Seed motifs

Melina Klostermann

25 April, 2023

Contents

1 Libraries and settings

```r
# ----------------------------------------
# libraries
# ----------------------------------------
library(tidyverse)
library(GenomicRanges)
library(colorspace)
library(gghalves)
library(BSgenome.Mmusculus.UCSC.mm10)
library(Biostrings)


# ----------------------------------------
# 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/I

seed_8mer_ac <- "UGAAUGUA"
seed_8mer_bd <- "UGAAUGUU"
seed_7mer_m8 <- "UGAAUGU"
seed_7mer_a1_ac <- "GAAUGUA"
seed_7mer_a1_bd <- "GAAUGUU"
seed_6mer <- "GAAUGU"
seed_6mer_wobble <- "GAUUGU"
seed_8mer_ac_wobble <- "UGAUUGUA"
seed_8mer_bd_wobble <- "UGAUUGUU"
seed_7mer_m8_wobble <- "UGAUUGU"
seed_7mer_a1_ac_wobble <- "GAUUGUA"
seed_7mer_a1_bd_wobble <- "GAUUGUU"

seed_list <- list(seed_8mer_ac, seed_8mer_bd, seed_7mer_m8, seed_7mer_a1_ac, seed_7mer_a1_bd, seed_6mer

seed_names_list <- list( "seed_8mer_ac", "seed_8mer_bd", "seed_7mer_m8", "seed_7mer_a1_ac", "seed_7mer_a
```

4   Position of seed relative to binding site

```
mir181_bs_area200nt <- makeGRangesFromDataFrame( mir181_bs, keep.extra.columns = T) + 200
BS_mir181_area200nt_seq <- getSeq(mir181_bs_area200nt, x = BSgenome.Mmusculus.UCSC.mm10) %>%
  RNAStringSet()

Seed_8 <- vmatchPattern(pattern = seed_8mer_ac, BS_mir181_area200nt_seq) %>% unlist()
