Bound genes and gene regions of MREs

Melina Klostermann

13 September, 2023

Contents

1 Libraries and settings

```
# ----------------------------------------
# libraries
# ----------------------------------------
library(tidyverse)
library(GenomicRanges)
library(colorspace)
library(eulerr)
library(gghalves)

# ----------------------------------------
# settings
# ----------------------------------------

here <- here::here()

source(paste0(here,"/Supporting_scripts/themes/theme_paper.R"))
source(paste0(here,"/Supporting_scripts/themes/CustomThemes.R"))


out <- paste0(here,"/Figure4/01_MRE_bound_gene_and_bound_region/")
```

```
# farben
farbe4 <- "#7AA6DCFF"

farbe6 <- "#003C67FF"
```

2 What was done?

- the genetype and gene region of the mir 181 binding sites (union) are ploted (Figure4)

3 Files

```
# ----------------------------------------
# MREs
# ----------------------------------------

mir181_bs <- readRDS(paste0(here, "/Figure2/01_mir181-enriched_binding_site_definition/mir181_bs.rds"))

names(mir181_bs) <- 1:NROW(mir181_bs)
mir181_bs <- as.data.frame(mir181_bs)

mir181_enriched_set <- mir181_bs %>%
  as.data.frame(.) %>%
  subset(set %in% c("ago_bs_mir181_chi&mir181_enriched", "mir181_enriched"))
```

