Fig4 ECDF plots

Nikita Verheyden

2023-05-16

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
```

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.g

#targets
larget <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure3/mir181_bs_with_seeds.rds")
largetframe <- as.data.frame(larget)

#targets with introns and other
tject <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure1/mir181_binding_sites__venn_types/mir
names(tject) <- 1:length(tject$geneName)
tframe <- as.data.frame(tject)
head(tframe)
```

```
##   seqnames   start     end width strand scoreSum scoreMean scoreMax
```

```
## 1     chr1 6245651 6245657     7      +  9.52553  4.762765  6.00678
## 2     chr1 6248341 6248347     7      + 92.68921 23.172303 48.76900
## 3     chr1 6248857 6248863     7      + 14.07133  7.035665  7.04425
## 4     chr1 6248918 6248924     7      + 38.91451 12.971503 20.65080
## 5     chr1 7170481 7170487     7      + 66.92218 13.384436 25.84490
## 6     chr1 9899605 9899611     7      + 25.15963  6.289907  8.61019
##         geneType geneName            geneID region BS_ID          mir_IP
## 1 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     5 mmu-miR-181a-5p
## 2 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     8 mmu-miR-181a-5p
## 3 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    10 mmu-miR-181a-5p
## 4 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    11 mmu-miR-181a-5p
## 5 protein_coding   Pcmtd1 ENSMUSG00000051285   utr3    19 mmu-miR-181a-5p
## 6 protein_coding     Sgk3 ENSMUSG00000025915   utr3    23 mmu-miR-181a-5p
##   n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d              set WT KO
## 1        1         1         0         0         0 ago_bs_mir181_chi  1  1
## 2        5         5         0         0         0 ago_bs_mir181_chi  1  1
## 3        6         6         0         0         0 ago_bs_mir181_chi  1  0
## 4        6         6         0         0         0 ago_bs_mir181_chi  1  1
## 5        4         4         0         0         0 ago_bs_mir181_chi  1  1
## 6        1         1         0         0         0 ago_bs_mir181_chi NA NA
##            geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 1 ENSMUSG00000025907     Rb1cc1      cds              4              3
## 2 ENSMUSG00000025907     Rb1cc1      cds             28             32
## 3 ENSMUSG00000025907     Rb1cc1      cds             13             11
## 4 ENSMUSG00000025907     Rb1cc1      cds             15             15
## 5 ENSMUSG00000051285     Pcmtd1     utr3             12             22
## 6               <NA>       <NA>     <NA>             NA             NA
##   counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 1              3              3             10              3
## 2             27             46             41             20
## 3              4             22             13             12
## 4             10             33             20             18
## 5             14             16             20              9
## 6             NA             NA             NA             NA
##            geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 1 ENSMUSG00000025907           1609           1973           1250
## 2 ENSMUSG00000025907           1609           1973           1250
## 3 ENSMUSG00000025907           1609           1973           1250
## 4 ENSMUSG00000025907           1609           1973           1250
## 5 ENSMUSG00000051285           1355           1706           1064
## 6               <NA>             NA             NA             NA
##   counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 1           2638           2231           1352       92.10645
## 2           2638           2231           1352      281.53271
## 3           2638           2231           1352      145.51107
## 4           2638           2231           1352      186.74162
## 5           1654           1348            755      151.36245
## 6             NA             NA             NA             NA
##   resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue resBs.padj
## 1           -0.1093039   0.5923673 0.03419066    0.8533018  0.9652601
## 2            0.2749428   0.2351157 1.35874137    0.2437557  0.6729889
## 3           -0.1805519   0.3623758 0.25017050    0.6169550  0.8961239
## 4           -0.2606282   0.3062717 0.73169661    0.3923338  0.7868678
## 5            0.1466485   0.3122905 0.22052922    0.6386370  0.9013566
```

```
## 6                   NA          NA         NA           NA         NA
##   resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 1             NA                   NA          NA         NA           NA
## 2             NA                   NA          NA         NA           NA
## 3             NA                   NA          NA         NA           NA
## 4             NA                   NA          NA         NA           NA
## 5             NA                   NA          NA         NA           NA
## 6             NA                   NA          NA         NA           NA
##   resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 1         NA           133.7259           117.9980           129.8669
## 2         NA           133.7259           117.9980           129.8669
## 3         NA           133.7259           117.9980           129.8669
## 4         NA           133.7259           117.9980           129.8669
## 5         NA           248.6210           225.2505           244.0445
## 6         NA                 NA                 NA                 NA
##   tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1           139.8635           146.2855           163.5360
## 2           139.8635           146.2855           163.5360
## 3           139.8635           146.2855           163.5360
## 4           139.8635           146.2855           163.5360
## 5           193.5994           195.1330           201.6149
## 6                 NA                 NA                 NA
##                   BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported  down
## 1  ENSMUSG00000025907.bs5              3              3          TRUE FALSE
## 2  ENSMUSG00000025907.bs8              3              3          TRUE FALSE
## 3 ENSMUSG00000025907.bs10              3              3          TRUE FALSE
## 4 ENSMUSG00000025907.bs11              3              3          TRUE FALSE
## 5  ENSMUSG00000051285.bs4              3              3          TRUE FALSE
## 6                   <NA>             NA             NA            NA    NA
```

```
#MMsat4
repeat_masker <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure2/MMsat4/repeat_masker.rds")
MMSAT4 <- repeat_masker[repeat_masker$repName == "MMSAT4"]
```

