# find_MMsat4

## Nikita Verheyden

### 2023-05-10

## Setup

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Figure2/MMsat4")
```

## packages

```
library(AnnotationHub)
```

```
## 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: BiocFileCache

## Loading required package: dbplyr
```

```
library(dplyr)
```

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

## The following objects are masked from 'package:dbplyr':
##
##     ident, sql

## 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(ggplot2)
library(ggpubr)
library(rtracklayer)
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: stats4
```

```
## Loading required package: S4Vectors
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, rename
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##
## Attaching package: 'IRanges'
```

```
## The following objects are masked from 'package:dplyr':
##
##     collapse, desc, slice
```

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

```
## Loading required package: GenomeInfoDb
```

```
##
## Attaching package: 'rtracklayer'
```

```
## The following object is masked from 'package:AnnotationHub':
##
##     hubUrl
```

**data**

```
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

#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"
```

## Get annotation

```
# ah = AnnotationHub()
# query(ah, c("RepeatMasker", "Mus musculus"))
# repeat_masker <- ah[["AH99012"]]

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

```
## GRanges object with 1693 ranges and 11 metadata columns:
##                  seqnames              ranges strand |   swScore   milliDiv
##                     <Rle>           <IRanges>  <Rle> | <integer> <numeric>
##       [1]            chr1   75624925-75625061      + |       347       293
##       [2]            chr1   75625261-75625377      + |       228       292
##       [3]            chr1   75625502-75625561      + |       257       233
##       [4]            chr1 116745797-116745936      + |       357       314
##       [5]            chr1 116745987-116746488      + |       600       228
##       ...             ...                 ...    ... .       ...       ...
##    [1689] chr9_KB469738_fix       24254-24454      - |       304       289
##    [1690] chr9_KB469738_fix       24588-24803      - |       440       281
##    [1691] chr9_KB469738_fix       25005-25142      - |       399       270
##    [1692] chr9_KB469738_fix       25257-25488      - |       389       275
##    [1693] chrY_JH792832_fix     349767-349817      + |       230       180
##           milliDel   milliIns   genoLeft      repName    repClass   repFamily
##          <numeric>  <numeric>  <integer>  <character> <character> <character>
##       [1]       29         29 -119846910       MMSAT4    Satellite   Satellite
##       [2]       34         34 -119846594       MMSAT4    Satellite   Satellite
##       [3]        0          0 -119846410       MMSAT4    Satellite   Satellite
##       [4]        7          0  -78726035       MMSAT4    Satellite   Satellite
##       [5]       48         22  -78725483       MMSAT4    Satellite   Satellite
##       ...      ...        ...        ...          ...          ...         ...
##    [1689]       24         16    -186187       MMSAT4    Satellite   Satellite
##    [1690]        0          0    -185838       MMSAT4    Satellite   Satellite
```

```
##    [1691]           7           7     -185499       MMSAT4    Satellite    Satellite
##    [1692]          21          21     -185153       MMSAT4    Satellite    Satellite
##    [1693]           0          20     -194372       MMSAT4    Satellite    Satellite
##            repStart     repEnd     repLeft
##           <integer>  <integer>  <integer>
##      [1]         32        168           0
##      [2]         32        148         -20
##      [3]         31         90         -78
##      [4]         27        167          -1
##      [5]          1        514           0
##      ...        ...        ...         ...
##    [1689]          0        202           1
##    [1690]          0        216           1
##    [1691]          0        168          31
##    [1692]          0        232           1
##    [1693]        115        164          -4
##    -------
##    seqinfo: 239 sequences (1 circular) from mm10 genome
```

## find overlaps

```r
# same strand
MMSAT4 <- repeat_masker[repeat_masker$repName == "MMSAT4"]

