

RNAhybrid fig 2-25 elite set

Nikita Verheyden

2023-04-17

directory

```
setwd("D:/Krueger_Lab/Publications/miR181_paper/Figure2/RNAhybrid")
```

```
source("D:/Krueger_Lab/Publications/miR181_paper_v21022023/Figure_theme/theme_paper.R")  
library(BSgenome.Mmusculus.UCSC.mm10)
```

```
## Loading required package: BSgenome
```

```
## Loading required package: BiocGenerics
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
##      table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##
```

```
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':
```

```
##
```

```
##      windows
```

```
## Loading required package: GenomeInfoDb
```

```
## Loading required package: GenomicRanges
```

```

## Loading required package: Biostrings
## Loading required package: XVector
##
## Attaching package: 'Biostrings'
## The following object is masked from 'package:base':
##
##     strsplit
## Loading required package: rtracklayer
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:Biostrings':
##
##     collapse, intersect, setdiff, setequal, union
## The following object is masked from 'package:XVector':
##
##     slice
## The following objects are masked from 'package:GenomicRanges':
##
##     intersect, setdiff, union
## The following object is masked from 'package:GenomeInfoDb':
##
##     intersect
## The following objects are masked from 'package:IRanges':
##
##     collapse, desc, intersect, setdiff, slice, union
## The following objects are masked from 'package:S4Vectors':
##
##     first, intersect, rename, setdiff, setequal, union
## The following objects are masked from 'package:BiocGenerics':
##
##     combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
library(ggplot2)
library(circlize)

## =====
## circlize version 0.4.15
## CRAN page: https://cran.r-project.org/package=circlize
## Github page: https://github.com/jokergoo/circlize
## Documentation: https://jokergoo.github.io/circlize\_book/book/

```

```

##
## If you use it in published research, please cite:
## Gu, Z. circlize implements and enhances circular visualization
##   in R. Bioinformatics 2014.
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(circlize))
## =====
library(ComplexHeatmap)

## Loading required package: grid

##
## Attaching package: 'grid'

## The following object is masked from 'package:Biostrings':
##
##   pattern
## =====
## ComplexHeatmap version 2.15.2
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite either one:
## - Gu, Z. Complex Heatmap Visualization. iMeta 2022.
## - Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##   genomic data. Bioinformatics 2016.
##
##
## The new InteractiveComplexHeatmap package can directly export static
## complex heatmaps into an interactive Shiny app with zero effort. Have a try!
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(ComplexHeatmap))
## =====
library(seqinr)

##
## Attaching package: 'seqinr'

## The following object is masked from 'package:dplyr':
##
##   count

## The following object is masked from 'package:Biostrings':
##
##   translate
library(GenomicRanges)
library(rGADEM)

## Loading required package: seqLogo

##
## Attaching package: 'seqLogo'

```

```
## The following object is masked from 'package:seqinr':
##
##      consensus
##
## Attaching package: 'rGADEM'
##
## The following object is masked from 'package:seqLogo':
##
##      consensus
##
## The following object is masked from 'package:seqinr':
##
##      consensus
mir181bs <- makeGRangesFromDataFrame(readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure1/mir181_b
names(mir181bs) <- 1:length(mir181bs)

#colours
farbeneg <- "#b4b4b4"
farbe1 <- "#0073C2FF"
farbe2 <- "#EFC000FF"
farbe3 <- "#CD534CFF"
farbe4 <- "#7AA6DCFF"
farbe5 <- "#868686FF"
farbe6 <- "#003C67FF"
farbe7 <- "#8F7700FF"
farbe8 <- "#3B3B3BFF"
farbe9 <- "#A73030FF"
farbe10 <- "#4A6990FF"
farbe11 <- "#FF6F00FF"
farbe12 <- "#C71000FF"
farbe13 <- "#008EAOFF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"

RNApcol <- "#b56504"
RNAncol <- "#027d73"
RPFpcol <- "#c4c404"
RPFncol <- "#8d0391"
```

only use elite set that overlaps between ago and mir181 enriched

```
table(mir181bs$set)

##
##      ago_bs_mir181_chi ago_bs_mir181_chi&mir181_enriched
##              7117              1318
##      mir181_enriched
##              5411
mir181bs <- mir181bs[mir181bs$set=="ago_bs_mir181_chi&mir181_enriched",]
head(mir181bs)
```

```
## GRanges object with 6 ranges and 56 metadata columns:
```