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 <- "#FF6F00FF"
farbe12 <- "#C71000FF"
farbe13 <- "#008EA0FF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"
```

```
RNApcol <- "#b56504"
RNAncol <- "#027d73"
RPFpcol <- "#c4c404"
RPFncol <- "#8d0391"
```

inspect targetdata

We're keeping all of those targets for now but will analyze the in ecdf plots

```
table(largetframe$set)
```

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

```
colnames(largetframe)
```

```
##  [1] "seqnames"                          "start"
##  [3] "end"                               "width"
##  [5] "strand"                            "scoreSum"
##  [7] "scoreMean"                         "scoreMax"
##  [9] "geneType"                          "geneName"
## [11] "geneID"                            "region"
## [13] "mir_IP"                            "n_mir181"
## [15] "n_mir181a"                         "n_mir181b"
## [17] "n_mir181c"                         "n_mir181d"
## [19] "set"                               "mir181BS_ID"
## [21] "WT"                                "KO"
## [23] "geneID.2"                          "geneName.1"
## [25] "region.1"                          "counts.bs.1_KO"
## [27] "counts.bs.2_KO"                    "counts.bs.3_KO"
## [29] "counts.bs.4_WT"                    "counts.bs.5_WT"
## [31] "counts.bs.6_WT"                    "geneID.1"
## [33] "counts.bg.1_KO"                    "counts.bg.2_KO"
## [35] "counts.bg.3_KO"                    "counts.bg.4_WT"
## [37] "counts.bg.5_WT"                    "counts.bg.6_WT"
## [39] "resBs.baseMean"                    "resBs.log2FoldChange"
## [41] "resBs.lfcSE"                       "resBs.stat"
## [43] "resBs.pvalue"                      "resBs.padj"
## [45] "resBg.baseMean"                    "resBg.log2FoldChange"
## [47] "resBg.lfcSE"                       "resBg.stat"
## [49] "resBg.pvalue"                      "resBg.padj"
## [51] "tpm.counts.bg.1_KO"                "tpm.counts.bg.2_KO"
## [53] "tpm.counts.bg.3_KO"                "tpm.counts.bg.4_WT"
## [55] "tpm.counts.bg.5_WT"                "tpm.counts.bg.6_WT"
## [57] "BS_ID"                             "tpm_support_KO"
## [59] "tpm_support_WT"                    "tpm_supported"
## [61] "down"                              "all_seeds_200down"
## [63] "first_seed_200down.start"          "first_seed_200down.end"
## [65] "first_seed_200down.width"          "first_seed_200down.type"
## [67] "first_seed_200down.wobble"         "seed_repetitions.200down"
## [69] "seed_repetitions.200down.wobble"   "all_seeds_200up"
```