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


# opposite strand
antiMMSAT4 <- MMSAT4
strand(antiMMSAT4) <- ifelse(strand(MMSAT4) == '+', '-', '+')

antiOLgenes <- as.data.frame(subsetByOverlaps(gff23, antiMMSAT4))

dim(OLgenes)
```

```
## [1] 3806    28
```

```r
dim(antiOLgenes)
```

```
## [1] 140   28
```

```r
RNA$MMsat4 <- "No_MMsat4"
RNA$MMsat4[RNA$gene_symbol %in% OLgenes$gene_name] <- "MMsat4"
RNA$MMsat4[RNA$gene_symbol %in% antiOLgenes$gene_name] <- "anti-MMsat4"
table(RNA$MMsat4)
```

```
##
## anti-MMsat4      MMsat4    No_MMsat4
##          12         238        13051
```

```r
RPF$MMsat4 <- "No_MMsat4"
RPF$MMsat4[RPF$gene_symbol %in% OLgenes$gene_name] <- "MMsat4"
RPF$MMsat4[RPF$gene_symbol %in% antiOLgenes$gene_name] <- "anti-MMsat4"
table(RPF$MMsat4)
```

```
##
```

```
## anti-MMsat4        MMsat4    No_MMsat4
##          12          225        11132
```

# plot into MA plots

```r
#RNA
RNAMA <- ggplot(RNA, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(MMsat4, levels = c("No_MMsat4
  geom_point(shape=21, size=2, colour=farbeneg) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe3))+
  coord_cartesian(ylim = c(-3,3), xlim = c(0,6))+
  geom_point(data=RNA[RNA$gene_symbol %in% OLgenes$gene_name,], aes(x=log10(baseMean), y=log2FoldChange
  geom_point(data=RNA[RNA$gene_symbol %in% antiOLgenes$gene_name,], aes(x=log10(baseMean), y=log2FoldCha
  geom_hline(yintercept = 0, colour= "black") +
  theme_bw() +
  theme(legend.title = element_blank()) +
  ggtitle("RNA")

#RPF
RPFMA <- ggplot(RPF, aes(x=log10(baseMean), y=log2FoldChange, fill=factor(MMsat4, levels = c("No_MMsat4
  geom_point(shape=21, size=2, colour=farbeneg) +
  scale_fill_manual(values = c(farbeneg, farbe1, farbe3))+
  coord_cartesian(ylim = c(-3,3), xlim = c(0,6))+
  geom_point(data=RPF[RPF$gene_symbol %in% OLgenes$gene_name,], aes(x=log10(baseMean), y=log2FoldChange
  geom_point(data=RPF[RPF$gene_symbol %in% antiOLgenes$gene_name,], aes(x=log10(baseMean), y=log2FoldCha
  geom_hline(yintercept = 0, colour= "black") +
  theme_bw() +
  theme(legend.title = element_blank()) +
  ggtitle("RPF")