##	seqnames	ranges	strand	scoreSum	scoreMean	scoreMax
##	<Rle>	<IRanges>	<Rle>	<numeric>	<numeric>	<numeric>
##	12529	chr1 6248277-6248283	+	29.0564	14.52818	22.20420
##	12530	chr1 6248959-6248965	+	35.1771	17.58855	24.10120
##	12531	chr1 59704254-59704260	+	42.8133	21.40665	31.15410
##	12532	chr1 60333013-60333019	+	25.1495	8.38316	12.80280
##	12533	chr1 60338110-60338116	+	23.9316	7.97721	9.56784
##	12534	chr1 60770687-60770693	+	115.4605	23.09209	57.46090
##	geneType	geneName	geneID	region	BS_ID	
##	<character>	<character>	<character>	<character>	<character>	
##	12529	protein_coding	Rb1cc1	ENSMUSG00000025907	cds	7
##	12530	protein_coding	Rb1cc1	ENSMUSG00000025907	cds	12
##	12531	protein_coding	Nop58	ENSMUSG00000026020	intron	142
##	12532	protein_coding	Nbeal1	ENSMUSG00000073664	utr3	147
##	12533	protein_coding	Nbeal1	ENSMUSG00000073664	utr3	148
##	12534	protein_coding	Cd28	ENSMUSG00000026012	utr3	153
##	mir_IP	n_mir181	n_mir181a	n_mir181b	n_mir181c	n_mir181d
##	<character>	<integer>	<integer>	<integer>	<integer>	<integer>
##	12529	mmu-miR-181b-5p	10	3	7	0
##	12530	mmu-miR-181a-5p	16	15	1	0
##	12531	mmu-miR-181a-5p	584	530	40	3
##	12532	mmu-miR-181a-5p	28	25	3	0
##	12533	mmu-miR-181a-5p	10	9	0	1
##	12534	mmu-miR-181a-5p	1	1	0	0
##	set	WT	KO	geneID.2		
##	<character>	<integer>	<integer>	<character>		
##	12529	ago_bs_mir181_chi&mi..	1	1	ENSMUSG00000025907	
##	12530	ago_bs_mir181_chi&mi..	1	1	ENSMUSG00000025907	
##	12531	ago_bs_mir181_chi&mi..	<NA>	<NA>	<NA>	
##	12532	ago_bs_mir181_chi&mi..	<NA>	<NA>	<NA>	
##	12533	ago_bs_mir181_chi&mi..	<NA>	<NA>	<NA>	
##	12534	ago_bs_mir181_chi&mi..	1	1	ENSMUSG00000026012	
##	geneName.1	region.1	counts.bs.1_KO	counts.bs.2_KO	counts.bs.3_KO	
##	<character>	<character>	<numeric>	<numeric>	<numeric>	
##	12529	Rb1cc1	cds	12	16	9
##	12530	Rb1cc1	cds	8	11	5
##	12531	<NA>	<NA>	NA	NA	NA
##	12532	<NA>	<NA>	NA	NA	NA
##	12533	<NA>	<NA>	NA	NA	NA
##	12534	Cd28	utr3	46	35	24
##	counts.bs.4_WT	counts.bs.5_WT	counts.bs.6_WT	geneID.1		
##	<numeric>	<numeric>	<numeric>	<character>		
##	12529	20	14	9	ENSMUSG00000025907	
##	12530	16	18	7	ENSMUSG00000025907	
##	12531	NA	NA	NA	<NA>	
##	12532	NA	NA	NA	<NA>	
##	12533	NA	NA	NA	<NA>	
##	12534	42	39	18	ENSMUSG00000026012	
##	counts.bg.1_KO	counts.bg.2_KO	counts.bg.3_KO	counts.bg.4_WT		
##	<numeric>	<numeric>	<numeric>	<numeric>		
##	12529	1609	1973	1250	2638	
##	12530	1609	1973	1250	2638	
##	12531	NA	NA	NA	NA	
##	12532	NA	NA	NA	NA	

##	12533	NA	NA	NA	NA
##	12534	552	714	444	654
##		counts.bg.5_WT	counts.bg.6_WT	resBs.baseMean	resBs.log2FoldChange
##		<numeric>	<numeric>	<numeric>	<numeric>
##	12529	2231	1352	152.021	0.352353
##	12530	2231	1352	132.183	-0.203986
##	12531	NA	NA	NA	NA
##	12532	NA	NA	NA	NA
##	12533	NA	NA	NA	NA
##	12534	641	299	251.718	0.174479
##		resBs.lfcSE	resBs.stat	resBs.pvalue	resBs.padj
##		<numeric>	<numeric>	<numeric>	<numeric>
##	12529	0.341807	1.055441	0.304257	0.725722
##	12530	0.387235	0.280266	0.596527	0.889407
##	12531	NA	NA	NA	NA
##	12532	NA	NA	NA	NA
##	12533	NA	NA	NA	NA
##	12534	0.240288	0.526789	0.467960	0.830004
##		resBg.log2FoldChange	resBg.lfcSE	resBg.stat	resBg.pvalue
##		<numeric>	<numeric>	<numeric>	<numeric>
##	12529	NA	NA	NA	NA
##	12530	NA	NA	NA	NA
##	12531	NA	NA	NA	NA
##	12532	NA	NA	NA	NA
##	12533	NA	NA	NA	NA
##	12534	NA	NA	NA	NA
##		tpm.counts.bg.1_KO	tpm.counts.bg.2_KO	tpm.counts.bg.3_KO	
##		<numeric>	<numeric>	<numeric>	
##	12529	133.726	117.998	129.867	
##	12530	133.726	117.998	129.867	
##	12531	NA	NA	NA	
##	12532	NA	NA	NA	
##	12533	NA	NA	NA	
##	12534	110.085	102.465	110.688	
##		tpm.counts.bg.4_WT	tpm.counts.bg.5_WT	tpm.counts.bg.6_WT	
##		<numeric>	<numeric>	<numeric>	
##	12529	139.8635	146.286	163.5360	
##	12530	139.8635	146.286	163.5360	
##	12531	NA	NA	NA	
##	12532	NA	NA	NA	
##	12533	NA	NA	NA	
##	12534	83.2024	100.853	86.7834	
##		BS_ID.1	tpm_support_KO	tpm_support_WT	tpm_supported
##		<character>	<integer>	<integer>	<logical>
##	12529	ENSMUSG00000025907.bs7	3	3	TRUE
##	12530	ENSMUSG00000025907.b..	3	3	TRUE
##	12531	<NA>	<NA>	<NA>	<NA>
##	12532	<NA>	<NA>	<NA>	<NA>
##	12533	<NA>	<NA>	<NA>	<NA>
##	12534	ENSMUSG00000026012.bs4	3	3	TRUE
##		down			
##		<logical>			
##	12529	FALSE			
##	12530	FALSE			

```
## 12531 <NA>
## 12532 <NA>
## 12533 <NA>
## 12534 FALSE
## -----
## seqinfo: 22 sequences from an unspecified genome; no seqlengths
```