```
## [71] "first_seed_200up.start"          "first_seed_200up.end"
## [73] "first_seed_200up.width"          "first_seed_200up.type"
## [75] "first_seed_200up.wobble"         "seed_repetitions.200up"
## [77] "seed_repetitions.200up.wobble"
```

ECDF plots

each code chunk is a split of the main target table that is then used for a specific ecdf plot

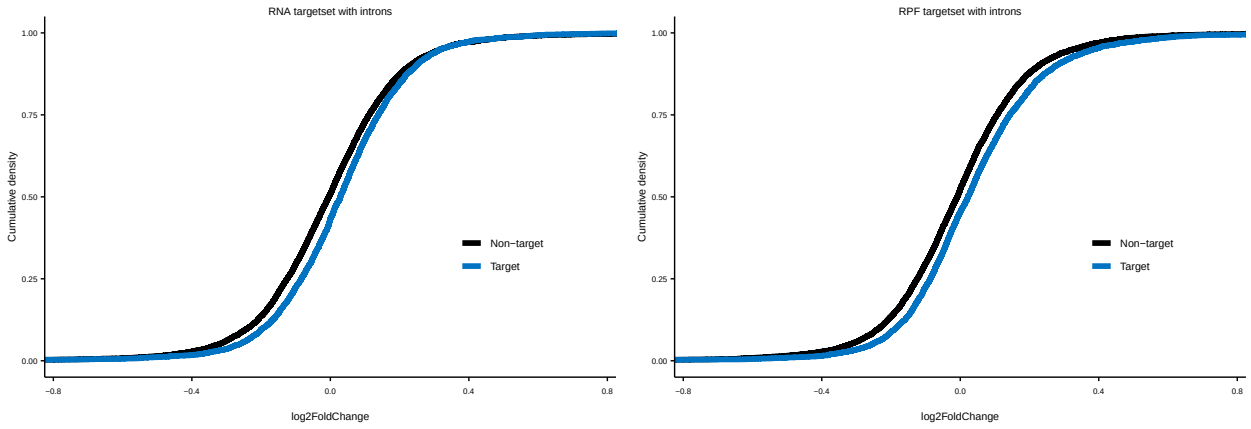
targetfiles

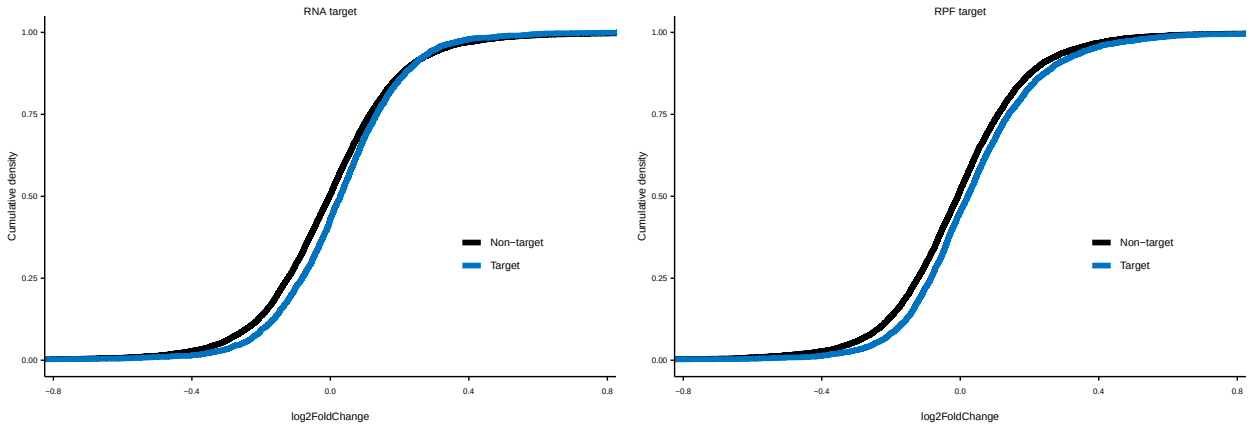
```
#old target with introns
#RNA
RNA$oldtarg <- "Non-target"
RNA$oldtarg[RNA$gene_symbol %in% tframe$geneName] <- "Target"
#RPF
RPF$oldtarg <- "Non-target"
RPF$oldtarg[RPF$gene_symbol %in% tframe$geneName] <- "Target"

#giant frame
#RNA
RNA$target <- "Non-target"
RNA$target[RNA$gene_symbol %in% largetframe$geneName] <- "Target"
#RPF
RPF$target <- "Non-target"
RPF$target[RPF$gene_symbol %in% largetframe$geneName] <- "Target"


#ECDF
#old targets with introns
#RNA
targetoldECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(oldtarg, levels = c("Non-
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA targetset with introns")
```

```
## 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.
```

```
targetoldECDFRNA

#RPF
targetoldECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(oldtarg, levels = c("Non-
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
```