Seed_7 <- vmatchPattern(pattern = seed_7mer_m8, BS_mir181_area200nt_seq) %>% unlist()
Seed_6 <- vmatchPattern(pattern = seed_6mer, BS_mir181_area200nt_seq) %>% unlist()

seed_gg <- data.frame(position_GAAUGU = start(Seed_8))
```
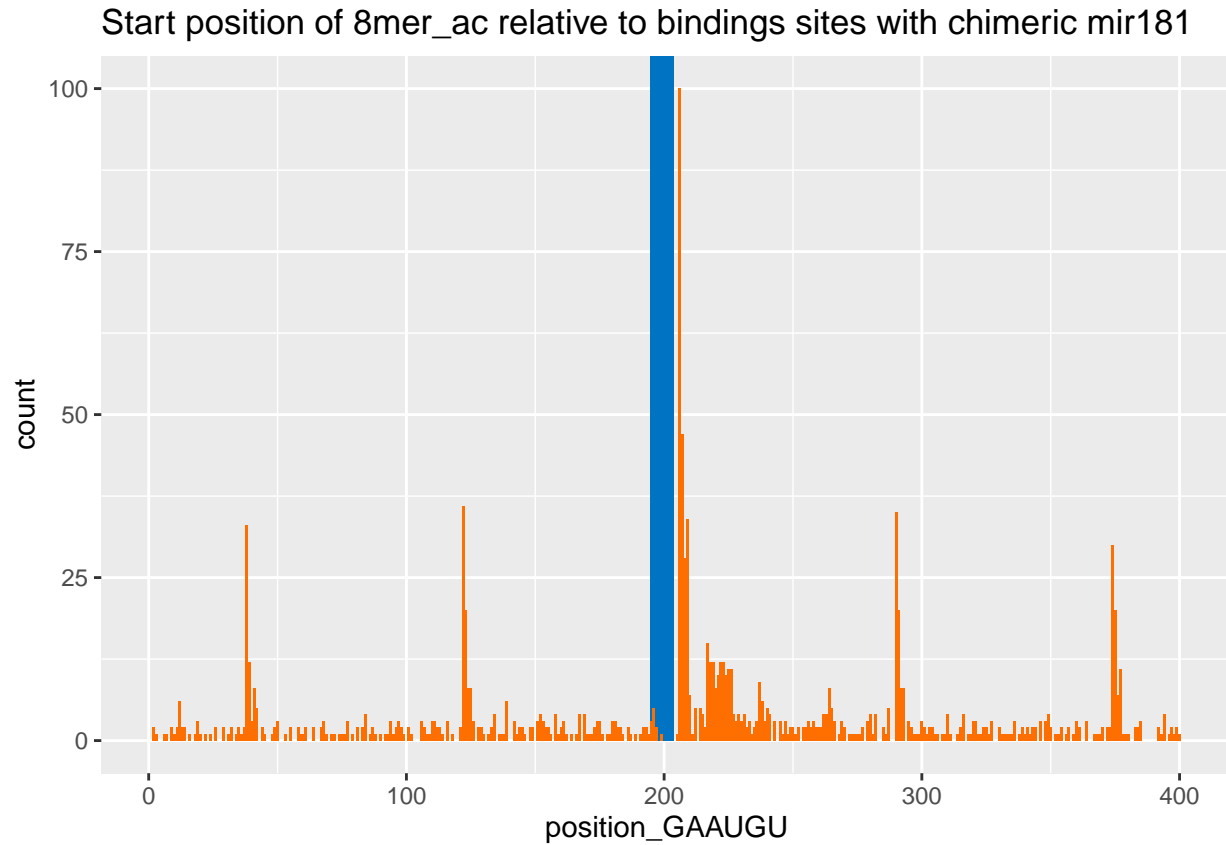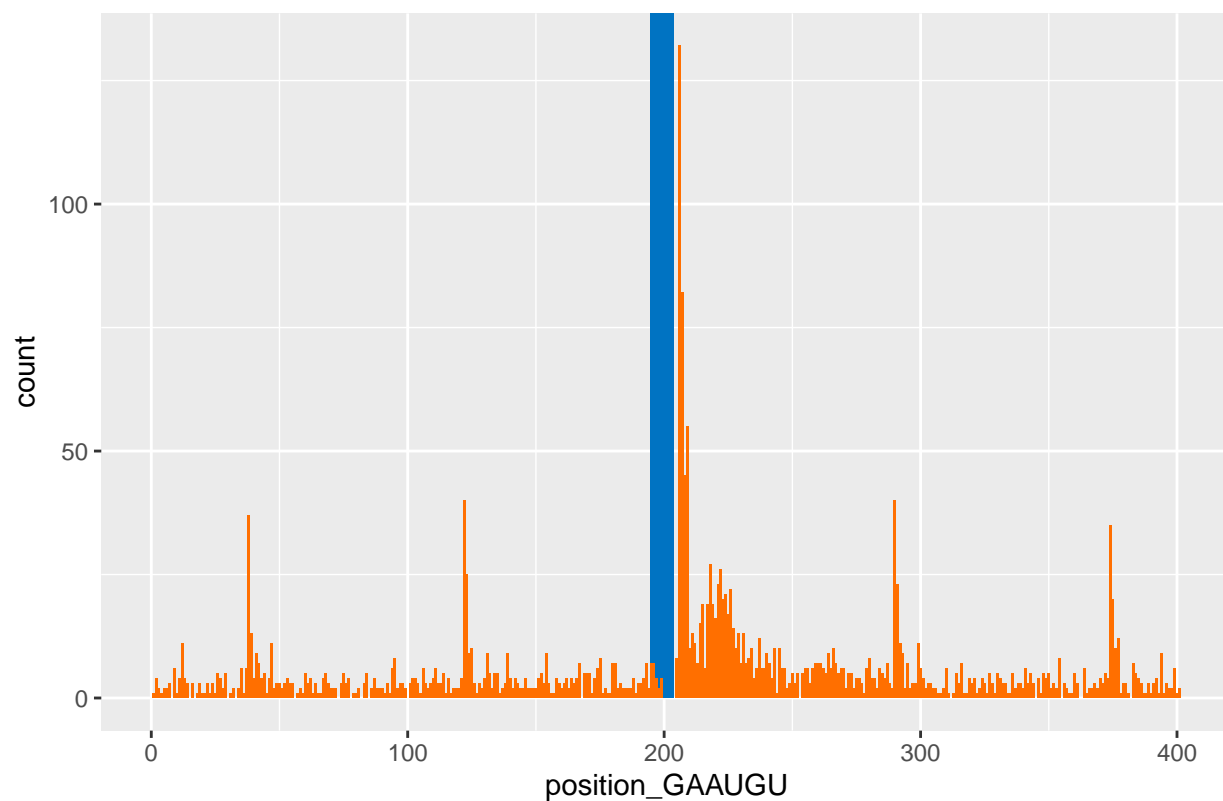
```
ggplot(seed_gg, aes(x = position_GAAUGU))+
  geom_rect(aes(xmin=194.5, xmax=203.5, ymin=0, ymax=Inf), fill = farbe1)+
  geom_bar(fill = farbe11)+
  ggtitle("Start position of 8mer_ac relative to bindings sites with chimeric mir181")
```

Start position of 8mer_ac relative to bindings sites with chimeric mir181


```
seed_gg <- data.frame(position_GAAUGU = start(Seed_7))

ggplot(seed_gg, aes(x = position_GAAUGU))+
  geom_rect(aes(xmin=194.5, xmax=203.5, ymin=0, ymax=Inf), fill = farbe1)+
  geom_bar(fill = farbe11)+
  ggtitle("Start position of 7mer_m8 relative to bindings sites with chimeric mir181")
```

Start position of 7mer_m8 relative to bindings sites with chimeric mir181


```
seed_gg <- data.frame(position_GAAUGU = start(Seed_6))

ggplot(seed_gg, aes(x = position_GAAUGU))+
  geom_rect(aes(xmin=194.5, xmax=203.5, ymin=0, ymax=Inf), fill = farbe1)+
  geom_bar(fill = farbe11)+
  ggtitle("Start position of GAAUGU relative to bindings sites with chimeric mir181")
```