Get DNA sequences

```
#resize ranges
```

```
mir181bs <- resize(mir181bs, width = width(mir181bs + 25), fix = "center")
mir181bs <- resize(mir181bs, width = width(mir181bs + 25), fix = "start")
```

```
# bsGene <- resize(bsGene, width = width(bsGene + 5), fix = "end")
```

```
df181 <- mutate(as.data.frame(mir181bs), Sequence = as.character(getSeq(BSgenome.Mmusculus.UCSC.mm10, s
```

```
#and turn T into Us
```

```
df181$Sequence <- gsub('T', 'U', df181$Sequence)
```

```
head(df181)
```

```
##      seqnames      start      end width strand  scoreSum scoreMean scoreMax
## 12529      chr1 6248252 6248358   107      + 29.05635 14.528175 22.20420
## 12530      chr1 6248934 6249040   107      + 35.17710 17.588550 24.10120
## 12531      chr1 59704229 59704335   107      + 42.81330 21.406650 31.15410
## 12532      chr1 60332988 60333094   107      + 25.14947  8.383157 12.80280
## 12533      chr1 60338085 60338191   107      + 23.93163  7.977210  9.56784
## 12534      chr1 60770662 60770768   107      + 115.46047 23.092094 57.46090
##      geneType geneName      geneID region BS_ID      mir_IP
## 12529 protein_coding Rb1cc1 ENSMUSG00000025907  cds        7 mmu-miR-181b-5p
## 12530 protein_coding Rb1cc1 ENSMUSG00000025907  cds       12 mmu-miR-181a-5p
## 12531 protein_coding  Nop58 ENSMUSG00000026020 intron    142 mmu-miR-181a-5p
## 12532 protein_coding Nbeal1 ENSMUSG00000073664  utr3    147 mmu-miR-181a-5p
## 12533 protein_coding Nbeal1 ENSMUSG00000073664  utr3    148 mmu-miR-181a-5p
## 12534 protein_coding  Cd28 ENSMUSG00000026012  utr3    153 mmu-miR-181a-5p
##      n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
## 12529      10         3         7         0         0
## 12530      16        15         1         0         0
## 12531     584       530        40         3        11
## 12532      28        25         3         0         0
## 12533      10         9         0         1         0
## 12534       1         1         0         0         0
##      set WT KO      geneID.2 geneName.1
## 12529 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000025907      Rb1cc1
## 12530 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000025907      Rb1cc1
## 12531 ago_bs_mir181_chi&mir181_enriched NA NA          <NA>          <NA>
## 12532 ago_bs_mir181_chi&mir181_enriched NA NA          <NA>          <NA>
## 12533 ago_bs_mir181_chi&mir181_enriched NA NA          <NA>          <NA>
## 12534 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000026012      Cd28
##      region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO counts.bs.4_WT
```

##	12529	cds	12	16	9	20
##	12530	cds	8	11	5	16
##	12531	<NA>	NA	NA	NA	NA
##	12532	<NA>	NA	NA	NA	NA
##	12533	<NA>	NA	NA	NA	NA
##	12534	utr3	46	35	24	42
##		counts.bs.5_WT	counts.bs.6_WT	geneID.1	counts.bg.1_KO	
##	12529	14	9	ENSMUSG00000025907	1609	
##	12530	18	7	ENSMUSG00000025907	1609	
##	12531	NA	NA	<NA>	NA	
##	12532	NA	NA	<NA>	NA	
##	12533	NA	NA	<NA>	NA	
##	12534	39	18	ENSMUSG00000026012	552	
##		counts.bg.2_KO	counts.bg.3_KO	counts.bg.4_WT	counts.bg.5_WT	
##	12529	1973	1250	2638	2231	
##	12530	1973	1250	2638	2231	
##	12531	NA	NA	NA	NA	
##	12532	NA	NA	NA	NA	
##	12533	NA	NA	NA	NA	
##	12534	714	444	654	641	
##		counts.bg.6_WT	resBs.baseMean	resBs.log2FoldChange	resBs.lfcSE	resBs.stat
##	12529	1352	152.0207	0.3523528	0.3418073	1.0554415
##	12530	1352	132.1833	-0.2039860	0.3872352	0.2802662
##	12531	NA	NA	NA	NA	NA
##	12532	NA	NA	NA	NA	NA
##	12533	NA	NA	NA	NA	NA
##	12534	299	251.7179	0.1744790	0.2402881	0.5267887
##		resBs.pvalue	resBs.padj	resBg.baseMean	resBg.log2FoldChange	resBg.lfcSE
##	12529	0.3042572	0.7257221	NA	NA	NA
##	12530	0.5965268	0.8894069	NA	NA	NA
##	12531	NA	NA	NA	NA	NA
##	12532	NA	NA	NA	NA	NA
##	12533	NA	NA	NA	NA	NA
##	12534	0.4679601	0.8300039	NA	NA	NA
##		resBg.stat	resBg.pvalue	resBg.padj	tpm.counts.bg.1_KO	tpm.counts.bg.2_KO
##	12529	NA	NA	NA	133.7259	117.9980
##	12530	NA	NA	NA	133.7259	117.9980
##	12531	NA	NA	NA	NA	NA
##	12532	NA	NA	NA	NA	NA
##	12533	NA	NA	NA	NA	NA
##	12534	NA	NA	NA	110.0847	102.4647
##		tpm.counts.bg.3_KO	tpm.counts.bg.4_WT	tpm.counts.bg.5_WT		
##	12529	129.8669	139.86348	146.2855		
##	12530	129.8669	139.86348	146.2855		
##	12531	NA	NA	NA		
##	12532	NA	NA	NA		
##	12533	NA	NA	NA		
##	12534	110.6878	83.20238	100.8529		
##		tpm.counts.bg.6_WT	BS_ID.1	tpm_support_KO	tpm_support_WT	
##	12529	163.53603	ENSMUSG00000025907.bs7	3	3	
##	12530	163.53603	ENSMUSG00000025907.bs12	3	3	
##	12531	NA	<NA>	NA	NA	
##	12532	NA	<NA>	NA	NA	
##	12533	NA	<NA>	NA	NA	


```
## 12534      86.78335  ENSMUSG00000026012.bs4      3      3
##      tpm_supported  down
## 12529      TRUE FALSE
## 12530      TRUE FALSE
## 12531      NA      NA
## 12532      NA      NA
## 12533      NA      NA
## 12534      TRUE FALSE
##
## 12529 UAAGACUUUAACUCUGUGUUUCAGGUAUCCACAAGUCAGGCAUCCCCACAGUCAGCUGCUUCUCCAAGAAUAGAAAGUACAACAGGCAUUACAAG
## 12530 AGGAAAAGCAUCAGCAAGAACUCCAAUCUUUAAAAAUAGAGUAUGAAUGUAAACUUGAUGCUCUAGUAAAAGACAGUGAAGAAAAUGUAAAUAA
## 12531 GCAGUUGAGGUCUAAGUCUGAUAAUUUGCAGGUGGCAAUGAUGAUUUUUUCUUAUUUGUUCACCUGACAAUAUAUGAAGGUGUUCAGUCACUACCU
## 12532 UGAAAAAUAGUGAUUUUGGAGUAGGGAUAUCACCCGACAGUUGAAUGCUUGCCUUCAUUGUUCAAGACUCUGGGUUUGCUCCCUCGCAUGGGUG
## 12533 AAGAUAUAUUUUUAUUUUGUCACAUAAAAUAUUUUUAAGAACUUCACUGAAUGUAUAUGAGUUAUUGUUUGAUUUUGUUCUGUAUAGUACGAUCA
## 12534 CUGUUUAGUAGCCUGUUUAUGUCAAAUCAAUUUAAGUUUCAACUUCUACAGUUGGCCACUUGUUGUCCUUUGUGGAUGGAUAUCUGAAAUUG
```