```
  ggtitle("RPF targetset with introns")

targetoldECDFRPF

#targets
#RNA
targetECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(target, levels = c("Non-targ
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA target")

targetECDFRNA

#RPF
targetECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(target, levels = c("Non-targ
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF target")

targetECDFRPF
```





datasets

```
#RNA
RNA$targetset <- "Non-target"
RNA$targetset[RNA$gene_symbol %in% largetframe[largetframe$set == "ago_bs_mir181_chi", "geneName"]] <-
RNA$targetset[RNA$gene_symbol %in% largetframe[largetframe$set == "mir181_enriched", "geneName"]] <- "mi
RNA$targetset[RNA$gene_symbol %in% largetframe[largetframe$set == "ago_bs_mir181_chi&mir181_enriched",

table(RNA$targetset)
```

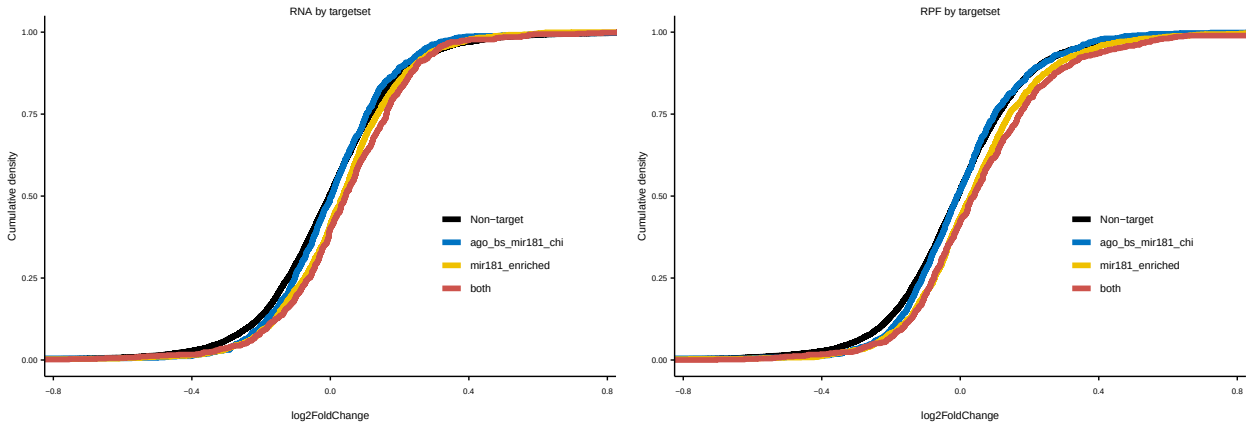
```
##
## ago_bs_mir181_chi              both   mir181_enriched        Non-target
##               783               667              1521             10330
```

```
#RPF
RPF$targetset <- "Non-target"
RPF$targetset[RPF$gene_symbol %in% largetframe[largetframe$set == "ago_bs_mir181_chi", "geneName"]] <-
RPF$targetset[RPF$gene_symbol %in% largetframe[largetframe$set == "mir181_enriched", "geneName"]] <- "mi
RPF$targetset[RPF$gene_symbol %in% largetframe[largetframe$set == "ago_bs_mir181_chi&mir181_enriched",

table(RPF$targetset)
```

```
##
## ago_bs_mir181_chi              both   mir181_enriched        Non-target
##               782               667              1508              8412
```

```
# ecdf plots
#RNA
setECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(targetset, levels = c("Non-targ
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA by targetset")

setECDFRNA

#RPF
setECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(targetset, levels = c("Non-targ
```

```
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF by targetset")

setECDFRPF
```



region (single targets)

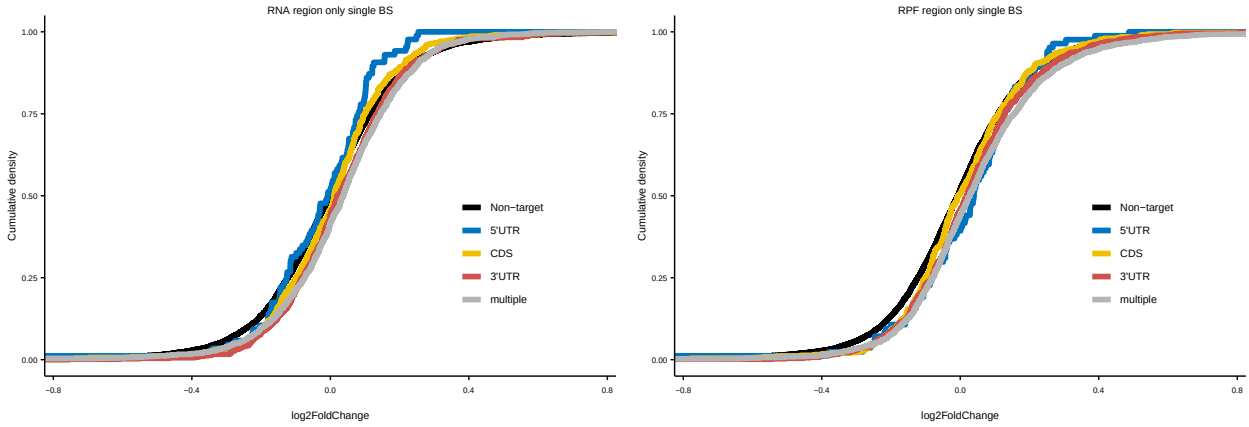
```
#get number of binding sites per gene to be able to sort for singles
bsnum <- as.data.frame(table(largetframe$geneName))
colnames(bsnum) <- c("geneName", "BS_number")