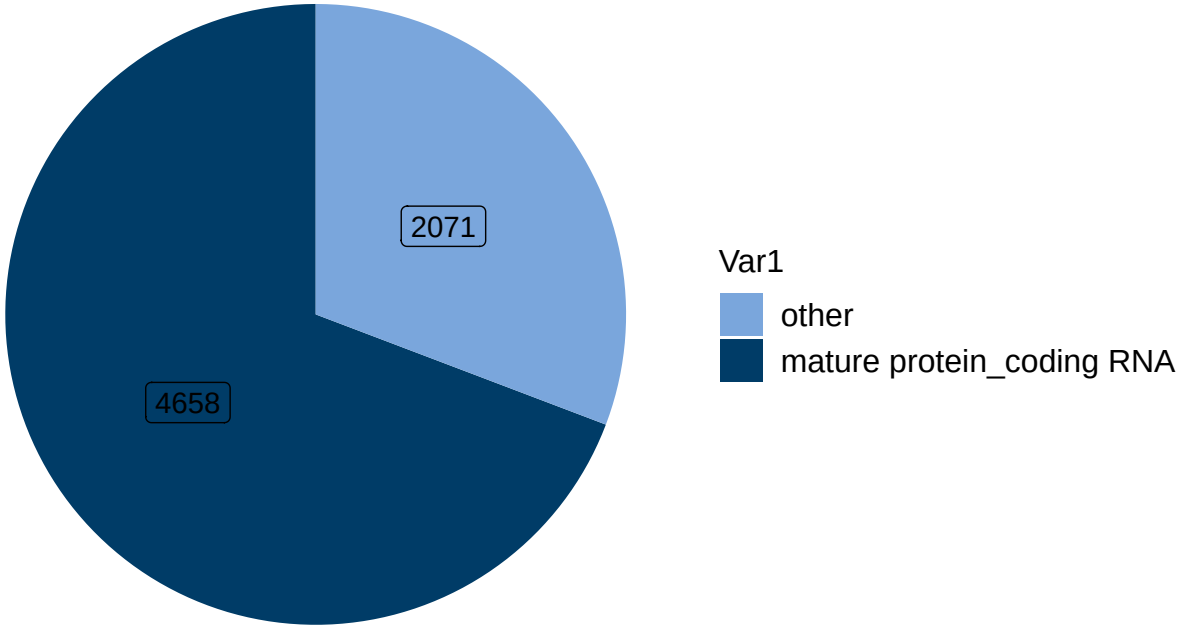
4 Characterise MRE

4.1 mir181 bound genes - Figure 4A

```
mir181_enriched_set <- mutate(mir181_enriched_set, geneType =
                          case_when(geneType != "protein_coding" ~ "other",
                                    region == "intron" ~ "other",
                                    region == "outside" ~ "other",
                                    is.na(region) ~ "other",
                                     T ~ "mature protein_coding RNA"))
gene_type_df <-   table(mir181_enriched_set$geneType) %>%
  as.data.frame(.)

p <- ggplot(gene_type_df, aes(y=Freq, x="", fill=Var1)) +
     geom_col()+
     coord_polar(theta="y") +
  #    xlim(c(2, 4)) +
  geom_label(data = gene_type_df, aes(y=Freq, x="", fill=Var1, label = Freq),
             position = position_stack(vjust = 0.5),
             show.legend = FALSE) +
  scale_fill_manual(values = c (farbe6, farbe4)) +
  theme_paper() +
  theme_nice_pie() +
  #theme(legend.position = "none") +
  guides(fill = guide_legend(reverse = TRUE)) +
  labs(y = NULL,
```

```
       x = NULL)
p
```




```
ggsave(p, filename = paste0(out, "Figure4A_bound_gene_types_miR181_BS.pdf"), width = unit(8, "cm"), hei
```

4.1.1 Remove non protein-coding binding sites

For all further analyses we removed binding sites on non protein-coding RNAs.

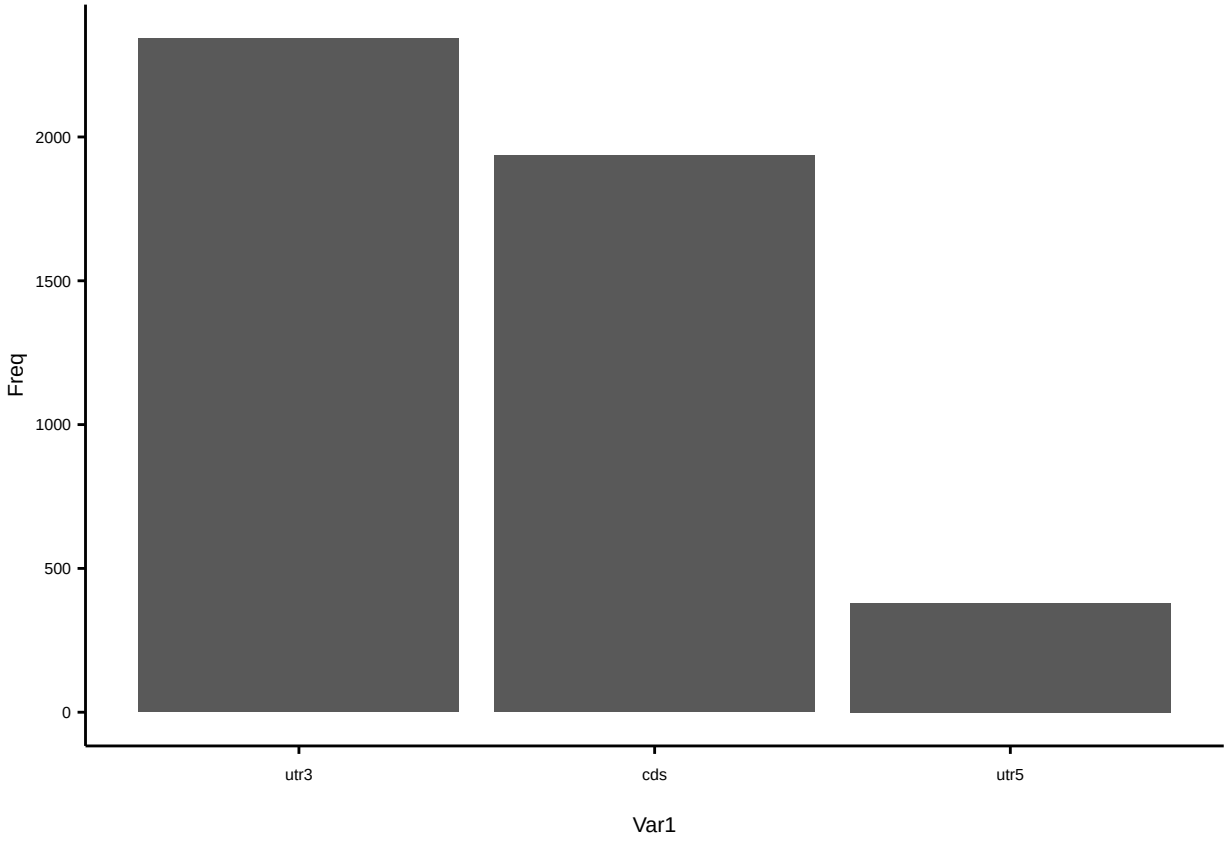
```
mir181_enriched_set <- subset(mir181_enriched_set, geneType == "mature protein_coding RNA")
```