Write to .fasta

this is deactivated for now because we only need it once right now just remove the eval if needed

```
candgeneName <- as.list(df181$geneName)
candname <- as.list(rownames(df181))
condgeneSeq <- as.list(df181$Sequence)

#change to output directory
setwd("D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/fastafiles-25_elite")

for (i in 1:length(candgeneName)) {
  write.fasta(condgeneSeq[i],candname[i],paste(candname[i], candgeneName[i], "miR_181", 'fasta', sep = "_"))
}
```

Import RNAhybrid results

files that are imported here were run with maximum loop of 3 in both sequences

```
Personalized_Reader <- function(lambda){
  read.table(lambda, sep = ":") %>% select(V1, V5, V6, V7, V9)}

reslistA <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/res")
reslistB <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/res")

myfilelistA <- lapply(reslistA, Personalized_Reader)
myfilelistB <- lapply(reslistB, Personalized_Reader)

resframeA <- bind_rows(myfilelistA)
resframeB <- bind_rows(myfilelistB)

colnames(resframeA) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases")
colnames(resframeB) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases")

head(resframeA)
```

```
##   rownumber   mfs   pvalue start_position      binding_bases
## 1    12529 -17.9 0.911034         10      ACUC  UG   CAG GU
## 2    12530 -21.5 0.243804         59    GCUC   A   GACAGUG GA  AUGU
## 3    12531 -19.9 0.517782         57      UUCACC GACA U   UGA  GUGUU
## 4    12532 -27.7 0.006767         29              UCA CCGACAGU UGAAUG
## 5    12533 -19.0 0.713812         41    UUCACUGA  UGU   UGA   AUGUU
## 6    12534 -16.1 0.999191         41     ACUU CU ACAGU  UGGG   UGUU
```

```
head(resframeB)
```

```
##   rownumber   mfs   pvalue start_position      binding_bases
## 1    12529 -17.9 0.911034         10      ACUC  UG   CAG GU
## 2    12530 -21.5 0.243804         59    GCUC   A   GACAGUG GA  AUGU
## 3    12531 -19.9 0.517782         57      UUCACC GACA U   UGA  GUGUU
## 4    12532 -27.7 0.006767         29              UCA CCGACAGU UGAAUG
## 5    12533 -19.0 0.713812         41    UUCACUGA  UGU   UGA   AUGUU
## 6    12534 -16.1 0.999191         41     ACUU CU ACAGU  UGGG   UGUU
```

merge with original df

```
# make seperate objects for each mature mirna just to see if they are much different
```

```
df181$rownumber <- as.character(rownames(df181))
resframeA$rownumber <- as.character(resframeA$rownumber)
resframeB$rownumber <- as.character(resframeB$rownumber)
```

```
bsseqHA <- left_join(df181, resframeA, by="rownumber")
bsseqHB <- left_join(df181, resframeB, by="rownumber")
```

```
head(bsseqHA)
```

```
##   seqnames      start      end width strand  scoreSum scoreMean scoreMax
## 1    chr1  6248252  6248358   107      +  29.05635  14.528175  22.20420
## 2    chr1  6248934  6249040   107      +  35.17710  17.588550  24.10120
## 3    chr1  59704229 59704335   107      +  42.81330  21.406650  31.15410
## 4    chr1  60332988 60333094   107      +  25.14947   8.383157  12.80280
## 5    chr1  60338085 60338191   107      +  23.93163   7.977210   9.56784
## 6    chr1  60770662 60770768   107      + 115.46047  23.092094  57.46090
##      geneType geneName      geneID region BS_ID      mir_IP
## 1 protein_coding Rb1cc1 ENSMUSG00000025907  cds      7 mmu-miR-181b-5p
## 2 protein_coding Rb1cc1 ENSMUSG00000025907  cds     12 mmu-miR-181a-5p
## 3 protein_coding  Nop58 ENSMUSG00000026020 intron   142 mmu-miR-181a-5p
## 4 protein_coding Nbeal1 ENSMUSG00000073664  utr3   147 mmu-miR-181a-5p
## 5 protein_coding Nbeal1 ENSMUSG00000073664  utr3   148 mmu-miR-181a-5p
## 6 protein_coding  Cd28 ENSMUSG00000026012  utr3   153 mmu-miR-181a-5p
##   n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
## 1      10         3         7         0         0
## 2      16        15         1         0         0
## 3     584       530       40         3        11
## 4      28        25         3         0         0
## 5      10         9         0         1         0
## 6       1         1         0         0         0
##                                     set WT KO      geneID.2 geneName.1
```

```

## 1 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000025907 Rb1cc1
## 2 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000025907 Rb1cc1
## 3 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 4 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 5 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 6 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000026012 Cd28
## region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO counts.bs.4_WT
## 1 cds 12 16 9 20
## 2 cds 8 11 5 16
## 3 <NA> NA NA NA NA
## 4 <NA> NA NA NA NA
## 5 <NA> NA NA NA NA
## 6 utr3 46 35 24 42
## counts.bs.5_WT counts.bs.6_WT geneID.1 counts.bg.1_KO
## 1 14 9 ENSMUSG00000025907 1609
## 2 18 7 ENSMUSG00000025907 1609
## 3 NA NA <NA> NA
## 4 NA NA <NA> NA
## 5 NA NA <NA> NA
## 6 39 18 ENSMUSG00000026012 552
## counts.bg.2_KO counts.bg.3_KO counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT
## 1 1973 1250 2638 2231 1352
## 2 1973 1250 2638 2231 1352
## 3 NA NA NA NA NA
## 4 NA NA NA NA NA
## 5 NA NA NA NA NA
## 6 714 444 654 641 299
## resBs.baseMean resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue
## 1 152.0207 0.3523528 0.3418073 1.0554415 0.3042572
## 2 132.1833 -0.2039860 0.3872352 0.2802662 0.5965268
## 3 NA NA NA NA NA
## 4 NA NA NA NA NA
## 5 NA NA NA NA NA
## 6 251.7179 0.1744790 0.2402881 0.5267887 0.4679601
## resBs.padj resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat
## 1 0.7257221 NA NA NA NA
## 2 0.8894069 NA NA NA NA
## 3 NA NA NA NA NA
## 4 NA NA NA NA NA
## 5 NA NA NA NA NA
## 6 0.8300039 NA NA NA NA
## resBg.pvalue resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO
## 1 NA NA 133.7259 117.9980
## 2 NA NA 133.7259 117.9980
## 3 NA NA NA NA
## 4 NA NA NA NA
## 5 NA NA NA NA
## 6 NA NA 110.0847 102.4647
## tpm.counts.bg.3_KO tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1 129.8669 139.86348 146.2855 163.53603
## 2 129.8669 139.86348 146.2855 163.53603
## 3 NA NA NA NA
## 4 NA NA NA NA
## 5 NA NA NA NA

