Bound genes and gene regions of MREs

Melina Klostermann

24 April, 2023

Contents

1 Libraries and settings

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

# ----------------------------------------
# settings
# ----------------------------------------
out <- "/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/Figure2/MRE_bou
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"
```

```
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

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

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

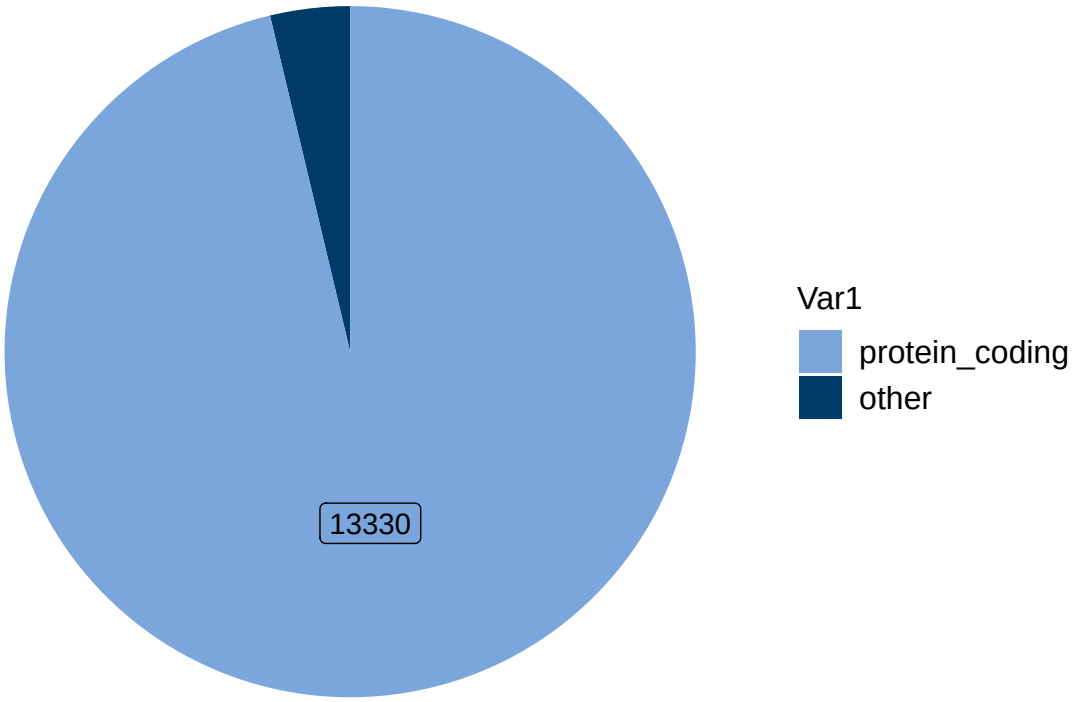
4 Characterise MRE

4.1 mir181 bound genes - Figure 2A

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

gene_type_df <- mutate(mir181_bs, geneType = case_when(geneType != "protein_coding" ~ "other", T ~ "pro
gene_type_df <-   table(gene_type_df$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 == "protein_coding"), aes(y=Freq, x="", fill=Va
             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 = uni
```

4.1.1 Remove non protein-coding binding sites

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

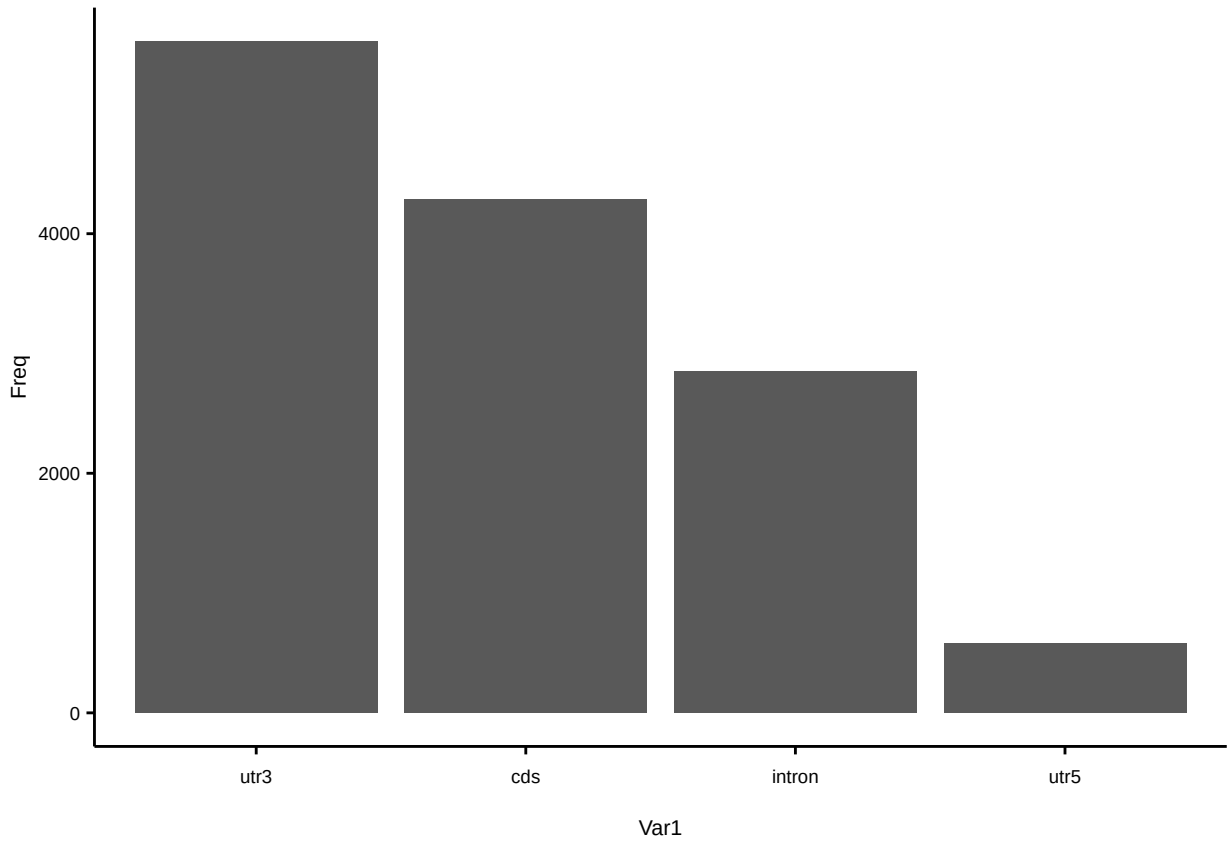
```
mir181_bs <- subset(mir181_bs, geneType == "protein_coding")
```

4.2 mir181 bound regions - Figure 2B

```
gene_region_df <- table(mir181_bs$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 =
```

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 <- subset(mir181_bs, !(region %in% c("outside", "intron") ))
```

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_1.34.7  IRanges_2.32.0