4.2 mir181 bound regions - Figure 4B

```
gene_region_df <- table(mir181_enriched_set$region) %>%
  as.data.frame(.) %>%
  arrange(desc(Freq))

gene_region_df$Var1 <- factor(gene_region_df$Var1, levels = gene_region_df$Var1)

p <- ggplot(gene_region_df %>% subset(Var1 != "outside"), aes(y=Freq, x=Var1)) +
     geom_col()+
  theme_paper()
p
```




```
ggsave(p, filename = paste0(out, "Figure4d_bound_gene_regions_miR181_BS.pdf"), width = unit(6, "cm"), he

sum(gene_region_df$Freq)
```

```
## [1] 4658
```

4.2.1 Remove binding sites in introns or outside of any gene region

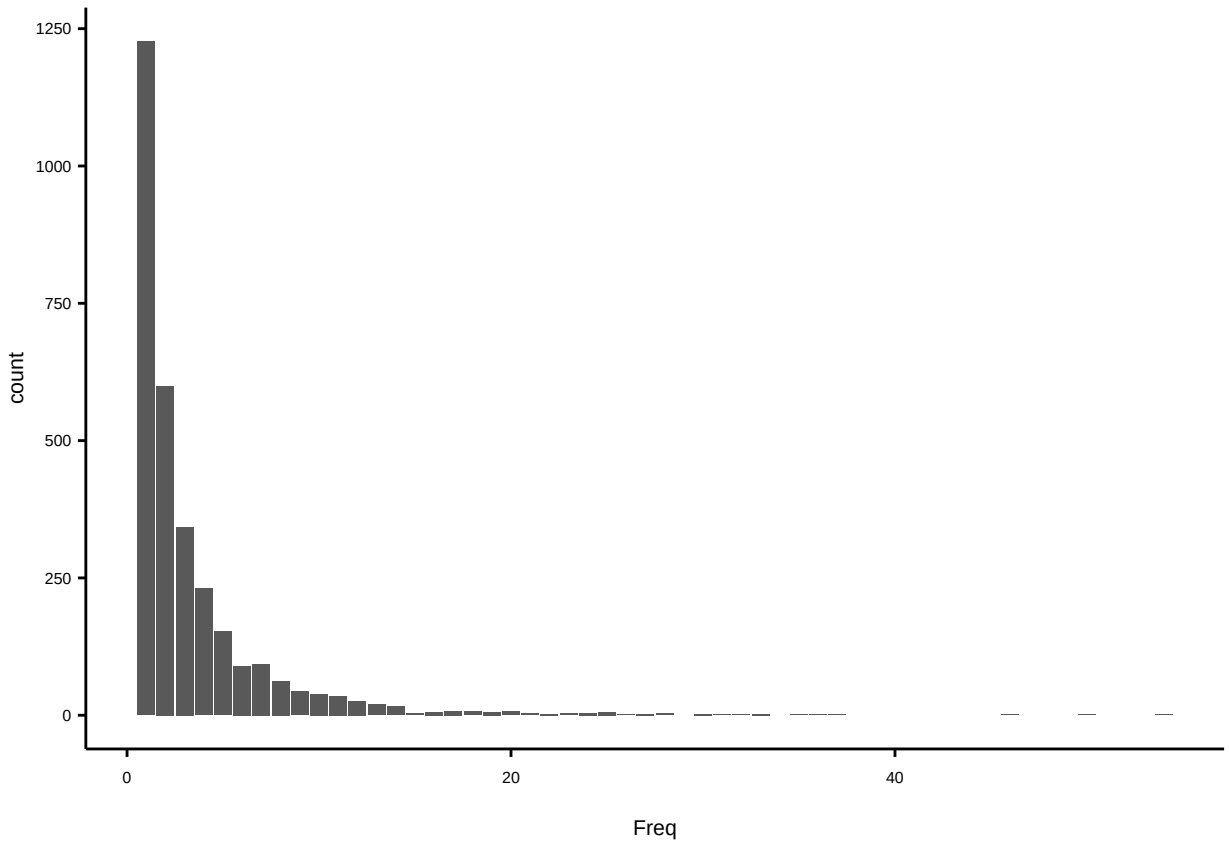
For all further analyses we removed binding sites in introns or in a gene regions with no region annotation.

```
mir181_bs <- mutate(mir181_bs, geneType =
                        case_when(geneType != "protein_coding" ~ "other",
                                  region == "intron" ~ "other",
                                  region == "outside" ~ "other",
                                  is.na(region) ~ "other",
                                   T ~ "mature protein_coding RNA")) %>%
  subset(.,  geneType == "mature protein_coding RNA")
```

