wobble) , "geneName"]] <- "Non-target"
RPF$wobble[RPF$gene_symbol %in% largetframe[largetframe$first_seed_200down.wobble == FALSE , "geneName"]] <- "Non-target"
RPF$wobble[RPF$gene_symbol %in% largetframe[largetframe$first_seed_200down.wobble == TRUE , "geneName"]] <- "Target"
RPF$wobble[RPF$gene_symbol %in% bsnum[bsnum$BS_number > 1, "gene_symbol"]]] <- "Multiple"
```

```
#RPF
```

```
wobbleecdfRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(wobble, levels = c("Non-target", "Target")))) +
  stat_ecdf(geom="step", linewidth=2) +
  geom_vline(xintercept = 0, linewidth=1, linetype="dashed", colour=farbeneg) +
  scale_colour_manual(values = c("black", farbeneg, farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-xlim, xlim)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange K0 vs WT") +
  ggtitle("Target genes RPFseq")
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
```

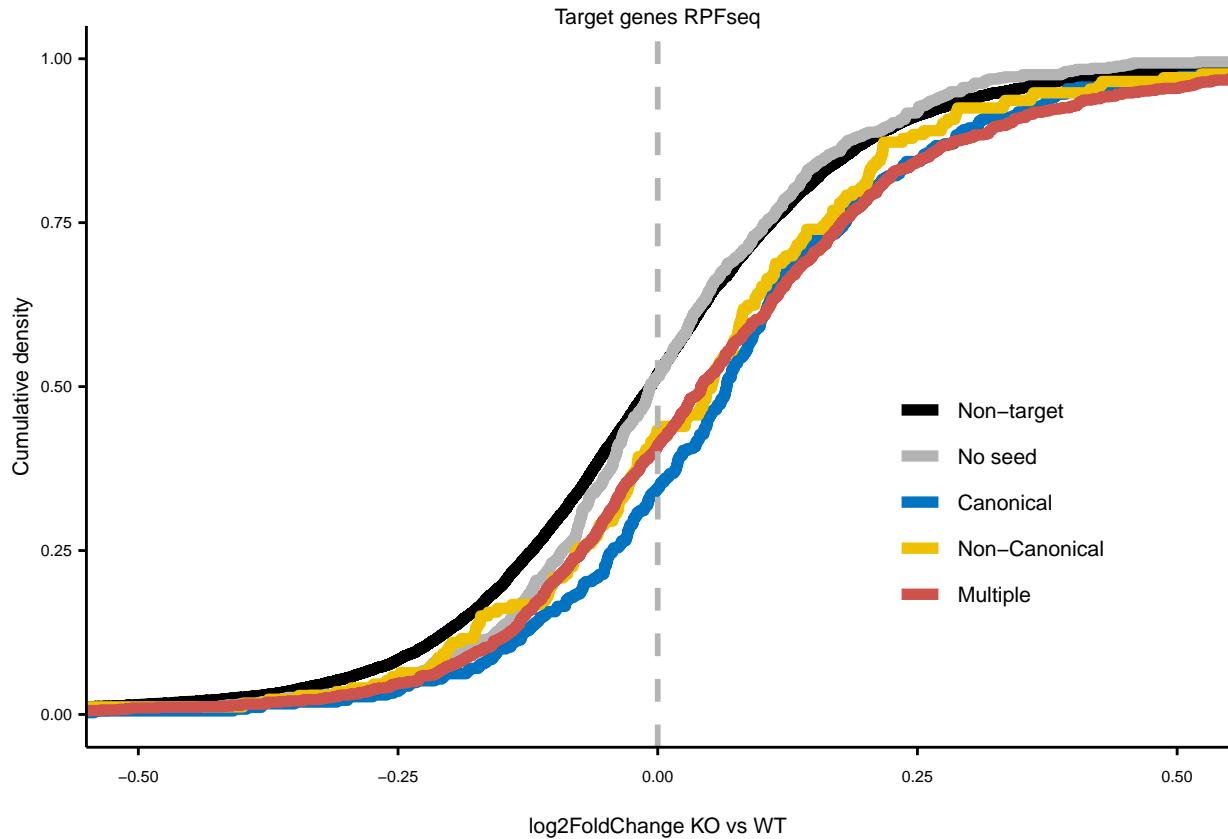
```
## i Please use the `linewidth` argument instead.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

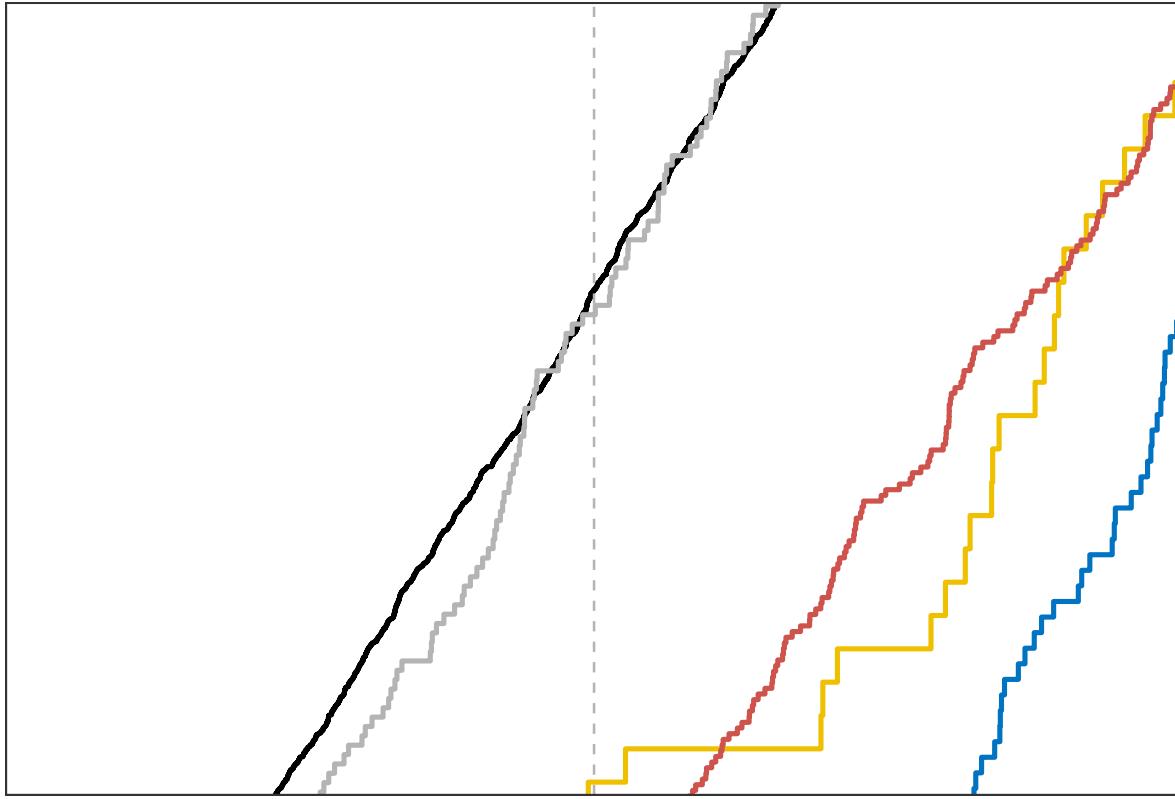
```
## generated.
```

```
wobbleecdfRPF
```



```
#zoom
wobbleecdfRPFZ <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(wobble, levels = c("Non-target", "No seed", "Canonical", "Non-Canonical", "Multiple")))) +
  stat_ecdf(geom="step", linewidth=2*zoomfac) +
  geom_vline(xintercept = 0, linewidth=1*zoomfac, linetype="dashed", colour=farbeneg) +
  scale_colour_manual(values = c("black", farbeneg, farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-(xlim/8), (xlim/8)), ylim = c(0.5-1/16, 0.5+1/16)) +
  theme_bw() +
  theme(legend.position = "none", axis.text = element_blank(), panel.grid = element_blank(), axis.ticks = element_blank(),
        panel.background = element_rect(fill='transparent'), #transparent panel bg
        plot.background = element_rect(fill='transparent', color=NA)) +
  scale_y_continuous("") + scale_x_continuous("")
```

wobbleecdfRPFZ



```
#export RPF
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/wobbleECDF_RPF.pdf", width=2, height = 2)
wobbleecdfRPF
dev.off()

## pdf
## 2
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/wobbleECDF_RPF_zoom.pdf", width=2*zoomfac, height=2*zoomfac)
wobbleecdfRPFZ
dev.off()

## pdf
## 2
table(RPF$wobble)

##
##      Canonical       Multiple       No seed Non-Canonical       Non-target
##            370           1016           616            173            9194
```