Start position of GAAUGU relative to bindings sites with chimeric mir181
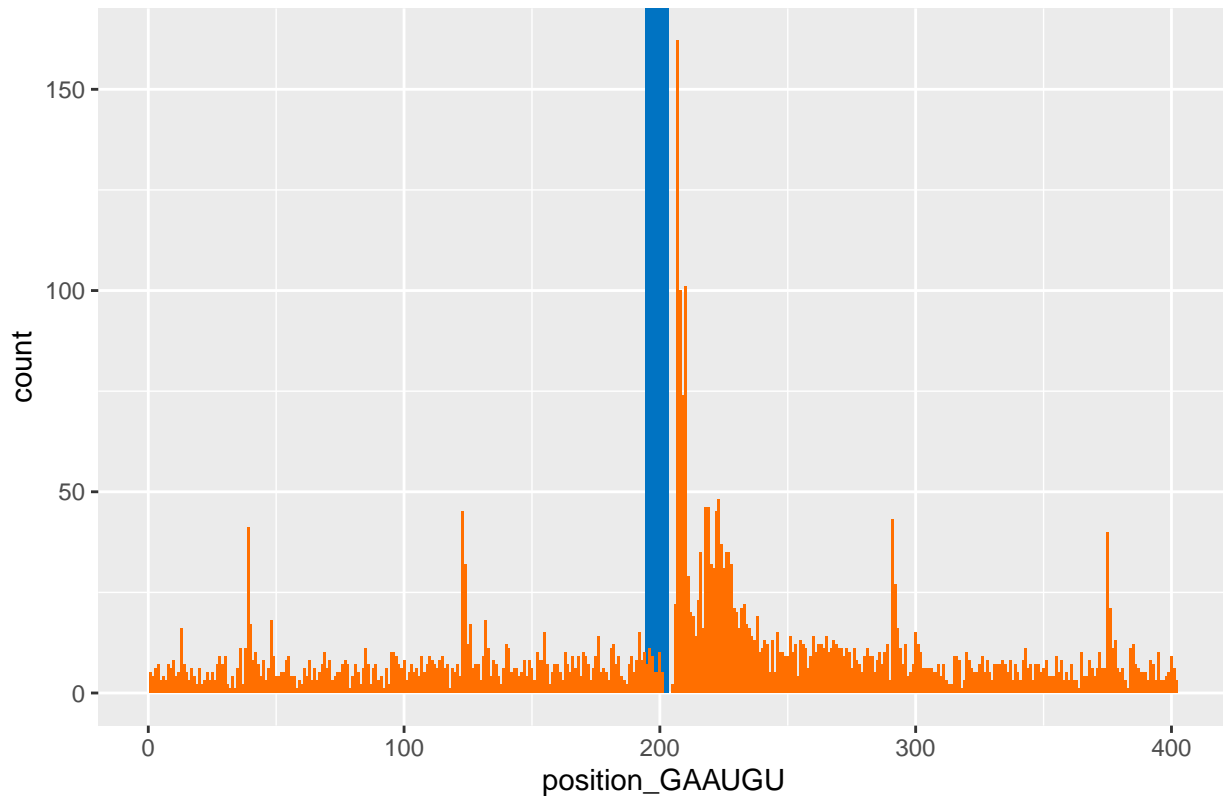
5 How many binding sites have a seed?

```r
mir181_bs_50down <- makeGRangesFromDataFrame( mir181_bs, keep.extra.columns = T) %>% flank(., width = 5
mir181_bs_50down_seq <- getSeq(mir181_bs_50down , x = BSgenome.Mmusculus.UCSC.mm10) %>%
  RNAStringSet()


Seeds <- lapply(seed_list, function(x) {
  vmatchPattern(pattern = x, mir181_bs_50down_seq) %>%
  lapply(., function(x) as.data.frame(x))})

BS_ID_list <- as.list(mir181_bs$BS_ID)

Seeds <- map(Seeds, ~map2(.x, BS_ID_list, ~mutate(.x, BS_ID = .y) ) %>%