```

```
## 6          110.6878          83.20238          100.8529          86.78335
##          BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported down
## 1 ENSMUSG00000025907.bs7          3          3          TRUE FALSE
## 2 ENSMUSG00000025907.bs12         3          3          TRUE FALSE
## 3          <NA>          NA          NA          NA    NA
## 4          <NA>          NA          NA          NA    NA
## 5          <NA>          NA          NA          NA    NA
## 6 ENSMUSG00000026012.bs4          3          3          TRUE FALSE
##
## 1 UAAGACUUUAACUCUGUGUUUCAGGUAUCCACAAGUCAGGCAUCCCCACAGUCAGCUGCUUCUCCAAGAAUAGAAAGUACAACAGGCAUUACAACCAC
## 2 AGGAAAAGCAUCAGCAAGAACUCCAAUCUUUAAAAAUAGAGUAUGAAUGUAAACUUGAUGCUCUAGUAAAAAGACAGUGAAGAAAAUGUAAAUAUUUU
## 3 GCAGUUGAGGUCUAAGUCUGAUAAUUUGCAGGUGGCAAUGAUGAUUUUUUCUUUUUUGUUCACCUGACAAUAUAUGAAGGUGUUCAGUCACUACCUCAU
## 4 UGAAAAAUAGUGAUUUUGGAGUAGGGAUAUACCCGACAGUUGAAUGCUUGCCUUCAUUGUUAAGACUCUGGGUUUGCUCCUCGCAUGGGUGCGAG
## 5 AAGAUUAUUUUUUUUUGUCACAUAAAAUUUUUUUAAGAACUUCACUGAAUGUAUAUGAGUUAAUGUUUGAUUUUGUUAUAGUACGAUAAAAA
## 6 CUGUUUAGUAGCCUGUUUAUGUCAAAUCAUUUUUAAGUUUCAACUUCUACAGUUGGGCCACUUGUUGCCUUGUGGAUGGAUAUCUGAAAUUGUGUC
## rownumber mfs pvalue start_position binding_bases
## 1 12529 -17.9 0.911034 10 ACUC UG CAG GU
## 2 12530 -21.5 0.243804 59 GCUC A GACAGUG GA AUGU
## 3 12531 -19.9 0.517782 57 UUCACC GACA U UGA GUGUU
## 4 12532 -27.7 0.006767 29 UCA CCGACAGU UGAAUG
## 5 12533 -19.0 0.713812 41 UUCACUGA UGU UGA AUGUU
## 6 12534 -16.1 0.999191 41 ACUU CU ACAGU UGGG UGUU
```

```
head(bsseqHB)
```

```
## seqnames start end width strand scoreSum scoreMean scoreMax
## 1 chr1 6248252 6248358 107 + 29.05635 14.528175 22.20420
## 2 chr1 6248934 6249040 107 + 35.17710 17.588550 24.10120
## 3 chr1 59704229 59704335 107 + 42.81330 21.406650 31.15410
## 4 chr1 60332988 60333094 107 + 25.14947 8.383157 12.80280
## 5 chr1 60338085 60338191 107 + 23.93163 7.977210 9.56784
## 6 chr1 60770662 60770768 107 + 115.46047 23.092094 57.46090
## geneType geneName geneID region BS_ID mir_IP
## 1 protein_coding Rb1cc1 ENSMUSG00000025907 cds 7 mmu-miR-181b-5p
## 2 protein_coding Rb1cc1 ENSMUSG00000025907 cds 12 mmu-miR-181a-5p
## 3 protein_coding Nop58 ENSMUSG00000026020 intron 142 mmu-miR-181a-5p
## 4 protein_coding Nbeal1 ENSMUSG00000073664 utr3 147 mmu-miR-181a-5p
## 5 protein_coding Nbeal1 ENSMUSG00000073664 utr3 148 mmu-miR-181a-5p
## 6 protein_coding Cd28 ENSMUSG00000026012 utr3 153 mmu-miR-181a-5p
## n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
## 1 10 3 7 0 0
## 2 16 15 1 0 0
## 3 584 530 40 3 11
## 4 28 25 3 0 0
## 5 10 9 0 1 0
## 6 1 1 0 0 0
## set WT KO geneID.2 geneName.1
## 1 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000025907 Rb1cc1
## 2 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000025907 Rb1cc1
## 3 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 4 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 5 ago_bs_mir181_chi&mir181_enriched NA NA <NA> <NA>
## 6 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG00000026012 Cd28
## region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO counts.bs.4_WT
## 1 cds 12 16 9 20
```