by type of MRE (wobble and non wobble combined)

```
#RNA
RNA$MREtype <- "Non-target"
RNA$MREtype [RNA$gene_symbol %in% targetframe[is.na(targetframe$first_seed_200down.type), "geneName"]] <-
RNA$MREtype [RNA$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_6mer", "geneName"]]
RNA$MREtype [RNA$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_7mer_a1", "geneName"]]
```

```

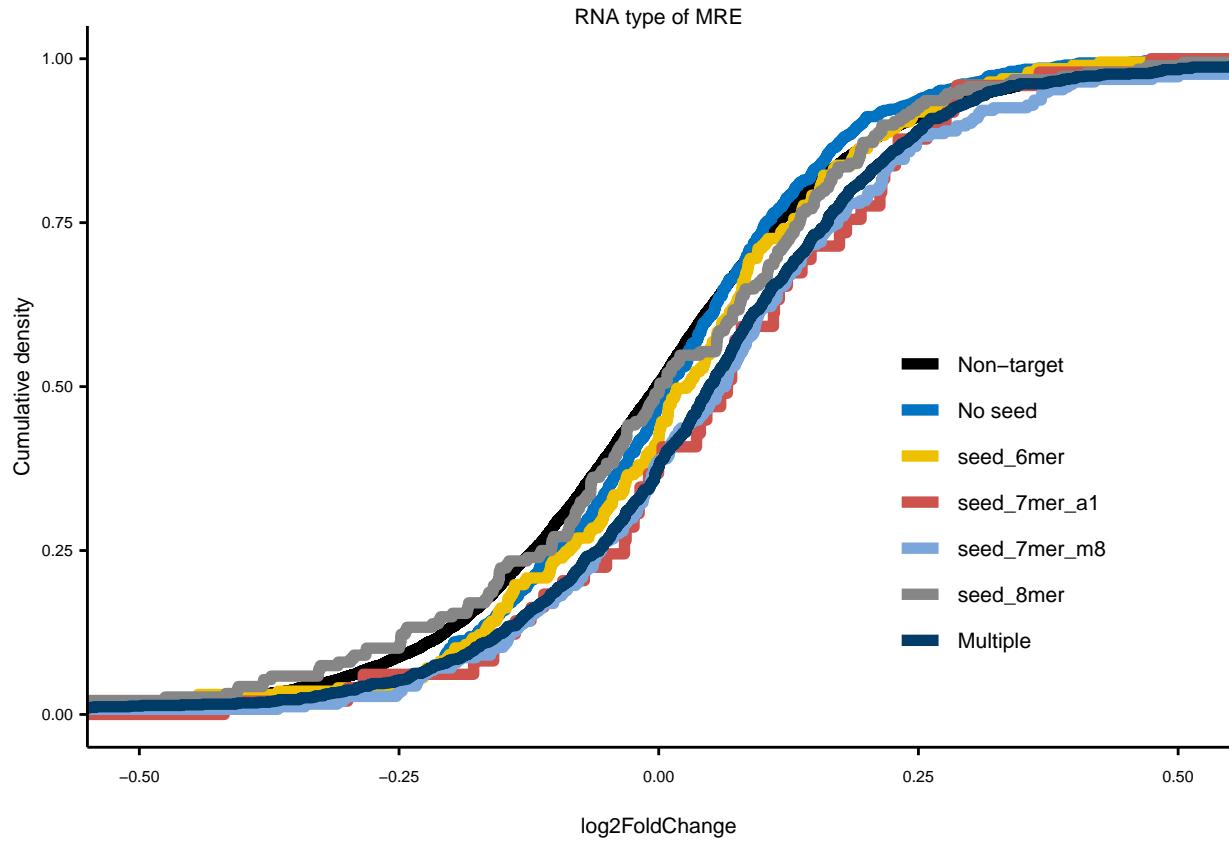
RNA$MREtype[RNA$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_7mer_m8","geneName"] ] <- "seed_7mer_m8"
RNA$MREtype[RNA$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_8mer","geneName"] ] <- "seed_8mer"
RNA$MREtype[RNA$gene_symbol %in% bsnum[bsnum$BS_number > 1, "gene_symbol"] ] <- "Multiple"

#RPF
RPF$MREtype <- "Non-target"
RPF$MREtype[RPF$gene_symbol %in% targetframe[is.na(targetframe$first_seed_200down.type), "geneName"] ] <- "No seed"
RPF$MREtype[RPF$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_6mer","geneName"] ] <- "seed_6mer"
RPF$MREtype[RPF$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_7mer_a1","geneName"] ] <- "seed_7mer_a1"
RPF$MREtype[RPF$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_7mer_m8","geneName"] ] <- "seed_7mer_m8"
RPF$MREtype[RPF$gene_symbol %in% targetframe[targetframe$first_seed_200down.type == "seed_8mer","geneName"] ] <- "seed_8mer"
RPF$MREtype[RPF$gene_symbol %in% bsnum[bsnum$BS_number > 1, "gene_symbol"] ] <- "Multiple"

# ecdf plots
#RNA
typeECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange),
                                colour=factor(MREtype, levels = c("Non-target", "No seed", "seed_6mer", "seed_7mer", "seed_8mer"))),
                        stat_ecdf(geom="step", linewidth=2) +
                        scale_colour_manual(values = c("black", farbe1, farbe2, farbe3, farbe4, farbe5, farbe6)) +
                        coord_cartesian(xlim = c(-xlim, xlim)) +
                        theme_paper() +
                        scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
                        ggtitle("RNA type of MRE"))

typeECDFRNA

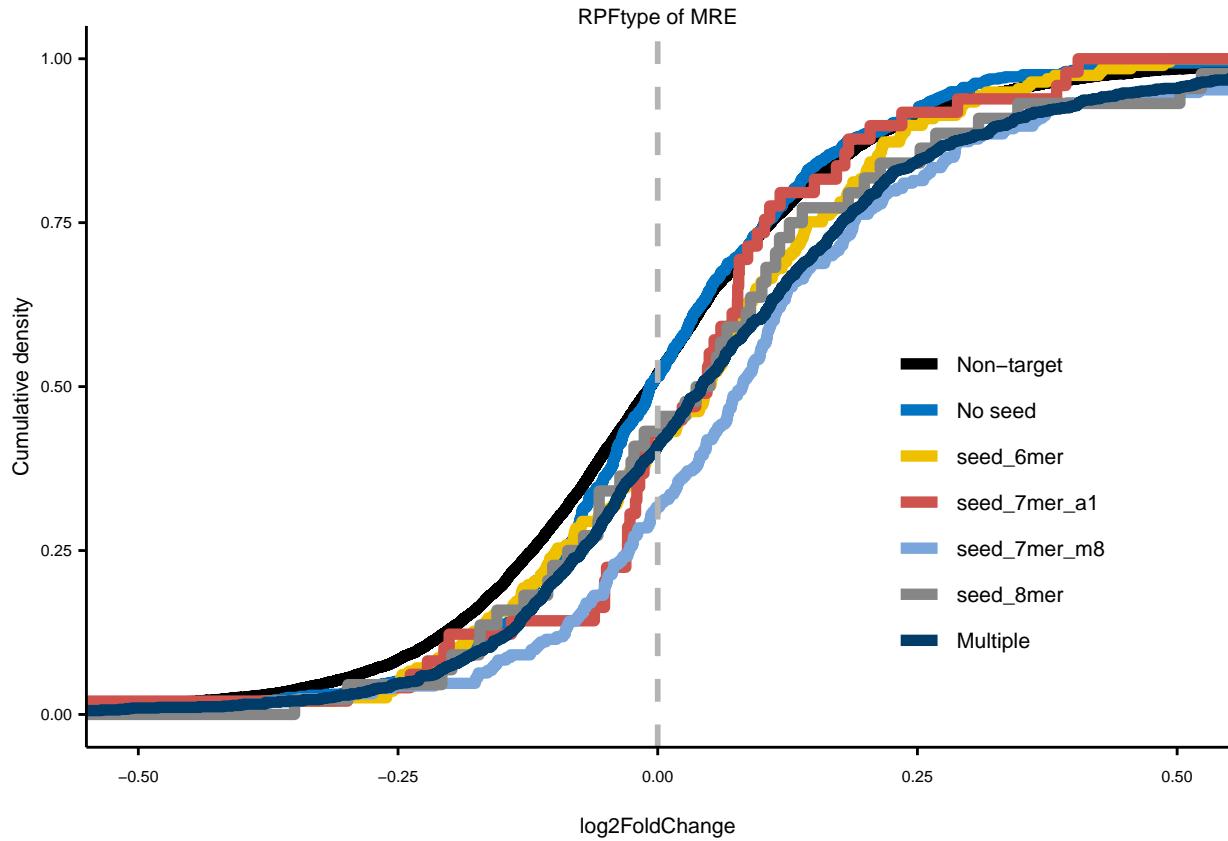
```



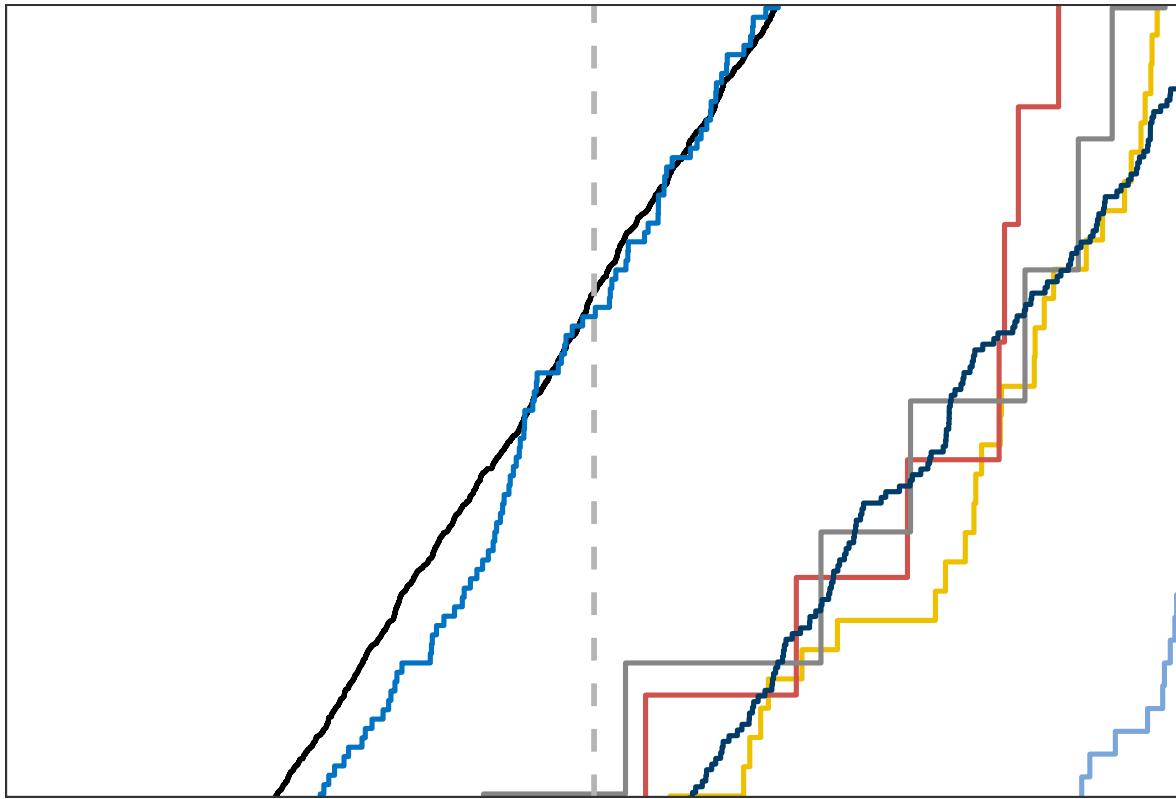
#RPF

```
typeECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange),
                               colour=factor(MREtype, levels = c("Non-target", "No seed", "seed_6mer", "seed_7mer_a1", "seed_7mer_m8", "seed_8mer", "Multiple")),
                               stat_ecdf(geom="step", linewidth=2) +
                               geom_vline(xintercept = 0, linewidth=1, linetype="dashed", colour=farbeneg) +
                               scale_colour_manual(values = c("black", farbe1, farbe2, farbe3, farbe4, farbe5, farbe6)) +
                               coord_cartesian(xlim = c(-xlim, xlim)) +
                               theme_paper() +
                               scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
                               ggtitle("RPFtype of MRE"))

typeECDFRPF
```



```
# zoom
typeECDFRPFZ <- ggplot(RPF, aes(as.numeric(log2FoldChange),
                                colour=factor(MREtype, levels = c("Non-target", "No seed", "seed_6mer", "seed_7mer_a1", "seed_7mer_m8", "seed_8mer", "Multiple")),
                                stat_ecdf(geom="step", linewidth=2*zoomfac) +
                                geom_vline(xintercept = 0, linewidth=1, linetype="dashed", colour=farbeneg) +
                                scale_colour_manual(values = c("black", farbe1, farbe2, farbe3, farbe4, farbe5, farbe6)) +
                                coord_cartesian(xlim = c(-xlim/8, (xlim/8)), ylim = c(0.5-1/16, 0.5+1/16)) +
                                theme_bw() +
                                theme(legend.position = "none", axis.text = element_blank(), panel.grid = element_blank(), axis.ticks = element_blank(),
                                      panel.background = element_rect(fill='transparent'), #transparent panel bg
                                      plot.background = element_rect(fill='transparent', color=NA)) +
                                scale_y_continuous("") + scale_x_continuous(""))
typeECDFRPFZ
```



```

#export RPF
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/typeECDF_RPF.pdf", width=2, height = 2)
typeECDFRPF
dev.off()

## pdf
## 2
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/typeECDF_RPF_zoom.pdf", width=2*zoomfac, height =
typeECDFRPFZ
dev.off()

## pdf
## 2
table(RPF$MREtype)

##
##      Multiple      No seed    Non-target    seed_6mer seed_7mer_a1 seed_7mer_m8
##      1016          616        9194         197           49          253
##      seed_8mer
##          44

