Bound genes and gene regions of MREs

Melina Klostermann

10 July, 2023

Contents

1 Libraries and settings

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


# ----------------------------------------
# settings
# ----------------------------------------
out <- "/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/Figure2/01_MRE_
source("/Users/melinaklostermann/Documents/projects/R_general_functions/theme_paper.R")
source("/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/mirko_files/mirECLIP/DifferentialBind


# farben
farbeneg <- "#B4B4B4"
farbe1 <- "#0073C2FF" #WT farbe
farbe2 <- "#EFC000FF"
farbe3 <- "#CD534CFF" #miR181KO farbe
farbe4 <- "#7AA6DCFF"
farbe5 <- "#868686FF"
farbe6 <- "#003C67FF"
farbe7 <- "#8F7700FF"
```

```r
farbe8 <- "#3B3B3BFF"
farbe9 <- "#A73030FF"
farbe10 <- "#4A6990FF"
farbe11 <- "#FF6F00FF"
farbe12 <- "#C71000FF"
farbe13 <- "#008EA0FF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"
```

2 What was done?

- the genetype and gene region of the mir 181 binding sites (union) are ploted (Figure2XX)

3 Files

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

mir181_bs <- readRDS("/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/

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 2A

```r
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 %>% subset(gene_type_df == "mature protein_coding RNA"), aes(y=Freq, x=
            position = position_stack(vjust = 0.5),
            show.legend = FALSE) +
  scale_fill_manual(values = c (farbe6, farbe4)) +
  theme_paper() +
```
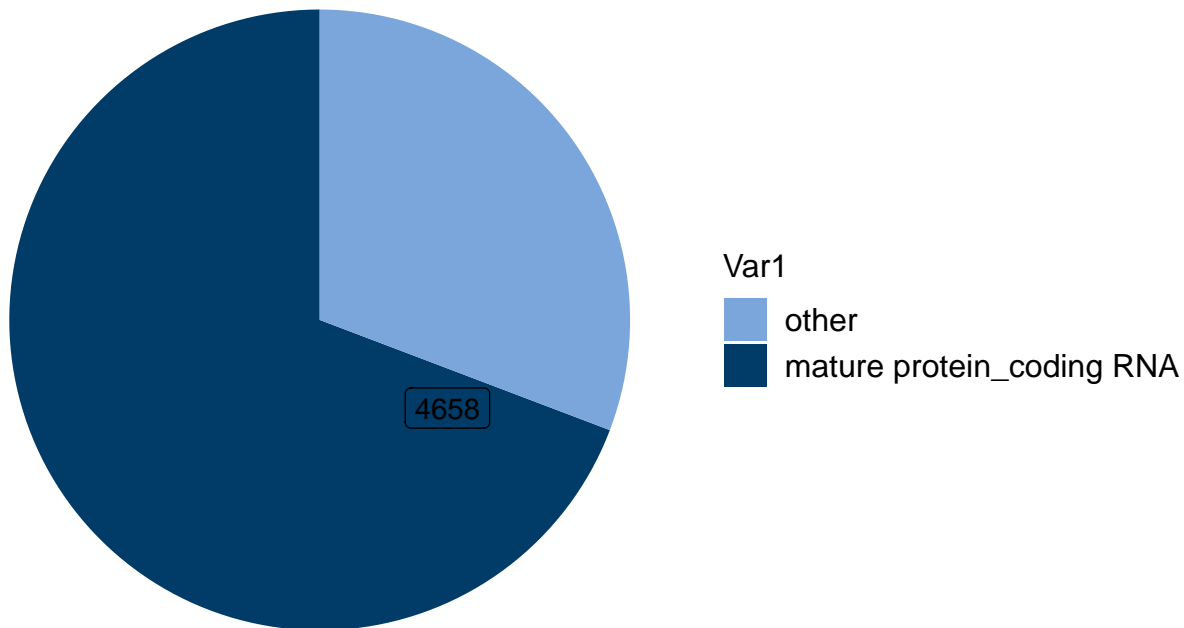
```
  theme_nice_pie() +
  #theme(legend.position = "none") +
  guides(fill = guide_legend(reverse = TRUE)) +
  labs(y = NULL,
       x = NULL)
p
```



```
ggsave(p, filename = paste0(out, "Figure2A_bound_gene_types_miR181_BS", Sys.Date(), ".pdf"), width = un
```

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 2B