#RNA
RNA$region_single <- "Non-target"
RNA$region_single[RNA$gene_symbol %in% largetframe[largetframe$region == "utr5", "geneName"]] <- "5'UTR
RNA$region_single[RNA$gene_symbol %in% largetframe[largetframe$region == "cds", "geneName"]] <- "CDS"
RNA$region_single[RNA$gene_symbol %in% largetframe[largetframe$region == "utr3", "geneName"]] <- "3'UTR
RNA$region_single[RNA$gene_symbol %in% bsnum[bsnum$BS_number > 1, "geneName"]] <- "multiple"

table(RNA$region_single)
```

```
##
##      3'UTR      5'UTR        CDS   multiple Non-target
##        659         86        451       1775      10330
```

```
#RPF
RPF$region_single <- "Non-target"
RPF$region_single[RPF$gene_symbol %in% largetframe[largetframe$region == "utr5", "geneName"]] <- "5'UTR
RPF$region_single[RPF$gene_symbol %in% largetframe[largetframe$region == "cds", "geneName"]] <- "CDS"
RPF$region_single[RPF$gene_symbol %in% largetframe[largetframe$region == "utr3", "geneName"]] <- "3'UTR
RPF$region_single[RPF$gene_symbol %in% bsnum[bsnum$BS_number > 1, "geneName"]] <- "multiple"

table(RPF$region_single)
```

```
##
##      3'UTR      5'UTR        CDS   multiple Non-target
##        656         84        450       1767       8412
```

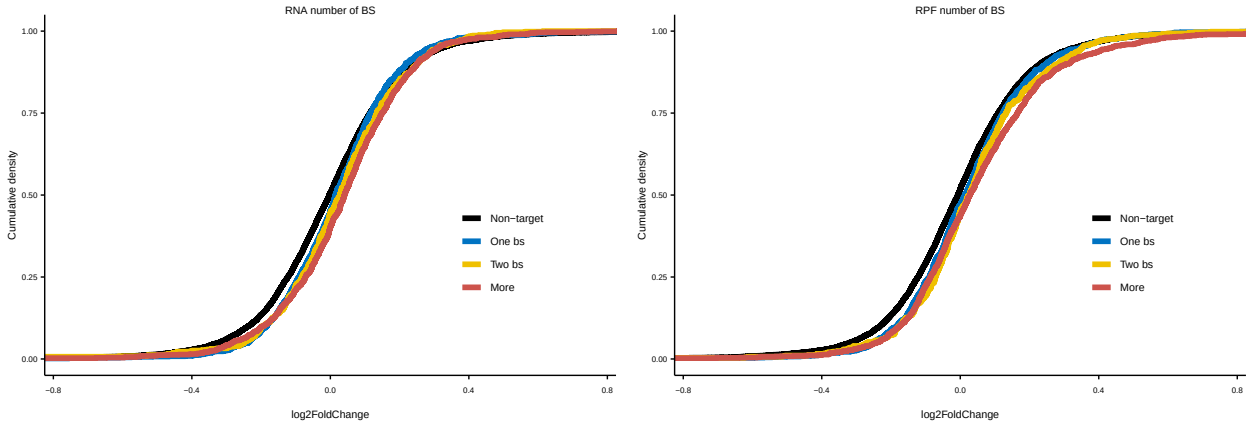
```
# ECDF plots
#RNA
regsingECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(region_single, levels = c("
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3, farbeneg)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA region only single BS")

regsingECDFRNA

#RPF
regsingECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(region_single, levels = c("
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3, farbeneg)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF region only single BS")

regsingECDFRPF
```



number of target sites

```
colnames(bsnum) <- c("gene_symbol", "BS_number")