            map_dfr(~.x))

Seeds <- map2(Seeds, seed_names_list, ~mutate(.x, seed = .y) ) %>% map_dfr(~.x)

seed_importance_order <- c( "seed_8mer_ac", "seed_8mer_bd", "seed_7mer_m8", "seed_7mer_a1_ac", "seed_7me


Seeds_1_per_BS <- Seeds %>%
  mutate(wobble = grepl("wobble", seed),
         seed = case_when(wobble  ~ substr(seed, 1, nchar(seed)-7), T ~ seed))
```
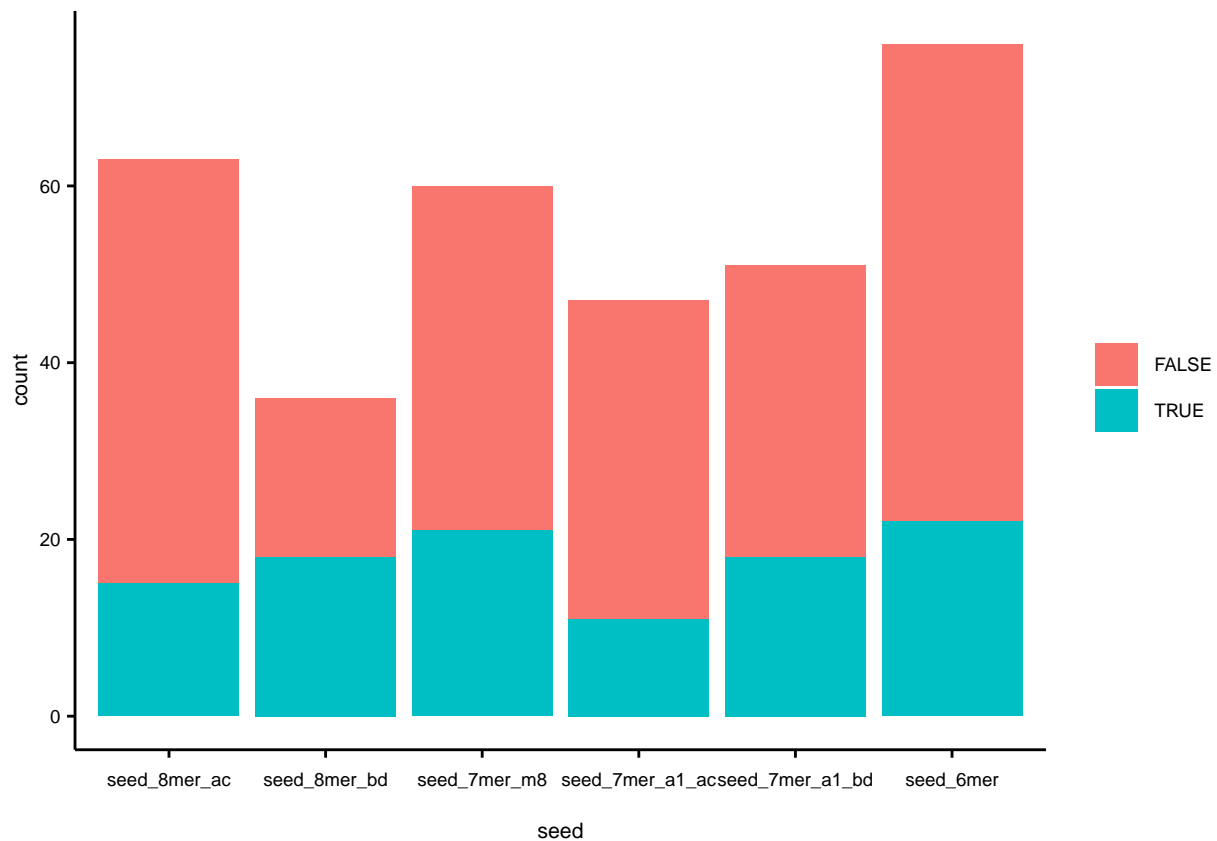
```
Seeds_1_per_BS$seed <- factor(Seeds_1_per_BS$seed, levels = seed_importance_order )

Seeds_1_per_BS <- Seeds_1_per_BS %>%
  group_by(BS_ID) %>%
 arrange(start, seed ) %>%
 dplyr::slice(1) %>%
 ungroup(.)