## 2	cds	8	11	5	16
## 3	<NA>	NA	NA	NA	NA
## 4	<NA>	NA	NA	NA	NA
## 5	<NA>	NA	NA	NA	NA
## 6	utr3	46	35	24	42
##	counts.bs.5_WT	counts.bs.6_WT	geneID.1	counts.bg.1_KO	
## 1	14	9	ENSMUSG00000025907	1609	
## 2	18	7	ENSMUSG00000025907	1609	
## 3	NA	NA	<NA>	NA	
## 4	NA	NA	<NA>	NA	
## 5	NA	NA	<NA>	NA	
## 6	39	18	ENSMUSG00000026012	552	
##	counts.bg.2_KO	counts.bg.3_KO	counts.bg.4_WT	counts.bg.5_WT	counts.bg.6_WT
## 1	1973	1250	2638	2231	1352
## 2	1973	1250	2638	2231	1352
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	714	444	654	641	299
##	resBs.baseMean	resBs.log2FoldChange	resBs.lfcSE	resBs.stat	resBs.pvalue
## 1	152.0207	0.3523528	0.3418073	1.0554415	0.3042572
## 2	132.1833	-0.2039860	0.3872352	0.2802662	0.5965268
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	251.7179	0.1744790	0.2402881	0.5267887	0.4679601
##	resBs.padj	resBg.baseMean	resBg.log2FoldChange	resBg.lfcSE	resBg.stat
## 1	0.7257221	NA	NA	NA	NA
## 2	0.8894069	NA	NA	NA	NA
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	0.8300039	NA	NA	NA	NA
##	resBg.pvalue	resBg.padj	tpm.counts.bg.1_KO	tpm.counts.bg.2_KO	
## 1	NA	NA	133.7259	117.9980	
## 2	NA	NA	133.7259	117.9980	
## 3	NA	NA	NA	NA	
## 4	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	
## 6	NA	NA	110.0847	102.4647	
##	tpm.counts.bg.3_KO	tpm.counts.bg.4_WT	tpm.counts.bg.5_WT	tpm.counts.bg.6_WT	
## 1	129.8669	139.86348	146.2855	163.53603	
## 2	129.8669	139.86348	146.2855	163.53603	
## 3	NA	NA	NA	NA	
## 4	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	
## 6	110.6878	83.20238	100.8529	86.78335	
##	BS_ID.1	tpm_support_KO	tpm_support_WT	tpm_supported	down
## 1	ENSMUSG00000025907.bs7	3	3	TRUE	FALSE
## 2	ENSMUSG00000025907.bs12	3	3	TRUE	FALSE
## 3	<NA>	NA	NA	NA	NA
## 4	<NA>	NA	NA	NA	NA
## 5	<NA>	NA	NA	NA	NA
## 6	ENSMUSG00000026012.bs4	3	3	TRUE	FALSE

```
##
## 1 UAAGACUUUAACUCUGUGUUUCAGGUAUCCACAAGUCAGGCAUCCCCACAGUCAGCUGCUUCUCCAAGAAUAGAAAGUACAACAGGCAUUACAACCAC
## 2 AGGAAAAGCAUCAGCAAGAACUCCAAUCUUUAAAAUUGAGUAUGAAUGUAAACUUGAUGCUCUAGUAAAAAGACAGUGAAGAAAAUGUAAAUAUUUUU
## 3 GCAGUUGAGGUCUAAGUCUGAUAAUUGCAGGUGGCAAUGAUGAUUUUUCUUAUUUGUUCACCUGACAAUAUAUGAAGGUGUUCAGUCACUACCUCAU
## 4 UGAAAAAUAGUGAUUUUGGAGUAGGGAUAUCACCCGACAGUUGAAUGCUUGCCUUCUUAUUGUUAAGACUCUGGGUUUGCUCUCCUCGCAUGGGUGCGAG
## 5 AAGAUUAUUUUUUAUUUGUCACAUAAAAUUUUUUAAGAACUUCACUGAAUGUAUAUGAGUUAAUGUUUGAUUUUGUUCUGUAUAGUACGAUCAAAAAU
## 6 CUGUUUAGUAGCCUGUUAUAGUCAAAUCAUUUUUAAGUUUCAACUUCUACAGUUGGGCCACUUGUUGUCCUUUGUGGAUGGAUAUCUGAAAUUGUGUC
##      rownumber      mfs      pvalue start_position      binding_bases
## 1      12529    -17.9  0.911034          10      ACUC  UG   CAG  GU
## 2      12530    -21.5  0.243804          59    GCUC      A   GACAGUG  GA   AUGU
## 3      12531    -19.9  0.517782          57      UUCACC GACA  U    UGA   GUGUU
## 4      12532    -27.7  0.006767          29      UCA  CCGACAGU  UGAAUG
## 5      12533    -19.0  0.713812          41    UUCACUGA  UGU   UGA   AUGUU
## 6      12534    -16.1  0.999191          41    ACUU  CU  ACAGU  UGGG      UGUU
```

adjust binding site info

```
bsseqHA$binding_bases <- chartr(" ", "N", bsseqHA$binding_bases)
bsseqHB$binding_bases <- chartr(" ", "N", bsseqHB$binding_bases)
```

```
head(bsseqHA)
```

```
##      seqnames      start      end width strand  scoreSum scoreMean scoreMax
## 1      chr1  6248252  6248358   107      +  29.05635  14.528175  22.20420
## 2      chr1  6248934  6249040   107      +  35.17710  17.588550  24.10120
## 3      chr1  59704229 59704335   107      +  42.81330  21.406650  31.15410
## 4      chr1  60332988 60333094   107      +  25.14947   8.383157  12.80280
## 5      chr1  60338085 60338191   107      +  23.93163   7.977210   9.56784
## 6      chr1  60770662 60770768   107      + 115.46047  23.092094  57.46090
##      geneType geneName      geneID region BS_ID      mir_IP
## 1 protein_coding  Rb1cc1 ENSMUSG00000025907  cds      7 mmu-miR-181b-5p
## 2 protein_coding  Rb1cc1 ENSMUSG00000025907  cds     12 mmu-miR-181a-5p
## 3 protein_coding   Nop58 ENSMUSG00000026020 intron   142 mmu-miR-181a-5p
## 4 protein_coding  Nbeal1 ENSMUSG00000073664  utr3    147 mmu-miR-181a-5p
## 5 protein_coding  Nbeal1 ENSMUSG00000073664  utr3    148 mmu-miR-181a-5p
## 6 protein_coding   Cd28 ENSMUSG00000026012  utr3    153 mmu-miR-181a-5p
##      n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
## 1      10      3      7      0      0
## 2      16     15      1      0      0
## 3     584     530     40      3     11
## 4      28     25      3      0      0
## 5      10      9      0      1      0
## 6       1      1      0      0      0
##      set WT KO      geneID.2 geneName.1
## 1 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000025907  Rb1cc1
## 2 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000025907  Rb1cc1
## 3 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 4 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 5 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 6 ago_bs_mir181_chi&mir181_enriched  1  1 ENSMUSG00000026012  Cd28
##      region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO counts.bs.4_WT
## 1      cds      12      16      9      20
## 2      cds       8      11      5      16
## 3     <NA>      NA      NA      NA      NA
```