#RNA
RNAnum <- left_join(RNA, bsnum, by="gene_symbol")
RNAnum$BS_number[is.na(RNAnum$BS_number)] <- "Non-target"
RNAnum$BS_num_plot <- ifelse(RNAnum$BS_number == "Non-target", "Non-target",
                             ifelse(RNAnum$BS_number == 1, "One bs",
                                    ifelse(RNAnum$BS_number == 2, "Two bs", "More")))


#RPF
RPFnum <- left_join(RPF, bsnum, by="gene_symbol")
RPFnum$BS_number[is.na(RPFnum$BS_number)] <- "Non-target"
RPFnum$BS_num_plot <- ifelse(RPFnum$BS_number == "Non-target", "Non-target",
```

```
                              ifelse(RPFnum$BS_number == 1, "One bs",
                                     ifelse(RPFnum$BS_number == 2, "Two bs", "More")))

#ecdf plots
#RNA
numECDFRNA <- ggplot(RNAnum, aes(as.numeric(log2FoldChange), colour=factor(BS_num_plot, levels = c("Non
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA number of BS")

numECDFRNA

#RPF
numECDFRPF <- ggplot(RPFnum, aes(as.numeric(log2FoldChange), colour=factor(BS_num_plot, levels = c("Non
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe2, farbe3)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF number of BS")

numECDFRPF
```



MMsat4

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

#RNA
RNA$tvsmmsat4 <- "Non-target"
RNA$tvsmmsat4[RNA$gene_symbol %in% mmsat4frame$gene_name] <- "MMsat4"
RNA$tvsmmsat4[RNA$gene_symbol %in% tframe$geneName] <- "miR-181 target"
RNA$tvsmmsat4[RNA$gene_symbol %in% tframe$geneName & RNA$gene_symbol %in% mmsat4frame$gene_name] <- "bo
table(RNA$tvsmmsat4)
```

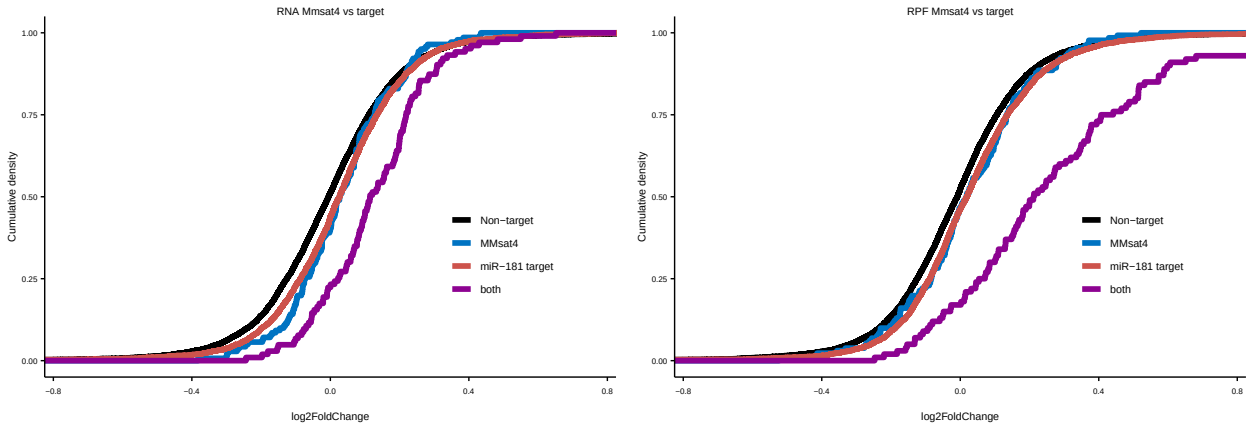
```
##
##           both miR-181 target         MMsat4     Non-target
```

```
##            103           3441            141           9616
```

```
#RPF
RPF$tvsmmsat4 <- "Non-target"
RPF$tvsmmsat4[RPF$gene_symbol %in% mmsat4frame$gene_name] <- "MMsat4"
RPF$tvsmmsat4[RPF$gene_symbol %in% tframe$geneName] <- "miR-181 target"
RPF$tvsmmsat4[RPF$gene_symbol %in% tframe$geneName & RPF$gene_symbol %in% mmsat4frame$gene_name] <- "bo
table(RPF$tvsmmsat4)
```

```
##
##           both miR-181 target         MMsat4     Non-target
##            100           3405            131           7733
```

```
#RNA
tolECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(tvsmmsat4, levels = c("Non-targ
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA Mmsat4 vs target")

tolECDFRNA


#RPF
tolECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(tvsmmsat4, levels = c("Non-targ
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF Mmsat4 vs target")

tolECDFRPF
```