```

mmsat4

```

mmsat4frame <- as.data.frame(subsetByOverlaps(gff23, MMSAT4))

```

```

#RNA
RNA$tvsymsat4 <- "Non-target"
RNA$tvsymsat4[RNA$gene_symbol %in% mmsat4frame$gene_name] <- "MMsat4"
RNA$tvsymsat4[RNA$gene_symbol %in% targetframe$geneName] <- "miR-181 target"
RNA$tvsymsat4[RNA$gene_symbol %in% targetframe$geneName & RNA$gene_symbol %in% mmsat4frame$gene_name] <- "both miR-181 target"
table(RNA$tvsymsat4)

##
##          both miR-181 target      MMsat4     Non-target
##            81             2107        163         10950

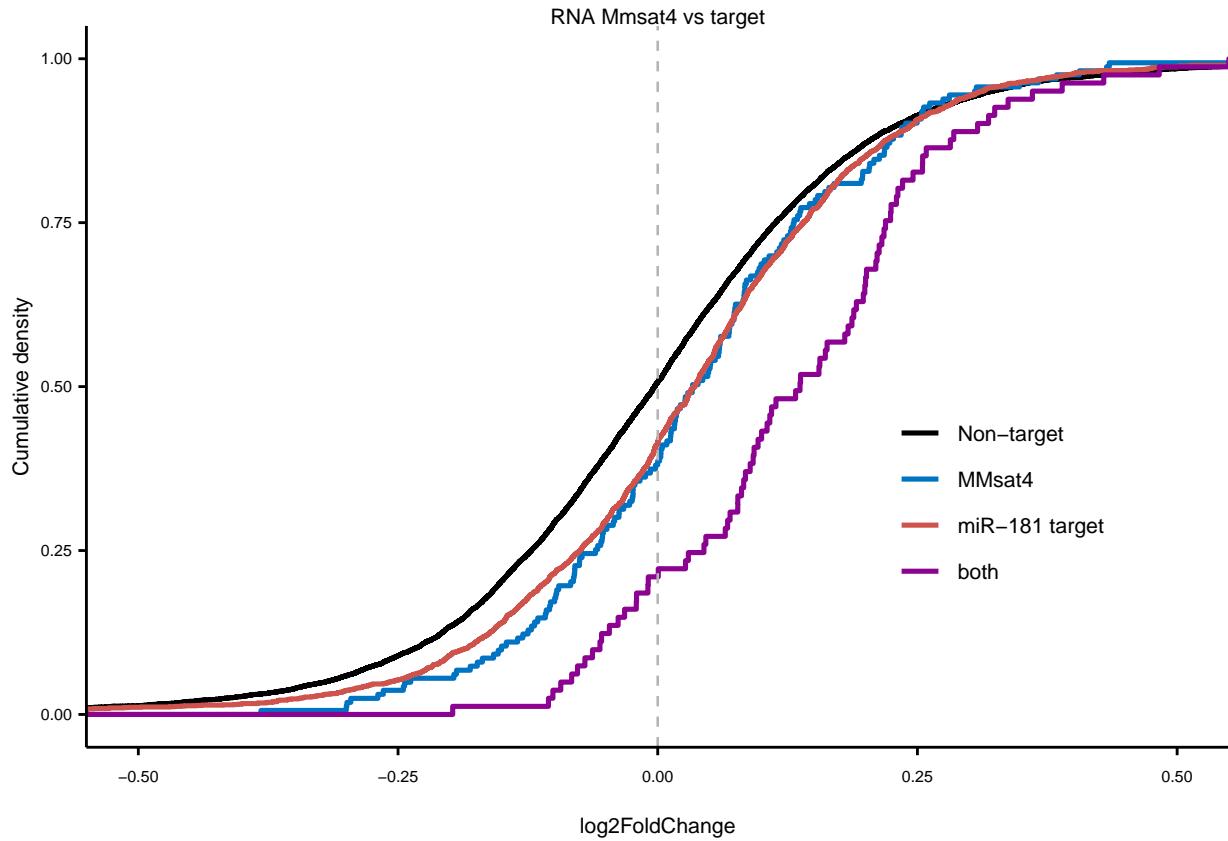
#RPF
RPF$tvsymsat4 <- "Non-target"
RPF$tvsymsat4[RPF$gene_symbol %in% mmsat4frame$gene_name] <- "MMsat4"
RPF$tvsymsat4[RPF$gene_symbol %in% targetframe$geneName] <- "miR-181 target"
RPF$tvsymsat4[RPF$gene_symbol %in% targetframe$geneName & RPF$gene_symbol %in% mmsat4frame$gene_name] <- "both miR-181 target"
table(RPF$tvsymsat4)

##
##          both miR-181 target      MMsat4     Non-target
##            79             2096        152         9042

#RNA
to1ECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(tvsymsat4, levels = c("Non-target", "MMsat4", "miR-181 target", "both miR-181 target")))) +
  stat_ecdf(geom="step", linewidth=2*zoomfac) +
  geom_vline(xintercept = 0, linewidth=1*zoomfac, linetype="dashed", colour=farbeneg) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-xlim, xlim)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA MMsat4 vs target")

to1ECDFRNA

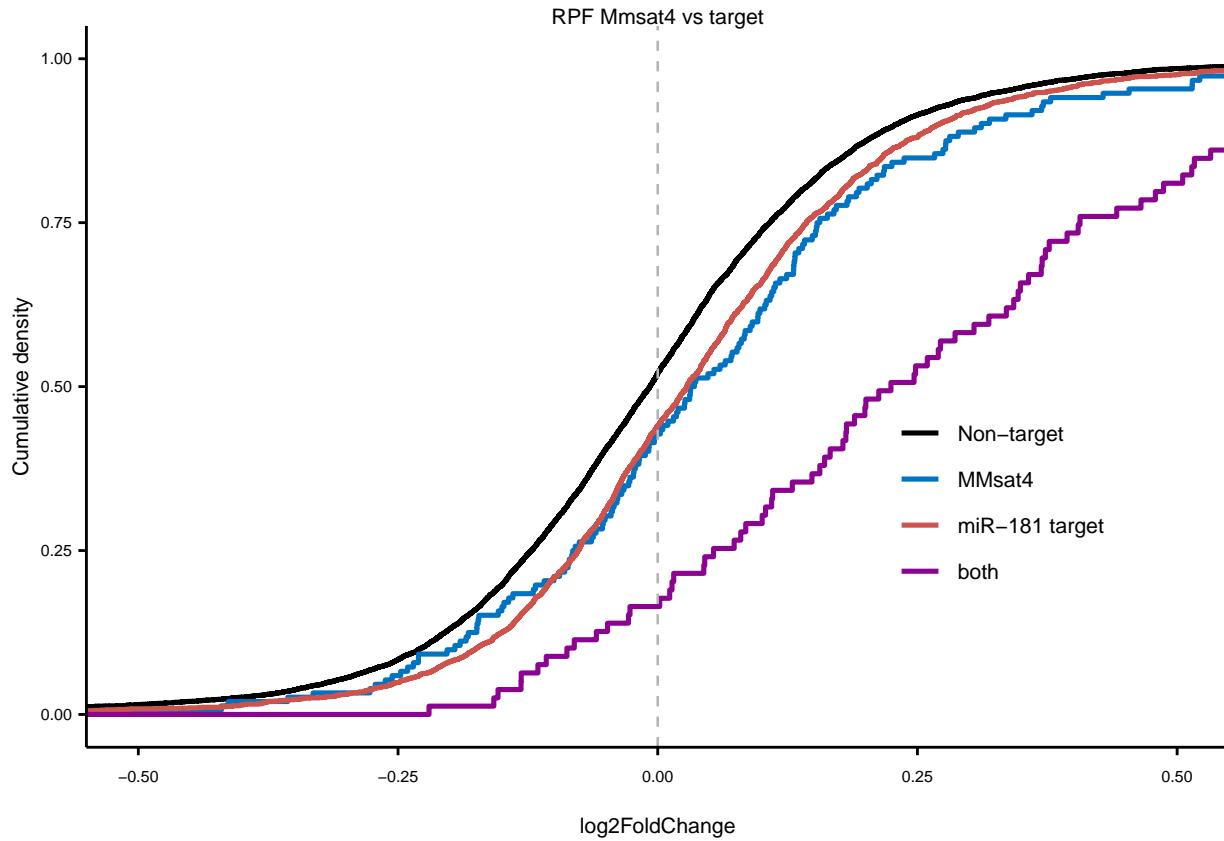
```



#RPF

```
to1ECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(tvsmmsat4, levels = c("Non-target", "MMsat4", "miR-181 target", "both")))) +  
  stat_ecdf(geom="step", linewidth=2*zoomfac) +  
  geom_vline(xintercept = 0, linewidth=1*zoomfac, linetype="dashed", colour=farbeneg) +  
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +  
  coord_cartesian(xlim = c(-xlim, xlim)) +  
  theme_paper() +  
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +  
  ggtitle("RPF Mmsat4 vs target")
```

to1ECDFRPF



```
#export RPF
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/MMSAT4ECDF_RPF.pdf", width=2, height = 2)
toECDFRPF
dev.off()

## pdf
## 2
table(RPF$tvsMmsat4)

##
##          both miR-181 target      MMsat4      Non-target
##            79        2096       152        9042
```