ggplot(Seeds_1_per_BS, aes(x = seed, fill = wobble))+
 geom_bar()+
 theme_paper()
```



```
nrow(Seeds_1_per_BS)/ nrow(mir181_bs)
```

```
## [1] 0.03179605
```

5.1 Distribution of seeds in sets

```
# Add seed info to BS
Seeds_1_per_BS <- Seeds_1_per_BS %>% select(c("BS_ID", "seed", "start", "wobble"))
colnames(Seeds_1_per_BS) <- c("BS_ID", "seed", "seed_start_from_BS", "wobble")

mir181_bs <- mir181_bs %>% left_join(., Seeds_1_per_BS, by = "BS_ID")

ggplot(mir181_bs, aes(x = seed, fill = wobble))+
 geom_bar()+
 theme_paper()+
```
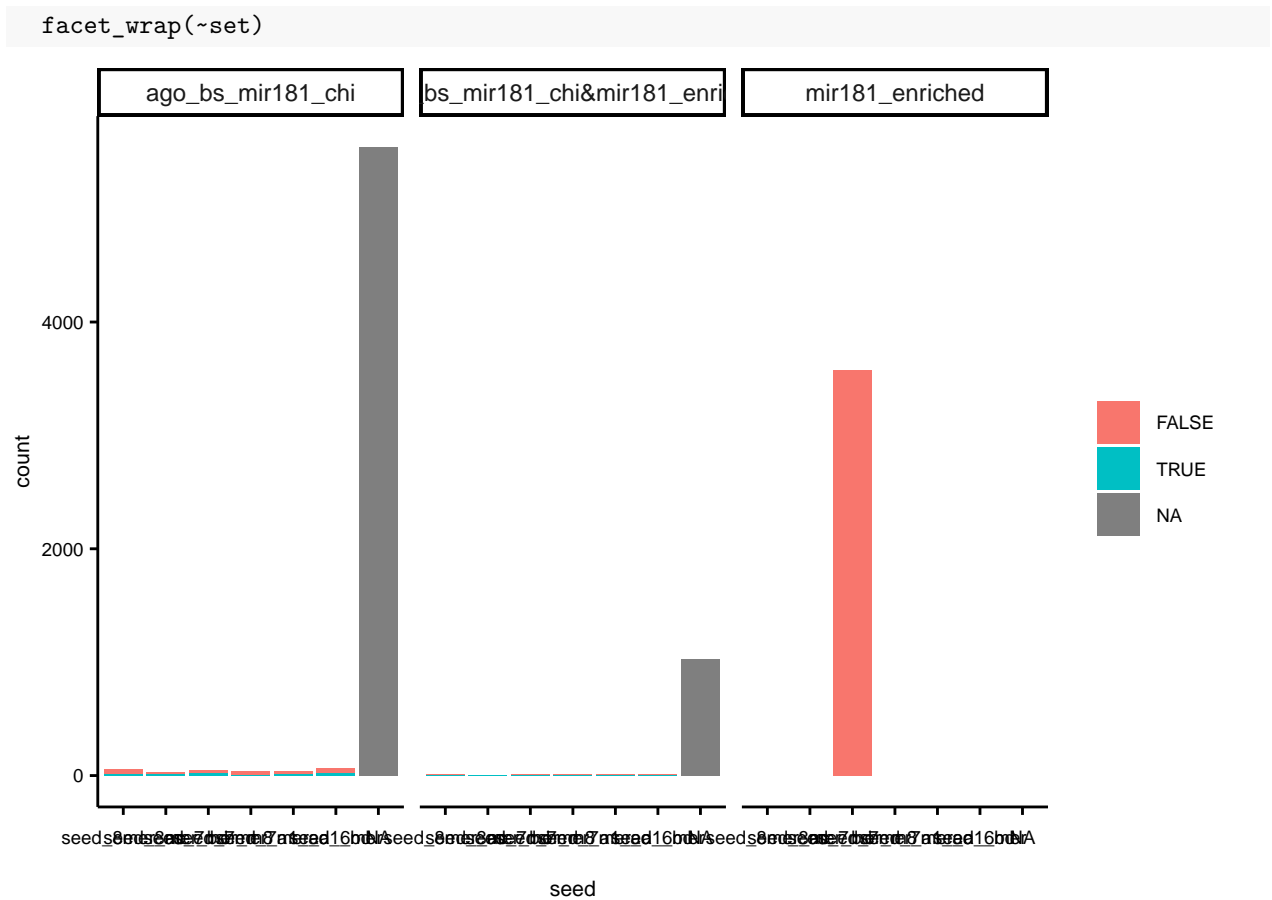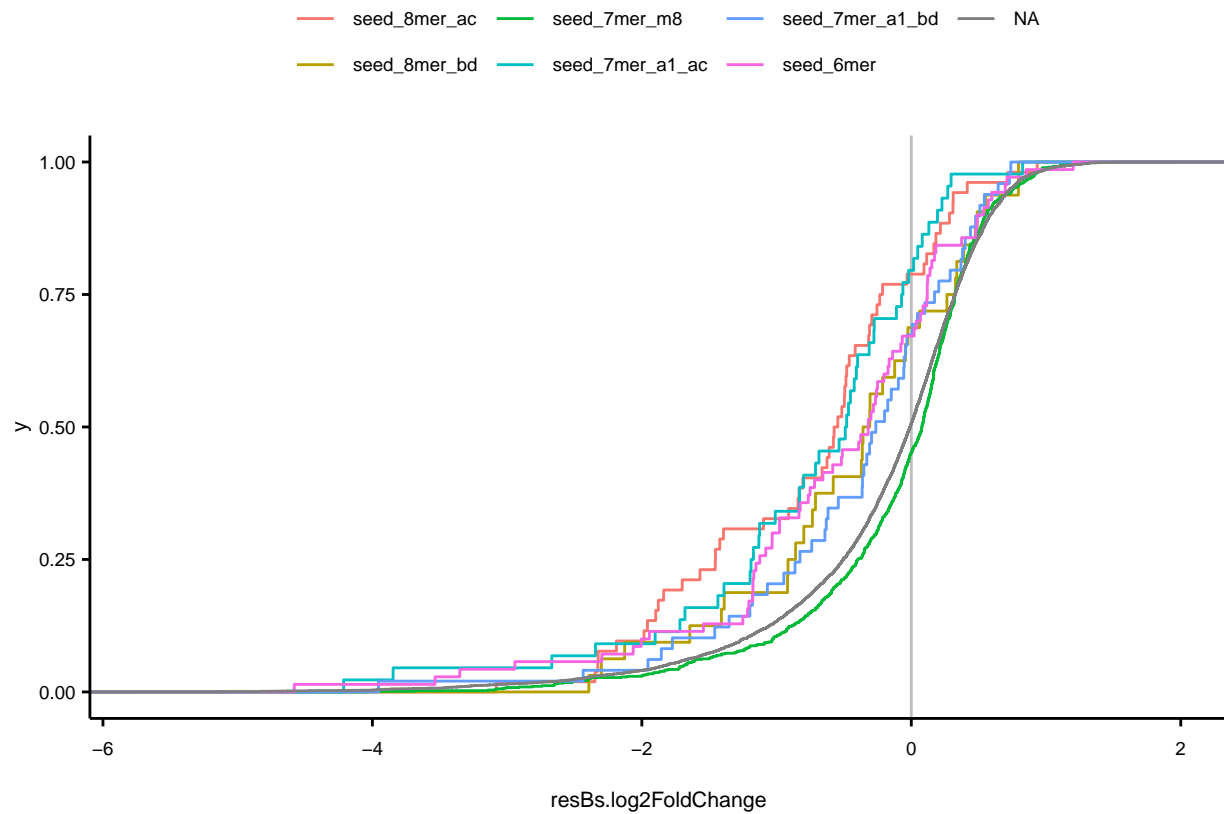
```
facet_wrap(~set)
```


