Assign mir181 binding sites to a specific transcript

Melina Klostermann

05 July, 2023

Contents

1 Libraries and settings

```
# ----------------------------------------
# libraries
# ----------------------------------------
library(tidyverse)
library(GenomicRanges)


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

2 What was done?

- The main expressed transcript isoform (as defined by APRIS) of each mir181 binding site is obtained
- Then the mir181 binding sites are mapped to their respective transcript
- The transcript annotations are later used for motiv discovery and structure predictions

```
#-------------------------------------
# Files
#-------------------------------------
anno <- readRDS("/Users/melinaklostermann/Documents/projects/AgoCLIP_miR181/R_github/miR181_paper/Metho

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

```
# get appris transcripts (when there are multiple take the longest)
transcripts <- anno[anno$type=="transcript"] %>% as.data.frame(.)

transcripts$transcript_id <- sub("\\..*", "", transcripts$transcript_id)

transcripts_appris <- transcripts[grepl(transcripts$tag, pattern= "appris_principal_1"),] %>%
  group_by(geneID) %>%
  arrange(desc(width), .by_group = T) %>%
```

```
  dplyr::slice(1)


# add transcript id to binding sites
transcripts_appris <- makeGRangesFromDataFrame(transcripts_appris, keep.extra.columns = T)
mir181_bs <- makeGRangesFromDataFrame(mir181_bs, keep.extra.columns = T)

idx <- findOverlaps(mir181_bs, transcripts_appris)

transcripts_appris <- as.data.frame(transcripts_appris) %>%
  select(seqnames, start, end, width, strand, geneID, transcript_id)

colnames(transcripts_appris) <- paste0(colnames(transcripts_appris), "_tx")

mir181_bs_appris <- as.data.frame(mir181_bs)

mir181_bs_appris <- cbind(mir181_bs_appris[queryHits(idx),], transcripts_appris[subjectHits(idx),])

# get mir181 bs position relativ to transcript
# (start of transcript is 1, strand is always +)

rel_mir181_bs_appris_p <- mir181_bs_appris %>%
  subset(strand == "+") %>%
  rowwise(.) %>%
  mutate(start = start - start_tx,
         end = end - start_tx)

rel_mir181_bs_appris_m <- mir181_bs_appris %>%
  subset(strand == "-") %>%
  rowwise(.) %>%
  mutate(start_genomic = start,
         start = abs(end - end_tx),
         end = abs(start_genomic - end_tx),
         strand = "+")

rel_mir181_bs_appris_m$start_genomic <- NULL

rel_mir181_bs_appris <- rbind(rel_mir181_bs_appris_p, rel_mir181_bs_appris_m)
```

```
saveRDS(rel_mir181_bs_appris, paste0(out,"mir181_bs_on_transcripts.rds"))
```