MurSatRep1

```
MurSatRep1frame <- as.data.frame(subsetByOverlaps(gff23, MurSatRep1))

#RNA
RNA$tvsMurSatRep1 <- "Non-target"
RNA$tvsMurSatRep1[RNA$gene_symbol %in% MurSatRep1frame$gene_name] <- "MurSatRep1"
RNA$tvsMurSatRep1[RNA$gene_symbol %in% largetframe$geneName] <- "miR-181 target"
RNA$tvsMurSatRep1[RNA$gene_symbol %in% largetframe$geneName & RNA$gene_symbol %in% MurSatRep1frame$gene]
table(RNA$tvsMurSatRep1)

##
##          both miR-181 target      MurSatRep1      Non-target
```

```

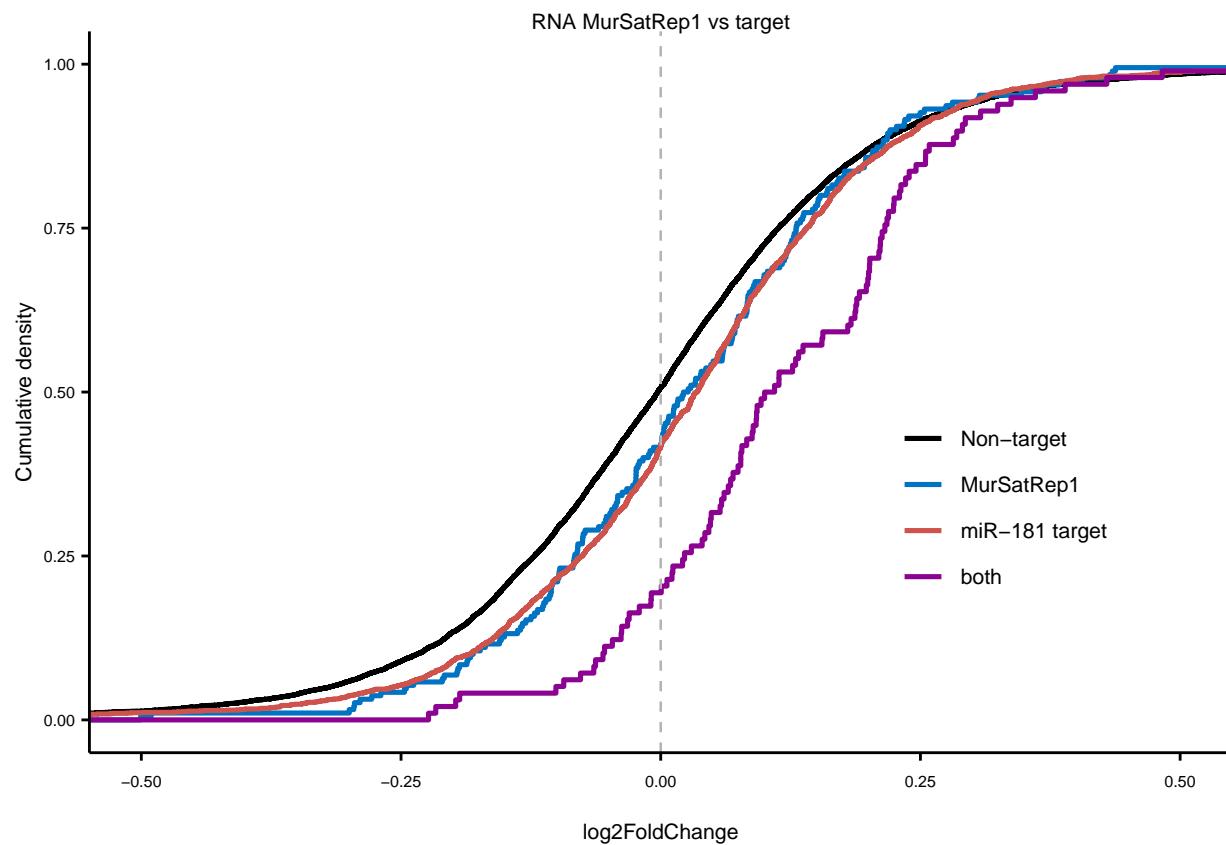
##          98        2090       190      10923
#RPF
RPF$tvsMurSatRep1 <- "Non-target"
RPF$tvsMurSatRep1[RPF$gene_symbol %in% MurSatRep1frame$gene_name] <- "MurSatRep1"
RPF$tvsMurSatRep1[RPF$gene_symbol %in% largetframe$geneName] <- "miR-181 target"
RPF$tvsMurSatRep1[RPF$gene_symbol %in% largetframe$geneName & RPF$gene_symbol %in% MurSatRep1frame$gene_name]
table(RPF$tvsMurSatRep1)

##
##          both miR-181 target      MurSatRep1      Non-target
##          96        2079        178        9016

#RNA
murECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(tvsMurSatRep1, levels = c("Non-target", "MurSatRep1", "miR-181 target", "both")))) +
  stat_ecdf(geom="step", linewidth=2*zoomfac) +
  geom_vline(xintercept = 0, linewidth=1*zoomfac, linetype="dashed", colour=farbeneg) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-xlim, xlim)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA MurSatRep1 vs target")

murECDFRNA

```



```

#RPF
murECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(tvsMurSatRep1, levels = c("Non-target", "MurSatRep1", "miR-181 target", "both")))) +
  stat_ecdf(geom="step", linewidth=2*zoomfac) +

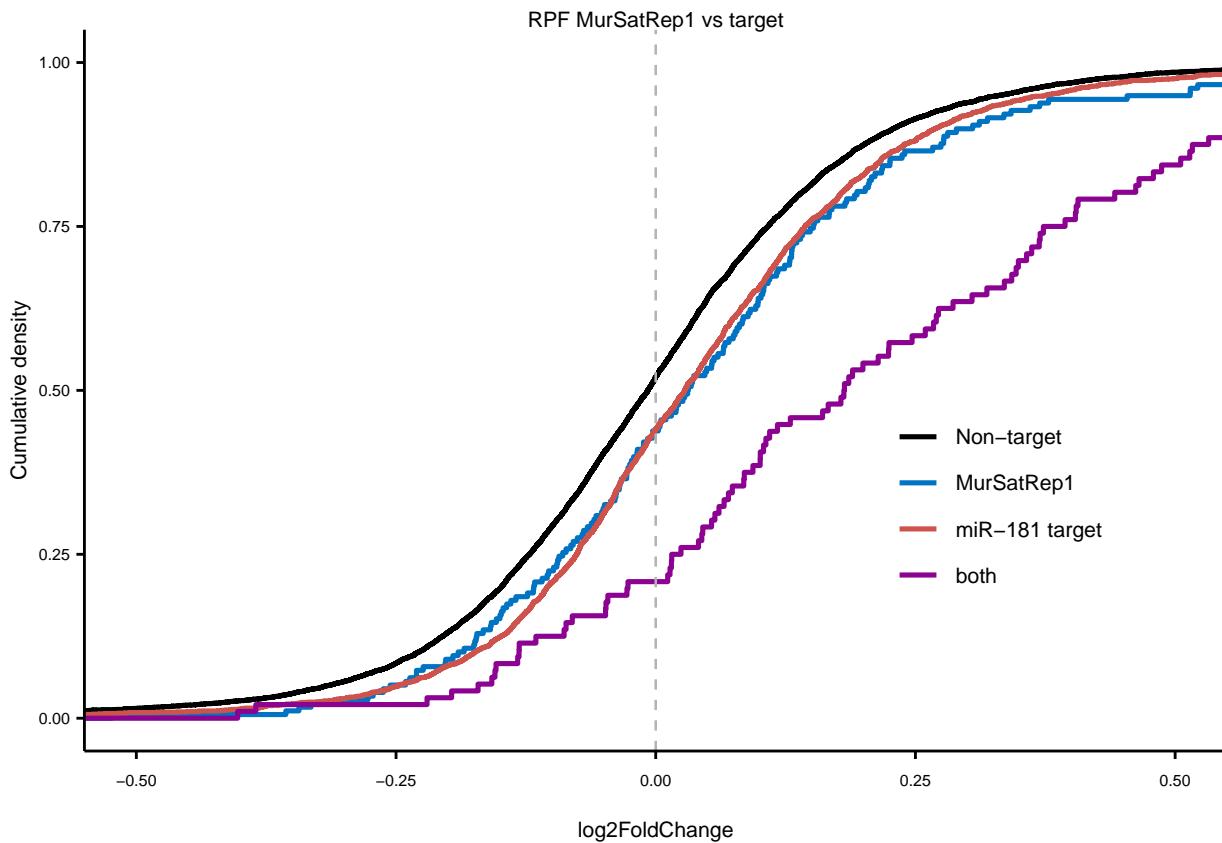
```

```

geom_vline(xintercept = 0, linewidth=1*zoomfac, linetype="dashed", colour=farbeneg) +
scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
coord_cartesian(xlim = c(-xlim, xlim)) +
theme_paper() +
scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
ggtitle("RPF MurSatRep1 vs target")

```

`murECDFRPF`



```

#export RPF
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/mursatrep1ECDF_RPF.pdf", width=2, height = 2)
murECDFRPF
dev.off()

```

```

## pdf
## 2
table(RPF$tvsMurSatRep1)

```

	both	miR-181 target	MurSatRep1	Non-target
##	96	2079	178	9016

metagene plot

```

MMSAT4$id <- "MMsat4"
MurSatRep1$id <- "MurSatRep1"

```

```

metaob <- c(MMSAT4, MurSatRep1)

metael <- metaGeneProfile(object = metaob, annotation = gff23path, group = "id")
metasat4 <- metaGeneProfile(object = MMSAT4, annotation = gff23path)
metamur <- metaGeneProfile(object = MurSatRep1, annotation = gff23path)

metasat4

## $Peaks
## GRanges object with 1693 ranges and 17 metadata columns:
##          seqnames      ranges strand |  swScore  milliDiv
##             <Rle>      <IRanges>  <Rle> | <integer> <numeric>
## [1]     chr1    49003992-49004114   - |      272     295
## [2]     chr1    49004329-49004605   - |      402     284
## [3]     chr1    82366411-82366554   - |      395     248
## [4]     chr1    86014673-86014899   - |      382     314
## [5]     chr1   117749987-117750291   - |      312     304
## ...
## [1689] ...       ...     ...   . |     ...     ...
## [1690] chr4_KZ289068_fix    531409-532565   + |     1261     87
## [1691] chr7_JH792828_fix    149909-150032   + |     282     294
## [1692] chr7_JH792828_fix    150332-150465   + |     262     358
## [1693] chrY_JH792832_fix    349767-349817   + |     230     180
##          milliDel milliIns genoLeft repName repClass repFamily
##             <numeric> <numeric> <integer> <character> <character> <character>
## [1]      57        8 -146467857  MMSAT4 Satellite  Satellite
## [2]      32       27 -146467366  MMSAT4 Satellite  Satellite
## [3]      42       21 -113105417  MMSAT4 Satellite  Satellite
## [4]       0        0 -109457072  MMSAT4 Satellite  Satellite
## [5]      30       27 -77721680  MMSAT4 Satellite  Satellite
## ...
## [1689]     0        0      -7  MMSAT4 Satellite  Satellite
## [1690]     40       40 -323706  MMSAT4 Satellite  Satellite
## [1691]     0        0 -323273  MMSAT4 Satellite  Satellite
## [1692]     22       29 -323017  MMSAT4 Satellite  Satellite
## [1693]     0       20 -194372  MMSAT4 Satellite  Satellite
##          repStart repEnd repLeft      id center location
##             <integer> <integer> <integer> <character> <integer> <character>
## [1]      -5     163     35 MMsat4 49004054      NO
## [2]     -13     278      1 MMsat4 49004467      NO
## [3]      0     168     22 MMsat4 82366483      NO
## [4]      0     227      1 MMsat4 86014787      NO
## [5]      0     305      1 MMsat4 117750139     NO
## ...
## [1689]     1    1157      0 MMsat4 531987      NO
## [1690]    31     154     -14 MMsat4 149971      NO
## [1691]    31     164     -4 MMsat4 150399      NO
## [1692]    31     167     -1 MMsat4 150653      NO
## [1693]   115     164     -4 MMsat4 349793      NO
##          Gene_ID Transcript_ID Position
##             <character> <character> <numeric>
## [1]      Nan      <NA>      5
## [2]      Nan      <NA>      5

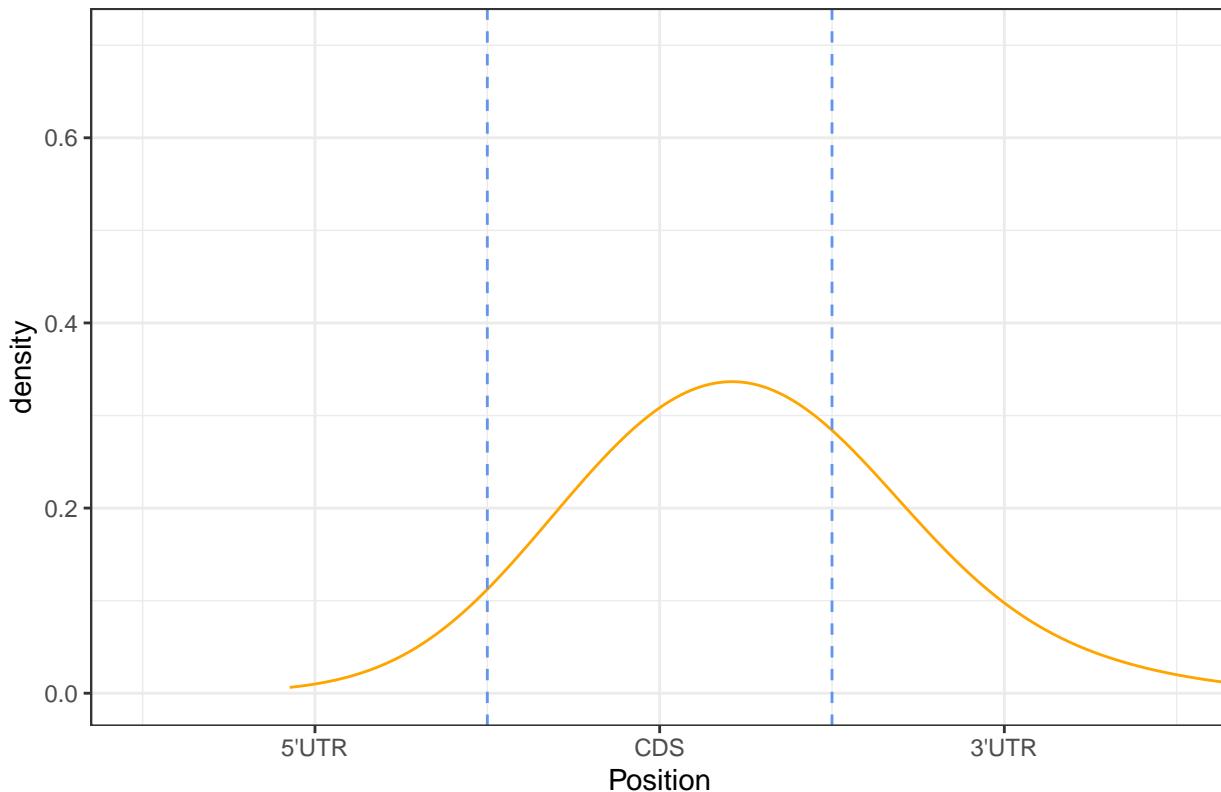
```

```

##      [3]     Nan     <NA>     5
##      [4]     Nan     <NA>     5
##      [5]     Nan     <NA>     5
##      ...     ...     ...     ...
##  [1689]     Nan     <NA>     5
##  [1690]     Nan     <NA>     5
##  [1691]     Nan     <NA>     5
##  [1692]     Nan     <NA>     5
##  [1693]     Nan     <NA>     5
##
## -----
## seqinfo: 239 sequences (1 circular) from mm10 genome
##
## $Plot