RNAMA
```
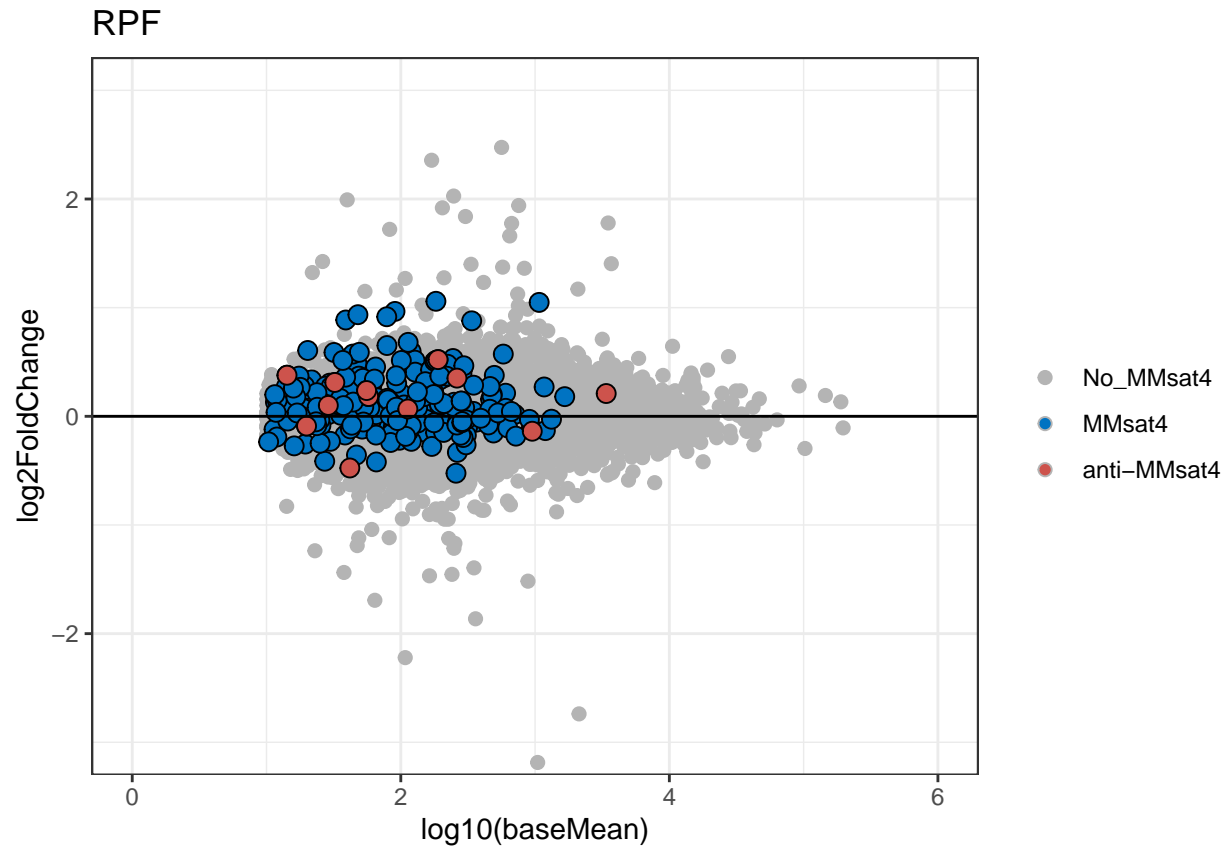
RPFMA

## RPF