## 4	<NA>	NA	NA	NA	NA
## 5	<NA>	NA	NA	NA	NA
## 6	utr3	46	35	24	42
##	counts.bs.5_WT	counts.bs.6_WT	geneID.1	counts.bg.1_KO	
## 1	14	9	ENSMUSG00000025907	1609	
## 2	18	7	ENSMUSG00000025907	1609	
## 3	NA	NA	<NA>	NA	
## 4	NA	NA	<NA>	NA	
## 5	NA	NA	<NA>	NA	
## 6	39	18	ENSMUSG00000026012	552	
##	counts.bg.2_KO	counts.bg.3_KO	counts.bg.4_WT	counts.bg.5_WT	counts.bg.6_WT
## 1	1973	1250	2638	2231	1352
## 2	1973	1250	2638	2231	1352
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	714	444	654	641	299
##	resBs.baseMean	resBs.log2FoldChange	resBs.lfcSE	resBs.stat	resBs.pvalue
## 1	152.0207	0.3523528	0.3418073	1.0554415	0.3042572
## 2	132.1833	-0.2039860	0.3872352	0.2802662	0.5965268
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	251.7179	0.1744790	0.2402881	0.5267887	0.4679601
##	resBs.padj	resBg.baseMean	resBg.log2FoldChange	resBg.lfcSE	resBg.stat
## 1	0.7257221	NA	NA	NA	NA
## 2	0.8894069	NA	NA	NA	NA
## 3	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA
## 6	0.8300039	NA	NA	NA	NA
##	resBg.pvalue	resBg.padj	tpm.counts.bg.1_KO	tpm.counts.bg.2_KO	
## 1	NA	NA	133.7259	117.9980	
## 2	NA	NA	133.7259	117.9980	
## 3	NA	NA	NA	NA	
## 4	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	
## 6	NA	NA	110.0847	102.4647	
##	tpm.counts.bg.3_KO	tpm.counts.bg.4_WT	tpm.counts.bg.5_WT	tpm.counts.bg.6_WT	
## 1	129.8669	139.86348	146.2855	163.53603	
## 2	129.8669	139.86348	146.2855	163.53603	
## 3	NA	NA	NA	NA	
## 4	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	
## 6	110.6878	83.20238	100.8529	86.78335	
##	BS_ID.1	tpm_support_KO	tpm_support_WT	tpm_supported	down
## 1	ENSMUSG00000025907.bs7	3	3	TRUE	FALSE
## 2	ENSMUSG00000025907.bs12	3	3	TRUE	FALSE
## 3	<NA>	NA	NA	NA	NA
## 4	<NA>	NA	NA	NA	NA
## 5	<NA>	NA	NA	NA	NA
## 6	ENSMUSG00000026012.bs4	3	3	TRUE	FALSE
##					
## 1	UAAGACUUUAACUCUGUGUUUCAGGUAUCCACAAGUCAGGCAUCCCCACAGUCAGCUGCUUCUCCAAGAAUAGAAAGUACAACAGGCAUUACAACCAC				

```
## 2 AGGAAAAGCAUCAGCAAGAACUCCAAUCUUUAAAAAUUGAGUAUGAAUGUAAACUUGAUGCUCUAGUAAAAGACAGUGAAGAAAUGUAAAUAUUUUU
## 3 GCAGUUGAGGUCUAAGUCUGAUAAUUUGCAGGUGGCAAUGAUGAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
## 4 UGAAAAAUAGUGAUUUUGGAGUAGGGAUAUCACCCGACAGUUGAAUGCUUGCCUUCUUAUUGUUCAAGACUCUGGGUUUUGCUCCCGCAUGGGUGCGAG
## 5 AAGAUAUAUUUUUUUUUUUGUCACAUAAAUAUUUUUUUAAGAACUUCACUGAAUGUAUAUGAGUUAAUGUUUGAUUUUUUUUUUUUUUUUUUUUUUUUUUU
## 6 CUGUUUAGUAGCCUGUUAUAGUCAAAUCAUUUUUAAGUUUCAACUUCUUACAGUUGGGCCACUUGUUGUCCUUUGUGGAUGGAUAUCUGAAAUUGUGUCU
##      rownumber      mfs      pvalue start_position      binding_bases
## 1      12529     -17.9    0.911034          10      NACUCNNUGNNNCAGNGUNNNNNNNN
## 2      12530     -21.5    0.243804          59      NGCUCNNNNNNANNGACAGUGNNNGANNAUGUN
## 3      12531     -19.9    0.517782          57      NUUCACCNGACANUNNNUGANNNGUGUUN
## 4      12532     -27.7    0.006767          29      NNUCANCCGACAGUNNUGAAUGNN
## 5      12533     -19.0    0.713812          41      NUUCACUGANNNUGUNNNUGANNNNNAUGUUN
## 6      12534     -16.1    0.999191          41      NACUUNCUNACAGUNNUGGGNNNNNNUGUUN
```

```
head(bsseqHB)
```

```
##      seqnames      start      end width strand      scoreSum scoreMean scoreMax
## 1      chr1    6248252    6248358    107      +    29.05635    14.528175    22.20420
## 2      chr1    6248934    6249040    107      +    35.17710    17.588550    24.10120
## 3      chr1    59704229    59704335    107      +    42.81330    21.406650    31.15410
## 4      chr1    60332988    60333094    107      +    25.14947     8.383157    12.80280
## 5      chr1    60338085    60338191    107      +    23.93163     7.977210     9.56784
## 6      chr1    60770662    60770768    107      +   115.46047    23.092094    57.46090
##      geneType geneName      geneID region BS_ID      mir_IP
## 1 protein_coding Rb1cc1 ENSMUSG000000025907      cds      7 mmu-miR-181b-5p
## 2 protein_coding Rb1cc1 ENSMUSG000000025907      cds     12 mmu-miR-181a-5p
## 3 protein_coding Nop58 ENSMUSG000000026020      intron    142 mmu-miR-181a-5p
## 4 protein_coding Nbeal1 ENSMUSG000000073664      utr3     147 mmu-miR-181a-5p
## 5 protein_coding Nbeal1 ENSMUSG000000073664      utr3     148 mmu-miR-181a-5p
## 6 protein_coding Cd28 ENSMUSG000000026012      utr3     153 mmu-miR-181a-5p
##      n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d
## 1      10      3      7      0      0
## 2      16     15      1      0      0
## 3     584     530     40      3     11
## 4      28     25      3      0      0
## 5      10      9      0      1      0
## 6       1      1      0      0      0
##      set WT KO      geneID.2 geneName.1
## 1 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG000000025907 Rb1cc1
## 2 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG000000025907 