##  [7] S4Vectors_0.36.1     BiocGenerics_0.44.0  forcats_0.5.2
## [10] stringr_1.5.0        dplyr_1.0.10         purrr_1.0.1
## [13] readr_2.1.3          tidyr_1.3.0          tibble_3.1.8
## [16] ggplot2_3.4.0        tidyverse_1.3.2      knitr_1.42
##
## loaded via a namespace (and not attached):
##  [1] httr_1.4.4             jsonlite_1.8.4         carData_3.0-5
##  [4] modelr_0.1.10          assertthat_0.2.1       highr_0.10
##  [7] GenomeInfoDbData_1.2.9 googlesheets4_1.0.1    cellranger_1.1.0
## [10] yaml_2.3.7             pillar_1.8.1           backports_1.4.1
## [13] glue_1.6.2             digest_0.6.31          XVector_0.38.0
## [16] ggsignif_0.6.4         rvest_1.0.3            htmltools_0.5.4
## [19] pkgconfig_2.0.3        broom_1.0.3            haven_2.5.1
## [22] zlibbioc_1.44.0        scales_1.2.1           tzdb_0.3.0
## [25] timechange_0.2.0       googledrive_2.0.0      farver_2.1.1
## [28] generics_0.1.3         car_3.1-1              ellipsis_0.3.2
## [31] ggpubr_0.5.0           withr_2.5.0            cli_3.6.0
## [34] magrittr_2.0.3         crayon_1.5.2           readxl_1.4.1
## [37] evaluate_0.20          fs_1.6.0               fansi_1.0.4
## [40] rstatix_0.7.1          xml2_1.3.3             textshaping_0.3.6
## [43] tools_4.2.2            hms_1.1.2              gargle_1.2.1
## [46] lifecycle_1.0.3        munsell_0.5.0          reprex_2.0.2
## [49] compiler_4.2.2         systemfonts_1.0.4      rlang_1.0.6
## [52] grid_4.2.2             RCurl_1.98-1.9         rstudioapi_0.14
## [55] bitops_1.0-7           labeling_0.4.2         rmarkdown_2.20
## [58] gtable_0.3.1           abind_1.4-5            DBI_1.1.3
## [61] R6_2.5.1               lubridate_1.9.1        fastmap_1.1.0
## [64] utf8_1.2.2             ragg_1.2.5             stringi_1.7.12
## [67] Rcpp_1.0.10            vctrs_0.5.2            dbplyr_2.3.0
## [70] tidyselect_1.2.0       xfun_0.36
```