```
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
```
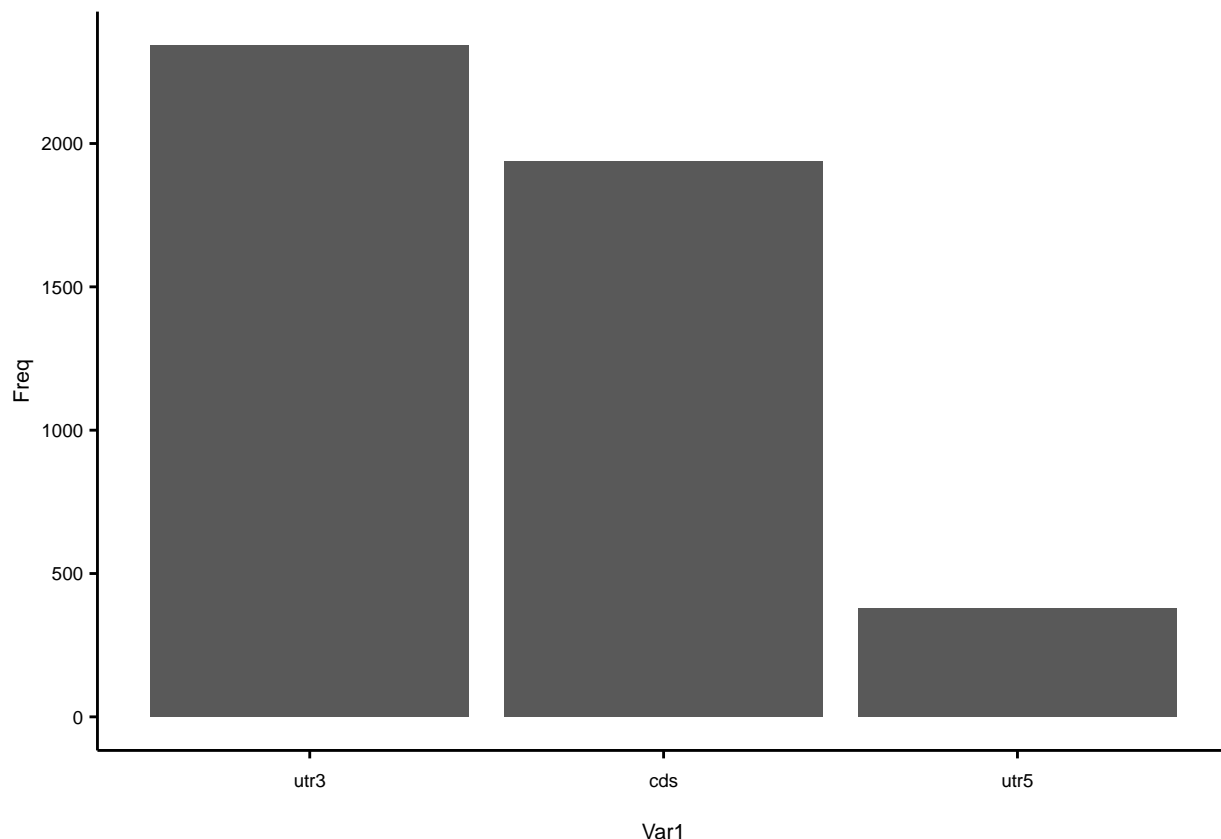
```
ggsave(p, filename = paste0(out, "Figure2B_bound_gene_regions_miR181_BS", Sys.Date(), ".pdf"), width = u

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 Save filtered BS

```
saveRDS(mir181_bs, paste0(out, "mir181_bs_afterFigure2B.rds"))
```

6 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] tidyselect_1.2.0   xfun_0.39               carData_3.0-5
##  [4] vctrs_0.6.3        generics_0.1.3          htmltools_0.5.5
##  [7] yaml_2.3.7         utf8_1.2.3              rlang_1.1.1
## [10] pillar_1.9.0       ggpubr_0.6.0            glue_1.6.2
## [13] withr_2.5.0        GenomeInfoDbData_1.2.9  lifecycle_1.0.3
## [16] zlibbioc_1.44.0    munsell_0.5.0           ggsignif_0.6.4
## [19] gtable_0.3.3       ragg_1.2.5              evaluate_0.21
## [22] labeling_0.4.2     tzdb_0.4.0              fastmap_1.1.1
## [25] fansi_1.0.4        highr_0.10              broom_1.0.5
## [28] Rcpp_1.0.10        backports_1.4.1         scales_1.2.1
## [31] XVector_0.38.0     abind_1.4-5             systemfonts_1.0.4
## [34] farver_2.1.1       textshaping_0.3.6       hms_1.1.3
## [37] digest_0.6.31      stringi_1.7.12          rstatix_0.7.2
## [40] grid_4.2.2         cli_3.6.1               tools_4.2.2
## [43] bitops_1.0-7       magrittr_2.0.3          RCurl_1.98-1.12
## [46] car_3.1-2          pkgconfig_2.0.3         timechange_0.2.0
## [49] rmarkdown_2.22     rstudioapi_0.14         R6_2.5.1
## [52] compiler_4.2.2
```