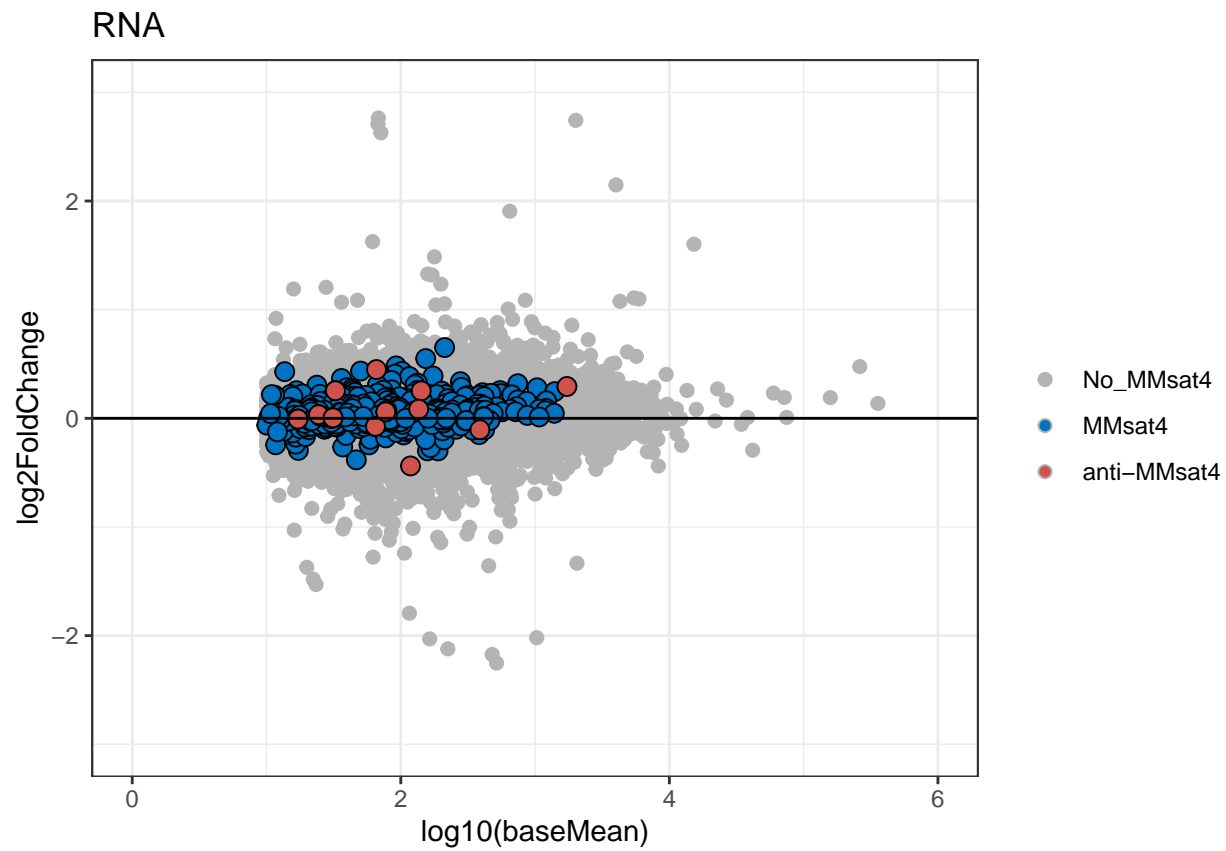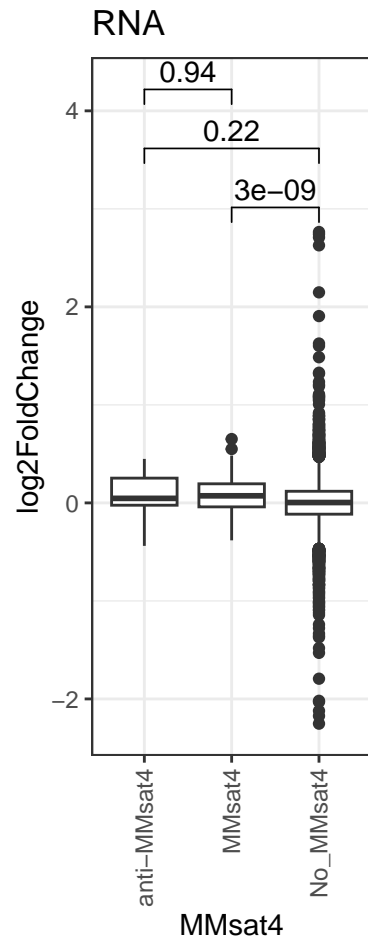