Mmsat4 in 3UTR and CDS

Here we took all targets (genes) that also contain a MMsat4 element and split them by 3'UTR or UTR. The location of the MMsat4 within the gene is not considered in this analysis.

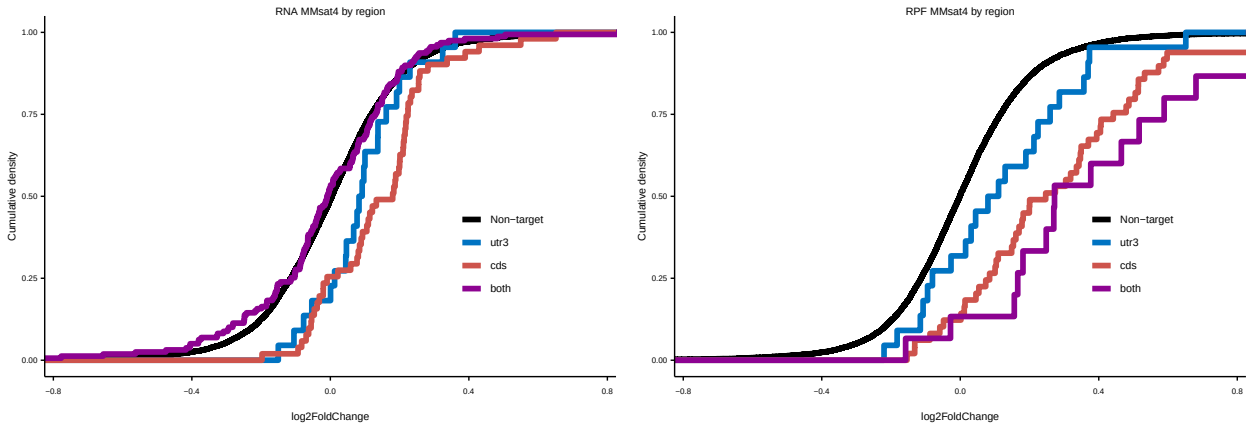
```
rframe <- tframe[tframe$geneName %in% mmsat4frame$gene_name,]

#RNA
RNA$regMMsat4 <- "Non-target"
RNA$regMMsat4[RNA$gene_symbol %in% rframe[rframe$region == "cds","geneName"]] <- "cds"
RNA$regMMsat4[RNA$gene_symbol %in% rframe[rframe$region == "utr3","geneName"]] <- "utr3"
RNA$regMMsat4[RNA$gene_symbol %in% rframe[rframe$region == "utr3","geneName"] &
                RNA$gene_symbol %in% rframe[rframe$region == "cds","geneName"]] <- "both"

table(RNA$regMMsat4)
```

```
##
##       both        cds Non-target       utr3
##        159         51      13069         22
```

```
#RPF
RPF$regMMsat4 <- "Non-target"
RPF$regMMsat4[RPF$gene_symbol %in% rframe[rframe$region == "cds","geneName"]] <- "cds"
RPF$regMMsat4[RPF$gene_symbol %in% rframe[rframe$region == "utr3","geneName"]] <- "utr3"
RPF$regMMsat4[RPF$gene_symbol %in% rframe[rframe$region == "utr3","geneName"] &
                RPF$gene_symbol %in% rframe[rframe$region == "cds","geneName"]] <- "both"

table(RPF$regMMsat4)
```

```
##
##       both        cds Non-target       utr3
##         15         49      11283         22
```

```
#ecdf plots
#RNA
posECDFRNA <- ggplot(RNA, aes(as.numeric(log2FoldChange), colour=factor(regMMsat4, levels = c("Non-targe
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RNA MMsat4 by region")

posECDFRNA

#RPF
posECDFRPF <- ggplot(RPF, aes(as.numeric(log2FoldChange), colour=factor(regMMsat4, levels = c("Non-targe
  stat_ecdf(geom="step", linewidth=2) +
  scale_colour_manual(values = c("black", farbe1, farbe3, RPFncol)) +
  coord_cartesian(xlim = c(-0.75, 0.75)) +
  theme_paper() +
  scale_y_continuous("Cumulative density") + scale_x_continuous("log2FoldChange") +
  ggtitle("RPF MMsat4 by region")

posECDFRPF
```