5 Number of MREs per transcript

```
df <- table(mir181_bs$geneID) %>%
  as.data.frame(.)

ggplot(df, aes(x = Freq))+
  geom_bar()+
  theme_paper()
```




```
ggsave(p, filename = paste0(out, "Figure4b_MREs_per_transcript.pdf"), width = unit(6, "cm"), height = u
```

6 Save filtered BS

```
saveRDS(mir181_bs, paste0(out, "mir181_bs_afterFigure4B.rds"))
```

7 Session Info

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
```

```
## other attached packages:
##  [1] gghalves_0.1.4       eulerr_7.0.0         colorspace_2.1-0
##  [4] GenomicRanges_1.50.2 GenomeInfoDb_1.34.9  IRanges_2.32.0
##  [7] S4Vectors_0.36.2     BiocGenerics_0.44.0  lubridate_1.9.2
## [10] forcats_1.0.0        stringr_1.5.0        dplyr_1.1.2
## [13] purrr_1.0.1          readr_2.1.4          tidyr_1.3.0
## [16] tibble_3.2.1         ggplot2_3.4.2        tidyverse_2.0.0
## [19] knitr_1.43
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.11            here_1.0.1             rprojroot_2.0.3
##  [4] digest_0.6.33          utf8_1.2.3             R6_2.5.1
##  [7] backports_1.4.1        evaluate_0.21          highr_0.10
## [10] pillar_1.9.0           zlibbioc_1.44.0        rlang_1.1.1
## [13] rstudioapi_0.15.0      car_3.1-2              rmarkdown_2.23
## [16] textshaping_0.3.6      labeling_0.4.2         RCurl_1.98-1.12
## [19] munsell_0.5.0          broom_1.0.5            compiler_4.2.2
## [22] xfun_0.39              pkgconfig_2.0.3        systemfonts_1.0.4
## [25] htmltools_0.5.5        tidyselect_1.2.0       GenomeInfoDbData_1.2.9
## [28] fansi_1.0.4            tzdb_0.4.0             withr_2.5.0
## [31] ggpubr_0.6.0           bitops_1.0-7           grid_4.2.2
## [34] gtable_0.3.3           lifecycle_1.0.3        magrittr_2.0.3
## [37] scales_1.2.1           cli_3.6.1              stringi_1.7.12
## [40] carData_3.0-5          farver_2.1.1           XVector_0.38.0
## [43] ggsignif_0.6.4         ragg_1.2.5             generics_0.1.3
## [46] vctrs_0.6.3            tools_4.2.2            glue_1.6.2
## [49] hms_1.1.3              abind_1.4-5            fastmap_1.1.1
## [52] yaml_2.3.7             timechange_0.2.0       rstatix_0.7.2
```