6 Do different seeds lead to different activity?

```
ggplot(mir181_bs, aes(x = resBs.log2FoldChange, color = seed))+
  geom_vline(xintercept = 0, color = "grey")+
  stat_ecdf()+
  #scale_color_manual(values = c(farbe1, farbe14, farbe2, farbeneg))+
  theme_paper()+
  theme(legend.position = "top")
```

```
ggplot(mir181_bs, aes(x = resBs.log2FoldChange, color = seed))+
  geom_vline(xintercept = 0, color = "grey")+
  stat_ecdf()+
  #scale_color_manual(values = c(farbe1, farbe14, farbe2, farbeneg))+
  theme_paper()+
  theme(legend.position = "top")+
  facet_wrap(~wobble)
```
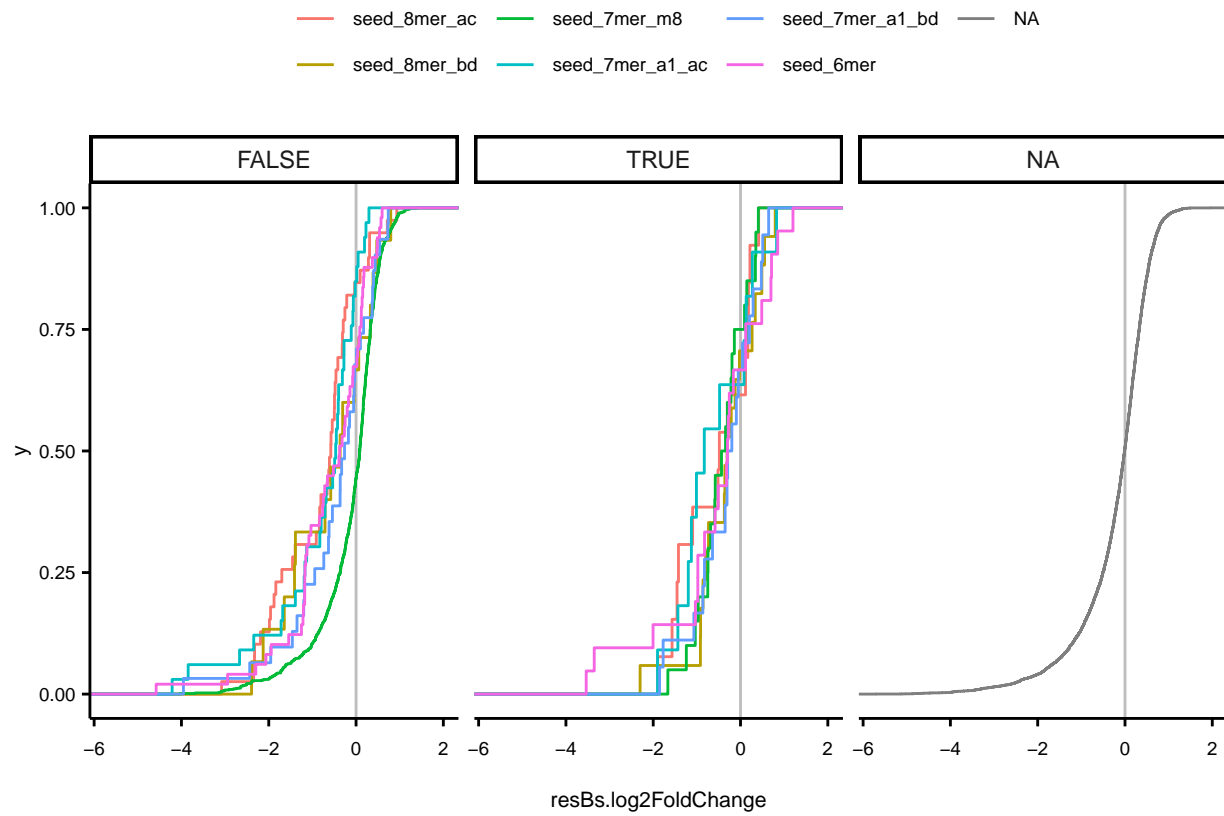
6.1 Save 50nt downstream for XSTREME search

```
writeXStringSet(mir181_bs_50down_seq, filepath = paste0(out, "mirBS_50downstream.fasta" ))
```