```r
my_comparisons <- list( c("No_MMsat4", "MMsat4"), c("No_MMsat4", "anti-MMsat4"), c("MMsat4", "anti-MMsat

RNAbox <- ggplot(RNA, aes(x=MMsat4, y=log2FoldChange)) +
  geom_boxplot() +
  stat_compare_means(comparisons = my_comparisons) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) +
  ggtitle("RNA")

RPFbox <- ggplot(RPF, aes(x=MMsat4, y=log2FoldChange)) +
  geom_boxplot() +
  stat_compare_means(comparisons = my_comparisons) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)) +
  ggtitle("RPF")

RNAbox
```
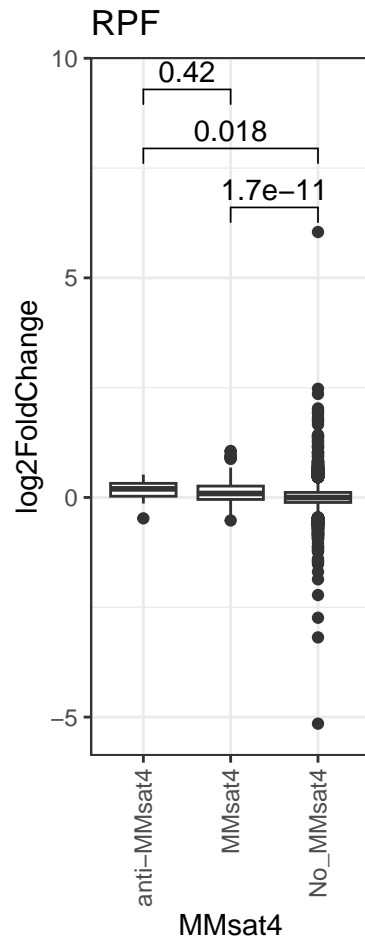
RNA

RPFbox

## RPF



Boxplot titled "RPF" with y-axis "log2FoldChange" (range −5 to 10) and x-axis groups "anti−MMsat4", "MMsat4", "No_MMsat4" under "MMsat4". Pairwise comparison p-values: 0.42, 0.018, 1.7e−11.

# 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] rtracklayer_1.58.0   GenomicRanges_1.50.2 GenomeInfoDb_1.34.9
##  [4] IRanges_2.32.0       S4Vectors_0.36.2     ggpubr_0.6.0
```

```
##  [7] ggplot2_3.4.2         dplyr_1.1.2              AnnotationHub_3.6.0
## [10] BiocFileCache_2.6.1 dbplyr_2.3.2              BiocGenerics_0.44.0
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-7                  matrixStats_0.63.0
##  [3] bit64_4.0.5                   filelock_1.0.2
##  [5] httr_1.4.6                    tools_4.2.3
##  [7] backports_1.4.1              utf8_1.2.3
##  [9] R6_2.5.1                      DBI_1.1.3
## [11] colorspace_2.1-0            withr_2.5.0
## [13] tidyselect_1.2.0            bit_4.0.5
## [15] curl_5.0.0                    compiler_4.2.3
## [17] cli_3.6.0                     Biobase_2.58.0
## [19] DelayedArray_0.23.2         labeling_0.4.2
## [21] scales_1.2.1                  rappdirs_0.3.3
## [23] digest_0.6.31                Rsamtools_2.14.0
## [25] rmarkdown_2.21                XVector_0.38.0
## [27] pkgconfig_2.0.3              htmltools_0.5.4
## [29] MatrixGenerics_1.10.0       highr_0.10
## [31] fastmap_1.1.1                rlang_1.1.0
## [33] rstudioapi_0.14              RSQLite_2.3.1
## [35] shiny_1.7.4                   farver_2.1.1
## [37] BiocIO_1.8.0                  generics_0.1.3
## [39] BiocParallel_1.32.6         car_3.1-2
## [41] RCurl_1.98-1.12              magrittr_2.0.3
## [43] GenomeInfoDbData_1.2.9      Matrix_1.5-3
## [45] Rcpp_1.0.10                   munsell_0.5.0
## [47] fansi_1.0.4                   abind_1.4-5
## [49] lifecycle_1.0.3              yaml_2.3.7
## [51] carData_3.0-5                SummarizedExperiment_1.28.0
## [53] zlibbioc_1.44.0              grid_4.2.3
## [55] blob_1.2.4                    parallel_4.2.3
## [57] promises_1.2.0.1            crayon_1.5.2
## [59] lattice_0.20-45              Biostrings_2.66.0
## [61] KEGGREST_1.38.0              knitr_1.42
## [63] pillar_1.9.0                  rjson_0.2.21
## [65] ggsignif_0.6.4                codetools_0.2-19
## [67] XML_3.99-0.14                glue_1.6.2
## [69] BiocVersion_3.16.0          evaluate_0.21
## [71] BiocManager_1.30.20         png_0.1-8
## [73] vctrs_0.6.2                   httpuv_1.6.11
## [75] gtable_0.3.3                  purrr_1.0.1
## [77] tidyr_1.3.0                   cachem_1.0.8
## [79] xfun_0.37                     mime_0.12
## [81] xtable_1.8-4                  broom_1.0.4
## [83] restfulr_0.0.15              rstatix_0.7.2
## [85] later_1.3.1                   tibble_3.2.1
## [87] GenomicAlignments_1.34.1    AnnotationDbi_1.60.2
## [89] memoise_2.0.1                ellipsis_0.3.2
## [91] interactiveDisplayBase_1.36.0
```