```

Meta Gene Profile



metamur

```

## $Peaks
## GRanges object with 2701 ranges and 17 metadata columns:
##           seqnames      ranges strand |  swScore  milliDiv
##           <Rle>      <IRanges> <Rle> | <integer> <numeric>
##      [1]    chr1  3950995-3951044   - |    249     220
##      [2]    chr1  7035216-7035274   - |    270     107
##      [3]    chr1 11458214-11458331   - |    564     188
##      [4]    chr1 49004231-49004449   - |    255     303
##      [5]    chr1 49004708-49004792   - |    281     286
##      ...     ...     ...     ... |    ...
##  [2697] chrY_JH584301_random 71770-71865   + |    247     302

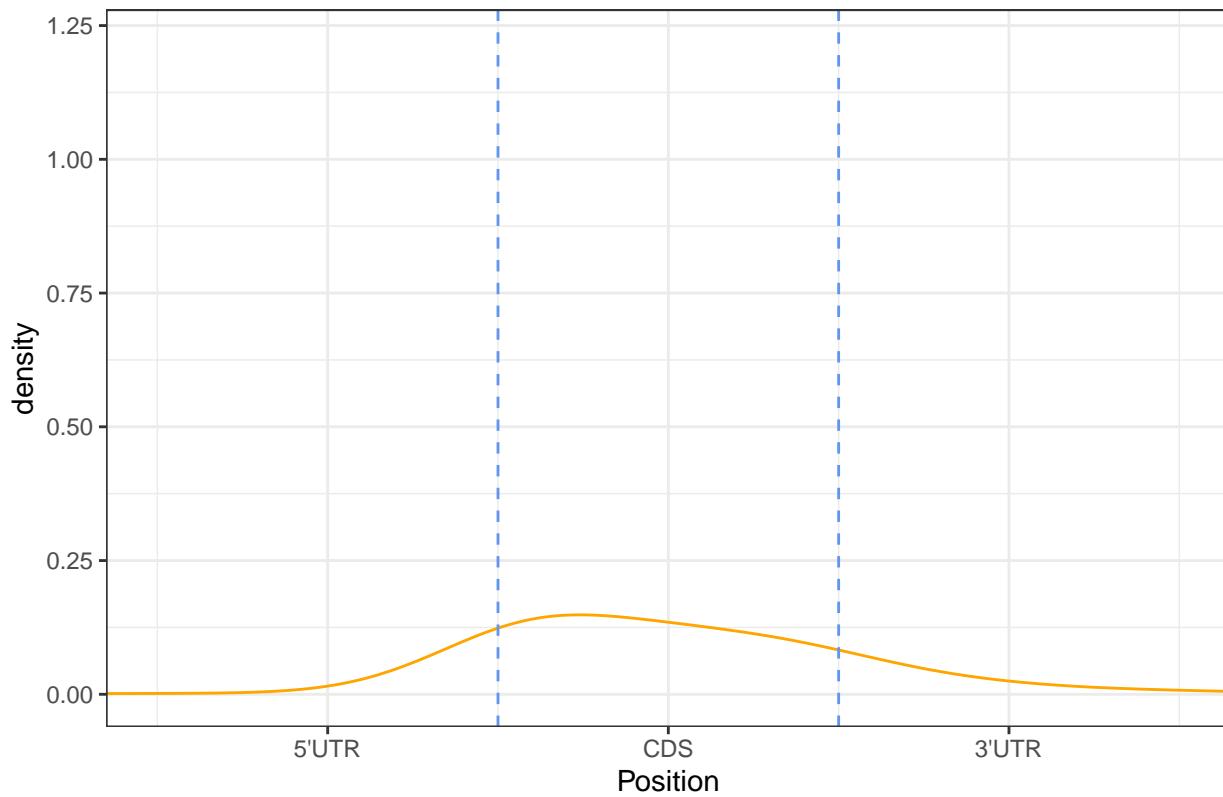
```

```

## [2698] chr3_KZ289066_fix 325612-325675 + | 385 172
## [2699] chr4_KZ289068_fix 110486-110611 + | 259 344
## [2700] chr16_KB469741_fix 436227-436289 + | 236 254
## [2701] chrY_JH792832_fix 347783-347884 + | 295 290
## milliDel milliIns genoLeft repName repClass repFamily
## <numeric> <numeric> <integer> <character> <character> <character>
## [1] 0 0 -191520927 MurSatRep1 Unknown Unknown
## [2] 85 51 -188436697 MurSatRep1 Unknown Unknown
## [3] 68 8 -184013640 MurSatRep1 Unknown Unknown
## [4] 91 50 -146467522 MurSatRep1 Unknown Unknown
## [5] 12 12 -146467179 MurSatRep1 Unknown Unknown
## ... ...
## [2697] 10 0 -188010 MurSatRep1 Unknown Unknown
## [2698] 0 0 -13559 MurSatRep1 Unknown Unknown
## [2699] 8 8 -421961 MurSatRep1 Unknown Unknown
## [2700] 45 0 -69587 MurSatRep1 Unknown Unknown
## [2701] 10 20 -196305 MurSatRep1 Unknown Unknown
## repStart repEnd repLeft id center location
## <integer> <integer> <integer> <character> <integer> <character>
## [1] -1395 742 693 MurSatRep1 3951020 NO
## [2] -1845 292 232 MurSatRep1 7035246 NO
## [3] -1868 269 145 MurSatRep1 11458273 NO
## [4] 0 2137 1910 MurSatRep1 49004341 NO
## [5] -1409 728 644 MurSatRep1 49004750 NO
## ...
## [2697] 627 723 -1414 MurSatRep1 71818 NO
## [2698] 220 283 -1854 MurSatRep1 325644 NO
## [2699] 620 745 -1392 MurSatRep1 110549 NO
## [2700] 201 266 -1871 MurSatRep1 436259 NO
## [2701] 627 727 -1410 MurSatRep1 347834 NO
## Gene_ID Transcript_ID Position
## <character> <character> <numeric>
## [1] Nan <NA> 5
## [2] Nan <NA> 5
## [3] Nan <NA> 5
## [4] Nan <NA> 5
## [5] Nan <NA> 5
## ...
## [2697] Nan <NA> 5
## [2698] Nan <NA> 5
## [2699] Nan <NA> 5
## [2700] Nan <NA> 5
## [2701] Nan <NA> 5
## -----
## seqinfo: 239 sequences (1 circular) from mm10 genome
##
## $Plot