Export

"D:/Krueger_Lab/Publications/miR181_paper/Figure4"

```
#target

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/targetoldintronECDFRNA.pdf", width=2, height = 2)
targetoldECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/targetoldintronECDFRPF.pdf", width=2, height = 2)
targetoldECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/targetECDF_RNA.pdf", width=2, height = 2)
targetECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/targetECDF_RPF.pdf", width=2, height = 2)
targetECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
#set

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/setECDF_RNA.pdf", width=2, height = 2)
setECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/setECDF_RPF.pdf", width=2, height = 2)
setECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
#region single targets

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/regsingECDF_RNA.pdf", width=2, height = 2)
regsingECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/regsingECDF_RPF.pdf", width=2, height = 2)
regsingECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
#number

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/numECDF_RNA.pdf", width=2, height = 2)
numECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/numECDF_RPF.pdf", width=2, height = 2)
numECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
# MMsat4 vs target

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/MMsat4vsTargetECDF_RNA.pdf", width=2, height = 2)
tolECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/MMsat4vsTargetECDF_RPF.pdf", width=2, height = 2)
tolECDFRPF
dev.off()
```

```
## pdf
##   2
```

```
# region with MMsat4 cds and 3'utr

pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/MMsat4byRegionECDF_RNA.pdf", width=2, height = 2)
posECDFRNA
dev.off()
```

```
## pdf
##   2
```

```
pdf("D:/Krueger_Lab/Publications/miR181_paper/Figure4/MMsat4byRegionECDF_RPF.pdf", width=2, height = 2)
posECDFRPF
dev.off()
```

```
## pdf
##   2
```

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] dplyr_1.1.2          rtracklayer_1.58.0   GenomicRanges_1.50.2
## [4] GenomeInfoDb_1.34.9  IRanges_2.32.0       S4Vectors_0.36.2
## [7] BiocGenerics_0.44.0  ggplot2_3.4.2
##
## loaded via a namespace (and not attached):
##  [1] lattice_0.20-45             tidyr_1.3.0
##  [3] Rsamtools_2.14.0            Biostrings_2.66.0
##  [5] digest_0.6.31               utf8_1.2.3
##  [7] R6_2.5.1                    backports_1.4.1
##  [9] evaluate_0.21               pillar_1.9.0
## [11] zlibbioc_1.44.0             rlang_1.1.0
## [13] rstudioapi_0.14             car_3.1-2
## [15] Matrix_1.5-3                rmarkdown_2.21
## [17] labeling_0.4.2              BiocParallel_1.32.6
## [19] RCurl_1.98-1.12             munsell_0.5.0
## [21] DelayedArray_0.23.2         broom_1.0.4
## [23] compiler_4.2.3              xfun_0.39
## [25] pkgconfig_2.0.3             htmltools_0.5.4
## [27] tidyselect_1.2.0            SummarizedExperiment_1.28.0
## [29] tibble_3.2.1                GenomeInfoDbData_1.2.9
## [31] codetools_0.2-19            matrixStats_0.63.0
## [33] XML_3.99-0.14               fansi_1.0.4
## [35] crayon_1.5.2                withr_2.5.0
```

```
## [37] ggpubr_0.6.0               GenomicAlignments_1.34.1
## [39] bitops_1.0-7               grid_4.2.3
## [41] gtable_0.3.3               lifecycle_1.0.3
## [43] magrittr_2.0.3             scales_1.2.1
## [45] cli_3.6.0                  carData_3.0-5
## [47] farver_2.1.1               XVector_0.38.0
## [49] ggsignif_0.6.4             generics_0.1.3
## [51] vctrs_0.6.2                rjson_0.2.21
## [53] restfulr_0.0.15            tools_4.2.3
## [55] Biobase_2.58.0             glue_1.6.2
## [57] purrr_1.0.1                MatrixGenerics_1.10.0
## [59] abind_1.4-5                parallel_4.2.3
## [61] fastmap_1.1.1              yaml_2.3.7
## [63] colorspace_2.1-0           rstatix_0.7.2
## [65] knitr_1.42                 BiocIO_1.8.0
```