```

Meta Gene Profile



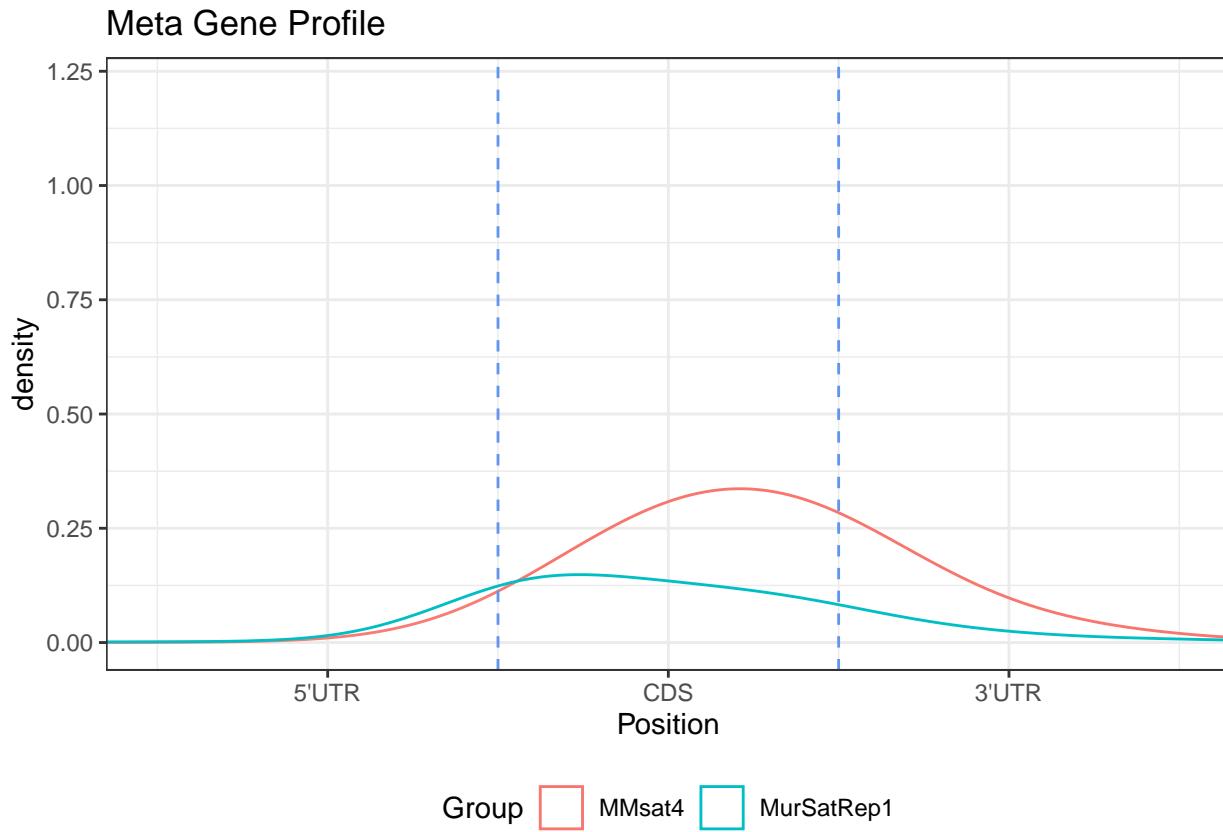
```
metael
```

```
## $Peaks
## GRanges object with 4394 ranges and 17 metadata columns:
##           seqnames      ranges strand |  swScore  milliDiv
##           <Rle>      <IRanges> <Rle> | <integer> <numeric>
## [1]     chr1    49003992-49004114   - |    272    295
## [2]     chr1    49004329-49004605   - |    402    284
## [3]     chr1    82366411-82366554   - |    395    248
## [4]     chr1    86014673-86014899   - |    382    314
## [5]     chr1   117749987-117750291   - |    312    304
## ...
## [4390] chrY_JH584301_random 71770-71865   + |    247    302
## [4391] chr3_KZ289066_fix   325612-325675   + |    385    172
## [4392] chr4_KZ289068_fix   110486-110611   + |    259    344
## [4393] chr16_KB469741_fix  436227-436289   + |    236    254
## [4394] chrY_JH792832_fix  347783-347884   + |    295    290
##     milliDel milliIns genoLeft repName repClass repFamily
##     <numeric> <numeric> <integer> <character> <character> <character>
## [1]      57        8 -146467857 MMSAT4 Satellite  Satellite
## [2]      32       27 -146467366 MMSAT4 Satellite  Satellite
## [3]      42       21 -113105417 MMSAT4 Satellite  Satellite
## [4]       0        0 -109457072 MMSAT4 Satellite  Satellite
## [5]      30       27 -77721680 MMSAT4 Satellite  Satellite
## ...
## [4390]     10        0 -188010 MurSatRep1 Unknown   Unknown
## [4391]     0        0 -13559 MurSatRep1 Unknown   Unknown
```

```

## [4392]      8      8   -421961 MurSatRep1    Unknown Unknown
## [4393]     45      0   -69587 MurSatRep1    Unknown Unknown
## [4394]     10     20  -196305 MurSatRep1    Unknown Unknown
##           repStart   repEnd   repLeft      id center location
##           <integer> <integer> <integer> <character> <integer> <character>
## [1]       -5      163      35 MMsat4 49004054    NO
## [2]      -13      278       1 MMsat4 49004467    NO
## [3]       0      168      22 MMsat4 82366483    NO
## [4]       0      227       1 MMsat4 86014787    NO
## [5]       0      305       1 MMsat4 117750139   NO
## ...
## [4390]     627      723    -1414 MurSatRep1 71818    NO
## [4391]     220      283    -1854 MurSatRep1 325644    NO
## [4392]     620      745    -1392 MurSatRep1 110549    NO
## [4393]     201      266    -1871 MurSatRep1 436259    NO
## [4394]     627      727    -1410 MurSatRep1 347834    NO
##           Gene_ID Transcript_ID Position
##           <character> <character> <numeric>
## [1]       Nan      <NA>      5
## [2]       Nan      <NA>      5
## [3]       Nan      <NA>      5
## [4]       Nan      <NA>      5
## [5]       Nan      <NA>      5
## ...
## [4390]     Nan      <NA>      5
## [4391]     Nan      <NA>      5
## [4392]     Nan      <NA>      5
## [4393]     Nan      <NA>      5
## [4394]     Nan      <NA>      5
##
## -----
## seqinfo: 239 sequences (1 circular) from mm10 genome
##
## $Plot

```



```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/metagene_profile_both.pdf", width=2, height = 2)
metael
```

```
## $Peaks
## GRanges object with 4394 ranges and 17 metadata columns:
##           seqnames      ranges strand |   swScore   milliDiv
##           <Rle>      <IRanges> <Rle> |   <integer> <numeric>
## [1]     chr1    49003992-49004114   - |     272    295
## [2]     chr1    49004329-49004605   - |     402    284
## [3]     chr1    82366411-82366554   - |     395    248
## [4]     chr1    86014673-86014899   - |     382    314
## [5]     chr1  117749987-117750291   - |     312    304
## ...
## ...
## [4390] chrY_JH584301_random    71770-71865   + |     247    302
## [4391] chr3_KZ289066_fix     325612-325675   + |     385    172
## [4392] chr4_KZ289068_fix     110486-110611   + |     259    344
## [4393] chr16_KB469741_fix    436227-436289   + |     236    254
## [4394] chrY_JH792832_fix    347783-347884   + |     295    290
##   milliDel milliIns genoLeft repName repClass repFamily
##   <numeric> <numeric> <integer> <character> <character> <character>
## [1]      57       8 -146467857   MMSAT4 Satellite  Satellite
## [2]      32      27 -146467366   MMSAT4 Satellite  Satellite
## [3]      42      21 -113105417   MMSAT4 Satellite  Satellite
## [4]       0       0 -109457072   MMSAT4 Satellite  Satellite
## [5]      30      27 -77721680   MMSAT4 Satellite  Satellite
## ...
## ...
## [4390]     10       0 -188010 MurSatRep1 Unknown    Unknown
```

```

## [4391]      0      0   -13559 MurSatRep1    Unknown Unknown
## [4392]      8      8   -421961 MurSatRep1    Unknown Unknown
## [4393]     45      0   -69587 MurSatRep1    Unknown Unknown
## [4394]     10     20  -196305 MurSatRep1    Unknown Unknown
##           repStart   repEnd   repLeft      id   center location
## <integer> <integer> <integer> <character> <integer> <character>
## [1]       -5     163      35 MMsat4 49004054      NO
## [2]      -13     278      1 MMsat4 49004467      NO
## [3]       0     168      22 MMsat4 82366483      NO
## [4]       0     227      1 MMsat4 86014787      NO
## [5]       0     305      1 MMsat4 117750139     NO
## ...
## [4390]     ...     ...     ...     ...     ...
## [4391]     220     283   -1854 MurSatRep1 71818      NO
## [4392]     620     745   -1392 MurSatRep1 325644      NO
## [4393]     201     266   -1871 MurSatRep1 436259      NO
## [4394]     627     727   -1410 MurSatRep1 347834      NO
##           Gene_ID Transcript_ID Position
## <character> <character> <numeric>
## [1]      Nan     <NA>      5
## [2]      Nan     <NA>      5
## [3]      Nan     <NA>      5
## [4]      Nan     <NA>      5
## [5]      Nan     <NA>      5
## ...
## [4390]     ...     ...
## [4391]     Nan     <NA>      5
## [4392]     Nan     <NA>      5
## [4393]     Nan     <NA>      5
## [4394]     Nan     <NA>      5
##
## -----
## seqinfo: 239 sequences (1 circular) from mm10 genome
##
## $Plot
dev.off()

## pdf
## 2

```

venn diagram

```

vset <- list(mmsat4frame[!duplicated(mmsat4frame$gene_name), "gene_name"], MurSatRep1frame[!duplicated(MurSatRep1frame$gene_name)], MurSatRep1frame[!duplicated(MurSatRep1frame$gene_name)])
names(vset) <- c("MMSAT4", "MurSatRep1")

vplot <- plot(euler(vset, shape = "ellipse"), quantities = TRUE, main="Genes by contained motifs" , fill=TRUE)

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/Venn_motif.pdf", width=2, height = 2)
vplot
dev.off()

## pdf
## 2

```

Sequence overlap mmsat4 mursatrep1

Just check to make sure these are not the same

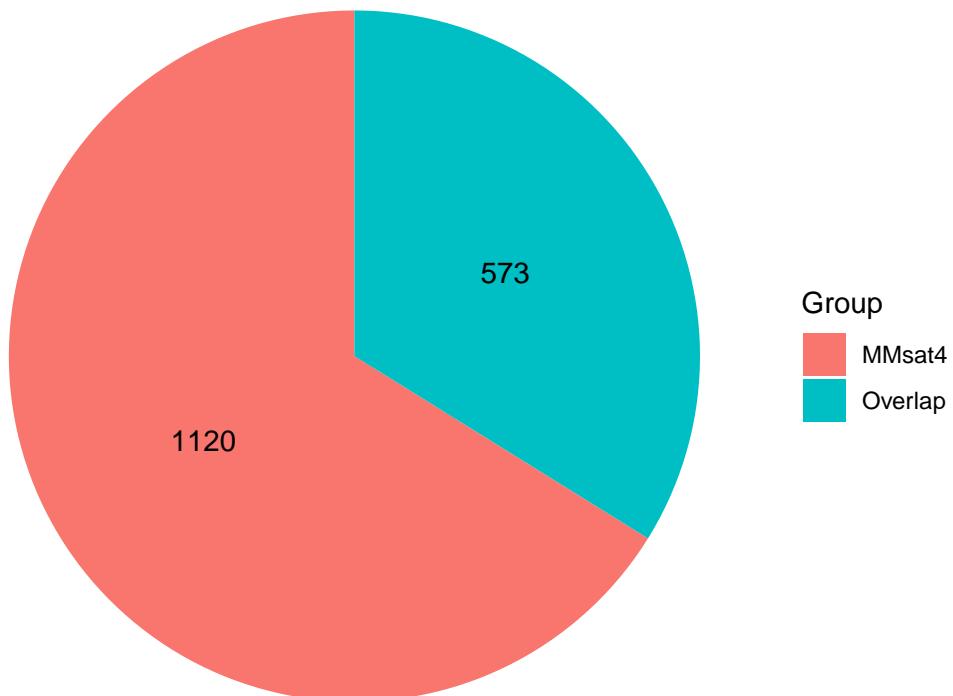
```
olmmsat4 <- subsetByOverlaps(MMSAT4, MurSatRep1)
olmursatrep1 <- subsetByOverlaps(MurSatRep1, MMSAT4)

#make pie charts
pimmsat <- data.frame(c("MMSat4", "Overlap"),
                        c(length(MMSAT4)-length(olmmsat4), length(olmmsat4)))
colnames(pimmsat) <- c("Group", "Count")

pimursatrep1 <- data.frame(c("MurSatRep1", "Overlap"),
                            c(length(MurSatRep1)-length(olmursatrep1), length(olmursatrep1)))
colnames(pimursatrep1) <- c("Group", "Count")

pimmsatP <- ggplot(pimmsat, aes(x="", y=Count, fill=Group)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  geom_text(aes(label = Count),
            position = position_stack(vjust = 0.5))

pimmsatP
```

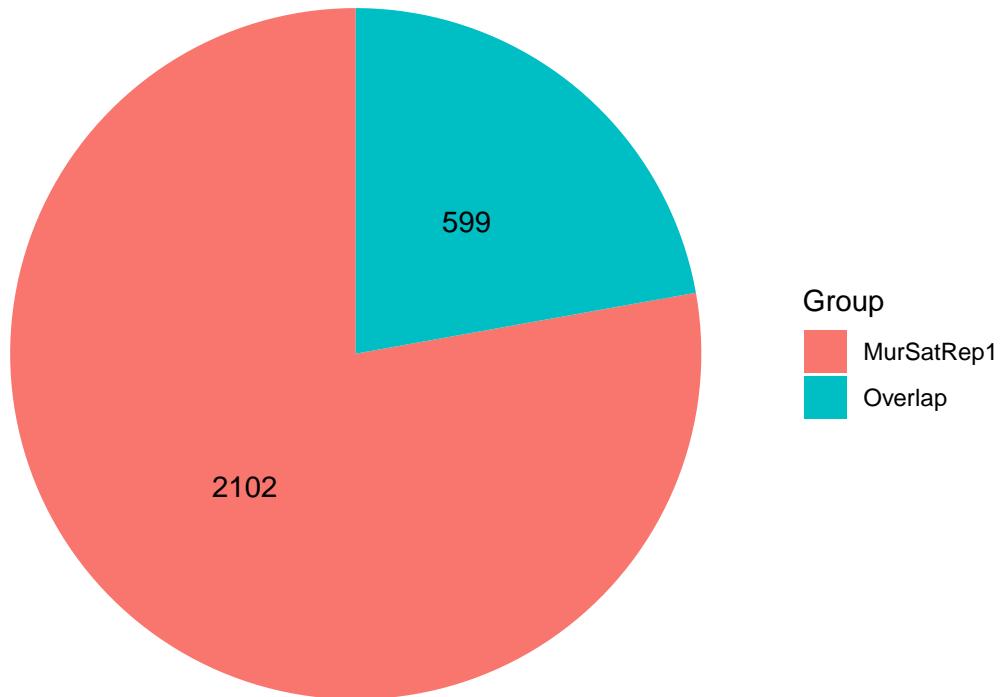


```

pimursatrep1P <- ggplot(pimursatrep1, aes(x="", y=Count, fill=Group)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  geom_text(aes(label = Count),
            position = position_stack(vjust = 0.5))

pimursatrep1P

```



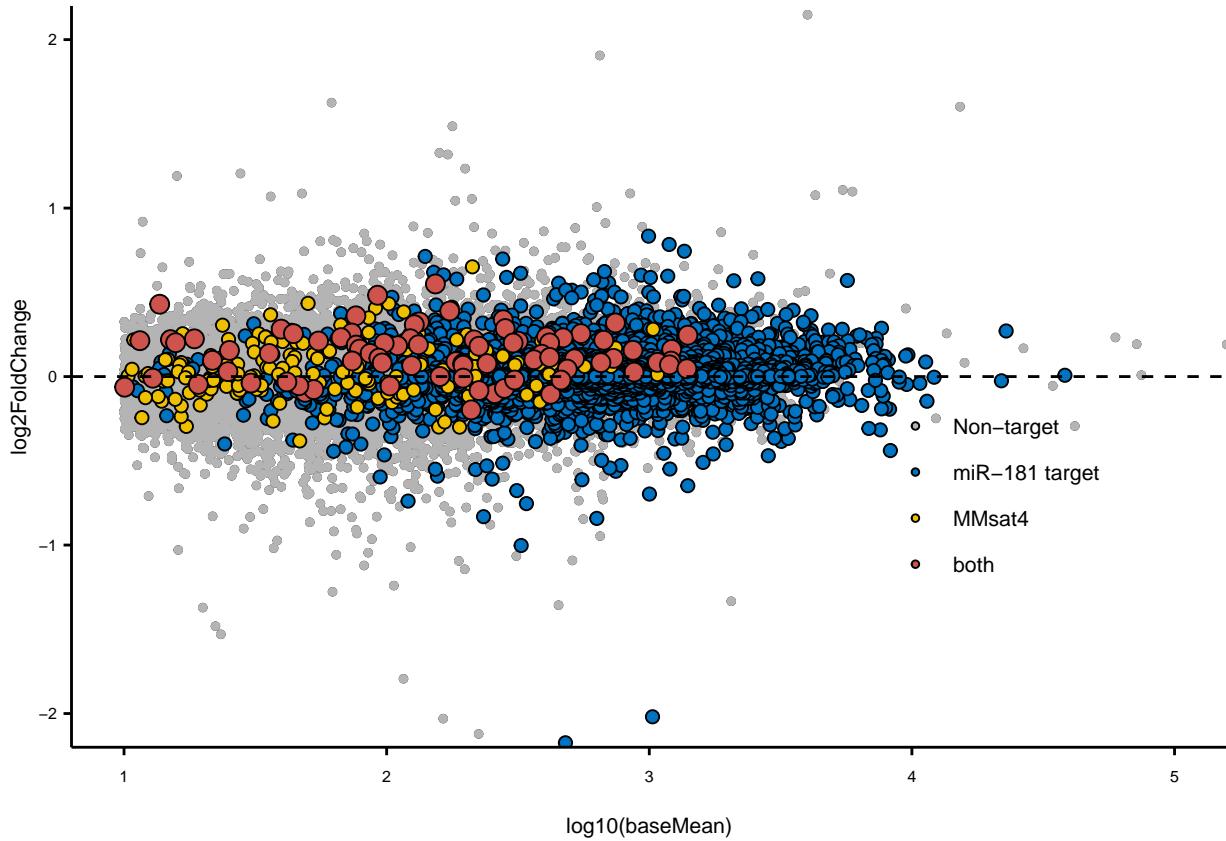
MAplots of targets with satellites

```

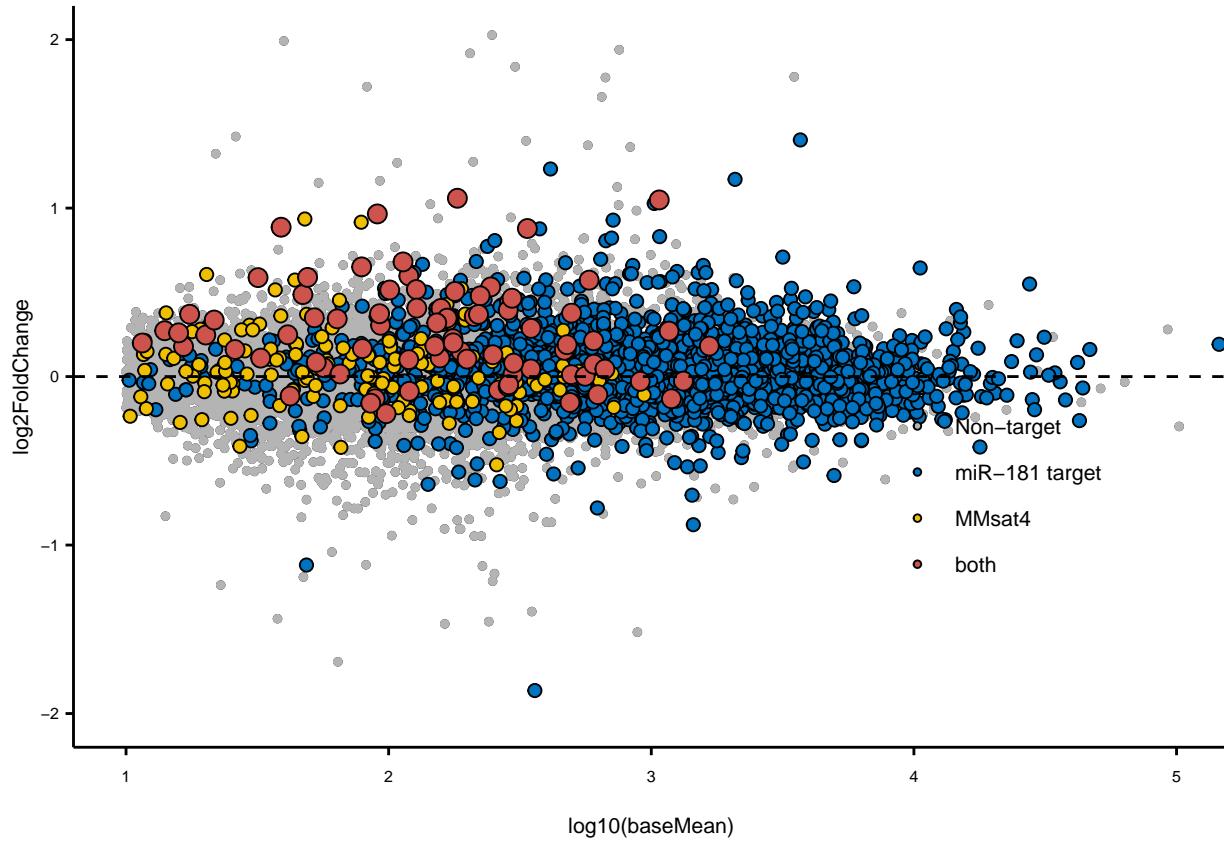
#mmsat4 RNA
MAmmsat4RNA <- ggplot(RNA, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(tvsmmsat4, levels = c("Non-target", "miR-181 target", "MMsat4", "both")))) +
  geom_point(shape=21, size=1) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe2, farbe3)) +
  geom_point(data = RNA[RNA$tvsmmsat4=="Non-target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="black") +
  geom_point(data = RNA[RNA$tvsmmsat4=="miR-181 target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="red") +
  geom_point(data = RNA[RNA$tvsmmsat4=="MMsat4",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="blue") +
  geom_point(data = RNA[RNA$tvsmmsat4=="both",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="green") +
  geom_hline(yintercept = 0, linetype="dashed", colour="black")+
  theme_paper() +
  coord_cartesian(ylim = c(-2,2), xlim = c(1,5))

MAmmsat4RNA

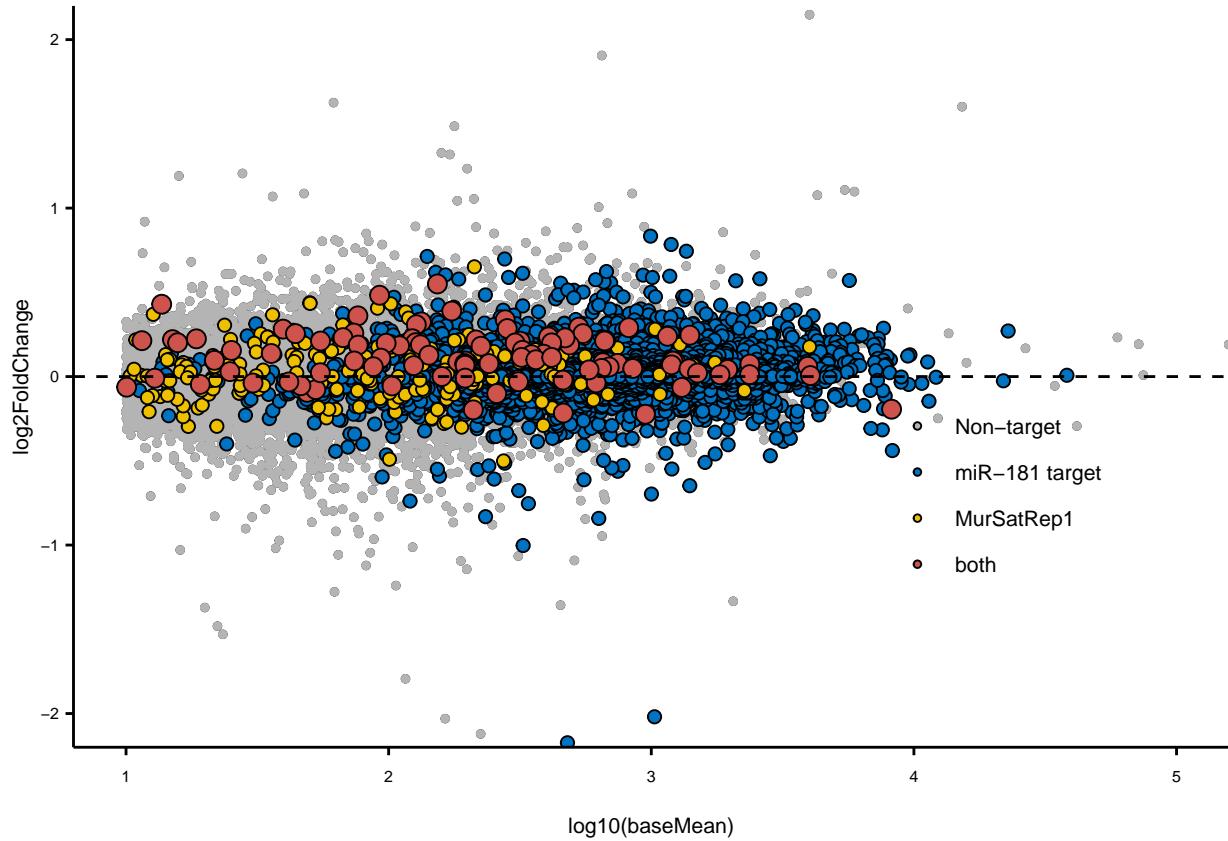
```



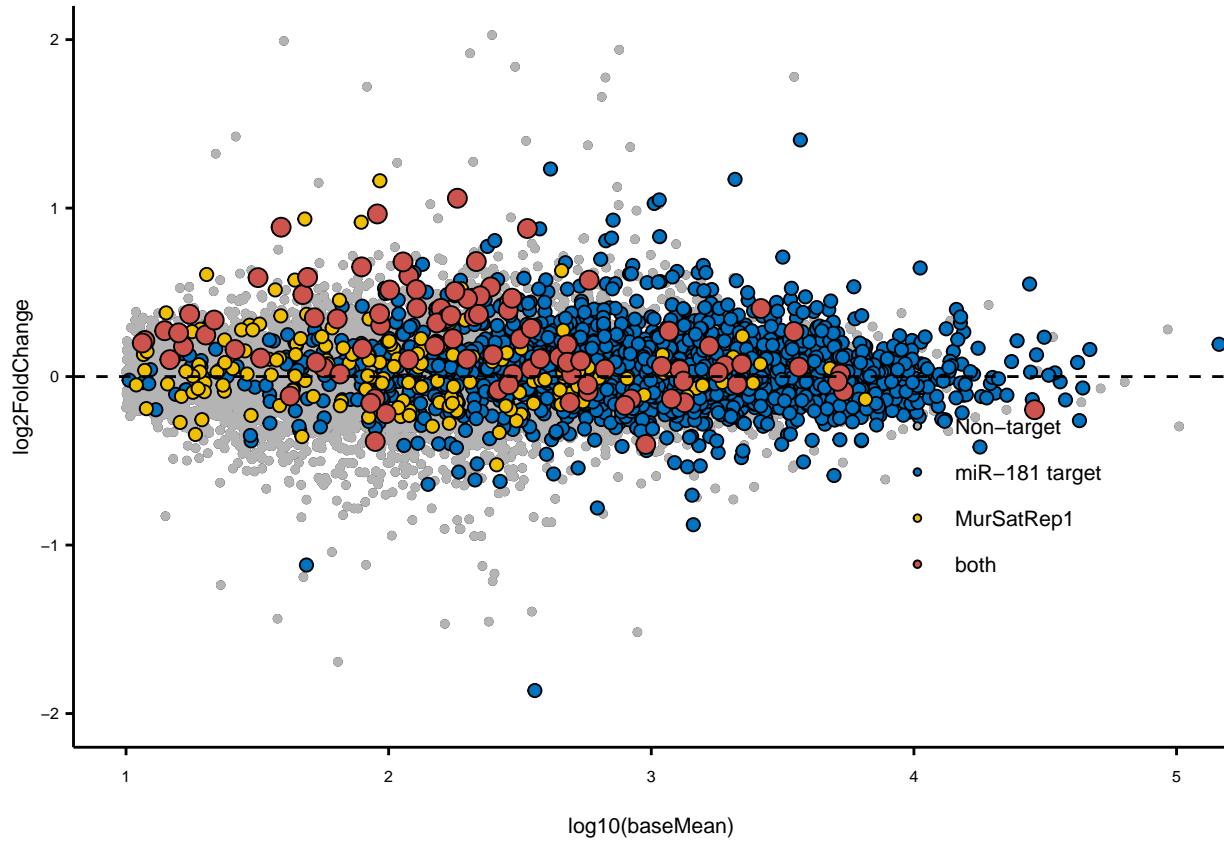
```
##mmsat4 RPF
MAmmsat4RPF <- ggplot(RPF, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(tvsmmsat4, levels = c("Non-target", "miR-181 target", "MMsat4", "both")))) +
  geom_point(shape=21, size=1) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe2, farbe3)) +
  geom_point(data = RPF[tvsmmsat4=="Non-target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="black") +
  geom_point(data = RPF[tvsmmsat4=="miR-181 target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="blue") +
  geom_point(data = RPF[tvsmmsat4=="MMsat4"], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="yellow") +
  geom_point(data = RPF[tvsmmsat4=="both"], aes(x=log10(baseMean), y=log2FoldChange), shape=21, colour="red") +
  geom_hline(yintercept = 0, linetype="dashed", colour="black") +
  theme_paper() +
  coord_cartesian(ylim = c(-2,2), xlim = c(1,5))
MAmmsat4RPF
```



```
#Mursatrep1 RNA
MAmursatrep1RNA <- ggplot(RNA, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(tvsMurSatRep1, levels=c("Non-target", "miR-181 target", "MMSat4", "both")))) +
  geom_point(shape=21, size=1) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe2, farbe3)) +
  geom_point(data = RNA[RNA$tvsmurSatRep1=="Non-target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RNA[RNA$tvsmurSatRep1=="miR-181 target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RNA[RNA$tvsmurSatRep1=="MMSat4",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RNA[RNA$tvsmurSatRep1=="both",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_hline(yintercept = 0, linetype="dashed", colour="black") +
  theme_paper() +
  coord_cartesian(ylim = c(-2,2), xlim = c(1,5))
MAmursatrep1RNA
```



```
#Mursatrep1 RPF
MAmursatrep1RPF <- ggplot(RPF, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(tvsMurSatRep1, levels=c("Non-target", "miR-181 target", "MurSatRep1", "both")))) +
  geom_point(shape=21, size=1) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe2, farbe3)) +
  geom_point(data = RPF[RPF$tvsmurSatRep1=="Non-target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RPF[RPF$tvsmurSatRep1=="miR-181 target",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RPF[RPF$tvsmurSatRep1=="MurSatRep1",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_point(data = RPF[RPF$tvsmurSatRep1=="both",], aes(x=log10(baseMean), y=log2FoldChange), shape=21) +
  geom_hline(yintercept = 0, linetype="dashed", colour="black") +
  theme_paper() +
  coord_cartesian(ylim = c(-2,2), xlim = c(1,5))
MAmursatrep1RPF
```



Session info

```
sessionInfo()

## R version 4.2.3 (2023-03-15 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8  LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## attached base packages:
## [1] stats4      stats       graphics    grDevices   utils       datasets   methods
## [8] base
##
## other attached packages:
## [1] eulerr_7.0.0        cowplot_1.1.1        cliProfiler_1.4.0
## [4] Rtsne_0.16          tibble_3.2.1         dplyr_1.1.2
## [7] rtracklayer_1.58.0  GenomicRanges_1.50.2  GenomeInfoDb_1.34.9
## [10] IRanges_2.32.0      S4Vectors_0.36.2     BiocGenerics_0.44.0
## [13] ggplot2_3.4.2
```

```

##
## loaded via a namespace (and not attached):
## [1] MatrixGenerics_1.10.0           Biobase_2.58.0
## [3] tidyR_1.3.0                     carData_3.0-5
## [5] highr_0.10                      BSgenome_1.66.3
## [7] GenomeInfoDbData_1.2.9         Rsamtools_2.14.0
## [9] yaml_2.3.7                       pillar_1.9.0
## [11] backports_1.4.1                 lattice_0.20-45
## [13] glue_1.6.2                      digest_0.6.31
## [15] ggsignif_0.6.4                  XVector_0.38.0
## [17] polyclip_1.10-4                 colorspace_2.1-0
## [19] htmltools_0.5.4                 Matrix_1.5-3
## [21] XML_3.99-0.14                  pkgconfig_2.0.3
## [23] broom_1.0.4                     zlibbioc_1.44.0
## [25] purrr_1.0.1                     scales_1.2.1
## [27] BiocParallel_1.32.6            generics_0.1.3
## [29] farver_2.1.1                   car_3.1-2
## [31] ggpubr_0.6.0                   withr_2.5.0
## [33] SummarizedExperiment_1.28.0   cli_3.6.0
## [35] magrittr_2.0.3                 crayon_1.5.2
## [37] evaluate_0.21                  fansi_1.0.4
## [39] rstatix_0.7.2                 tools_4.2.3
## [41] BiocIO_1.8.0                   lifecycle_1.0.3
## [43] matrixStats_0.63.0            munsell_0.5.0
## [45] DelayedArray_0.23.2          Biostrings_2.66.0
## [47] compiler_4.2.3                rlang_1.1.0
## [49] grid_4.2.3                     RCurl_1.98-1.12
## [51] rstudioapi_0.14               rjson_0.2.21
## [53] bitops_1.0-7                  labeling_0.4.2
## [55] rmarkdown_2.22                 restfulr_0.0.15
## [57] gtable_0.3.3                  codetools_0.2-19
## [59] abind_1.4-5                   R6_2.5.1
## [61] GenomicAlignments_1.34.1     knitr_1.43
## [63] fastmap_1.1.1                 utf8_1.2.3
## [65] polylabelr_0.2.0              parallel_4.2.3
## [67] Rcpp_1.0.10                    vctrs_0.6.2
## [69] tidyselect_1.2.0              xfun_0.39

```