Rb1cc1
## 3 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 4 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 5 ago_bs_mir181_chi&mir181_enriched NA NA      <NA>      <NA>
## 6 ago_bs_mir181_chi&mir181_enriched 1 1 ENSMUSG000000026012 Cd28
##      region.1 counts.bs.1_KO counts.bs.2_KO counts.bs.3_KO counts.bs.4_WT
## 1      cds      12      16      9      20
## 2      cds      8      11      5      16
## 3      <NA>      NA      NA      NA      NA
## 4      <NA>      NA      NA      NA      NA
## 5      <NA>      NA      NA      NA      NA
## 6      utr3     46     35     24     42
##      counts.bs.5_WT counts.bs.6_WT      geneID.1 counts.bg.1_KO
## 1      14      9 ENSMUSG000000025907      1609
## 2      18      7 ENSMUSG000000025907      1609
## 3      NA      NA      <NA>      NA
## 4      NA      NA      <NA>      NA
```



```

## 5          NA          NA          <NA>          NA
## 6          39          18 ENSMUSG00000026012          552
## counts.bg.2_KO counts.bg.3_KO counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT
## 1          1973          1250          2638          2231          1352
## 2          1973          1250          2638          2231          1352
## 3          NA          NA          NA          NA          NA
## 4          NA          NA          NA          NA          NA
## 5          NA          NA          NA          NA          NA
## 6          714          444          654          641          299
## resBs.baseMean resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue
## 1          152.0207          0.3523528          0.3418073          1.0554415          0.3042572
## 2          132.1833          -0.2039860          0.3872352          0.2802662          0.5965268
## 3          NA          NA          NA          NA          NA
## 4          NA          NA          NA          NA          NA
## 5          NA          NA          NA          NA          NA
## 6          251.7179          0.1744790          0.2402881          0.5267887          0.4679601
## resBs.padj resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat
## 1          0.7257221          NA          NA          NA          NA
## 2          0.8894069          NA          NA          NA          NA
## 3          NA          NA          NA          NA          NA
## 4          NA          NA          NA          NA          NA
## 5          NA          NA          NA          NA          NA
## 6          0.8300039          NA          NA          NA          NA
## resBg.pvalue resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO
## 1          NA          NA          133.7259          117.9980
## 2          NA          NA          133.7259          117.9980
## 3          NA          NA          NA          NA
## 4          NA          NA          NA          NA
## 5          NA          NA          NA          NA
## 6          NA          NA          110.0847          102.4647
## tpm.counts.bg.3_KO tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1          129.8669          139.86348          146.2855          163.53603
## 2          129.8669          139.86348          146.2855          163.53603
## 3          NA          NA          NA          NA
## 4          NA          NA          NA          NA
## 5          NA          NA          NA          NA
## 6          110.6878          83.20238          100.8529          86.78335
## BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported down
## 1 ENSMUSG00000025907.bs7          3          3          TRUE FALSE
## 2 ENSMUSG00000025907.bs12          3          3          TRUE FALSE
## 3          <NA>          NA          NA          NA NA
## 4          <NA>          NA          NA          NA NA
## 5          <NA>          NA          NA          NA NA
## 6 ENSMUSG00000026012.bs4          3          3          TRUE FALSE
##
## 1 UAAGACUUUAACUCUGUGUUUCAGGUAUCCACAAGUCAGGCAUCCCCACAGUCAGCUGCUUCUCCAAGAAUAGAAAGUACAACAGGCAUUACAACCAC
## 2 AGGAAAAGCAUCAGCAAGAACUCCAUCUUUAAAAUUGAGUAUGAAUGUAAACUUGAUGCUCUAGUAAAAAGACAGUGAAGAAAAUGUAAAUAUUUUU
## 3 GCAGUUGAGGUCUAAGUCUGAUAAUUUGCAGGUGGCAUGAUGAUUUUUUCUUAUUUGUUCACCUGACAAUAUAUGAAGGUGUUCAGUCACUACCUCAU
## 4 UGAAAAAUAGUGAUUUUGGAGUAGGGGAUAUACCCGACAGUUGAAUGCUUGCCUUCAUUGUUAAGACUCUGGGUUUGCUCCCUCGCAUGGGUGCGAG
## 5 AAGAUUAUUUUUUAUUUGUCACAUAAUAUUUUUUAAGAACUUCACUGAAUGUAUAUGAGUUAUGUUUGAUUUUGUUGUUGUUAUAGUACGAUCAAAT
## 6 CUGUUUAGUAGCCUGUUUAUGUCAAUAUUUUAAGUUUAACUUCUACAGUUGGGCCACUUGUUGUCCUUGUGGAUGGAUAUCUGAAAUUGUGUCU
## rownumber mfs pvalue start_position binding_bases
## 1          12529 -17.9 0.911034          10 NACUCNNUGNNNCAGNGUNNNNNNNNN
## 2          12530 -21.5 0.243804          59 NGCUCNNNNNNANNGACAGUGNNGANNAUGUN

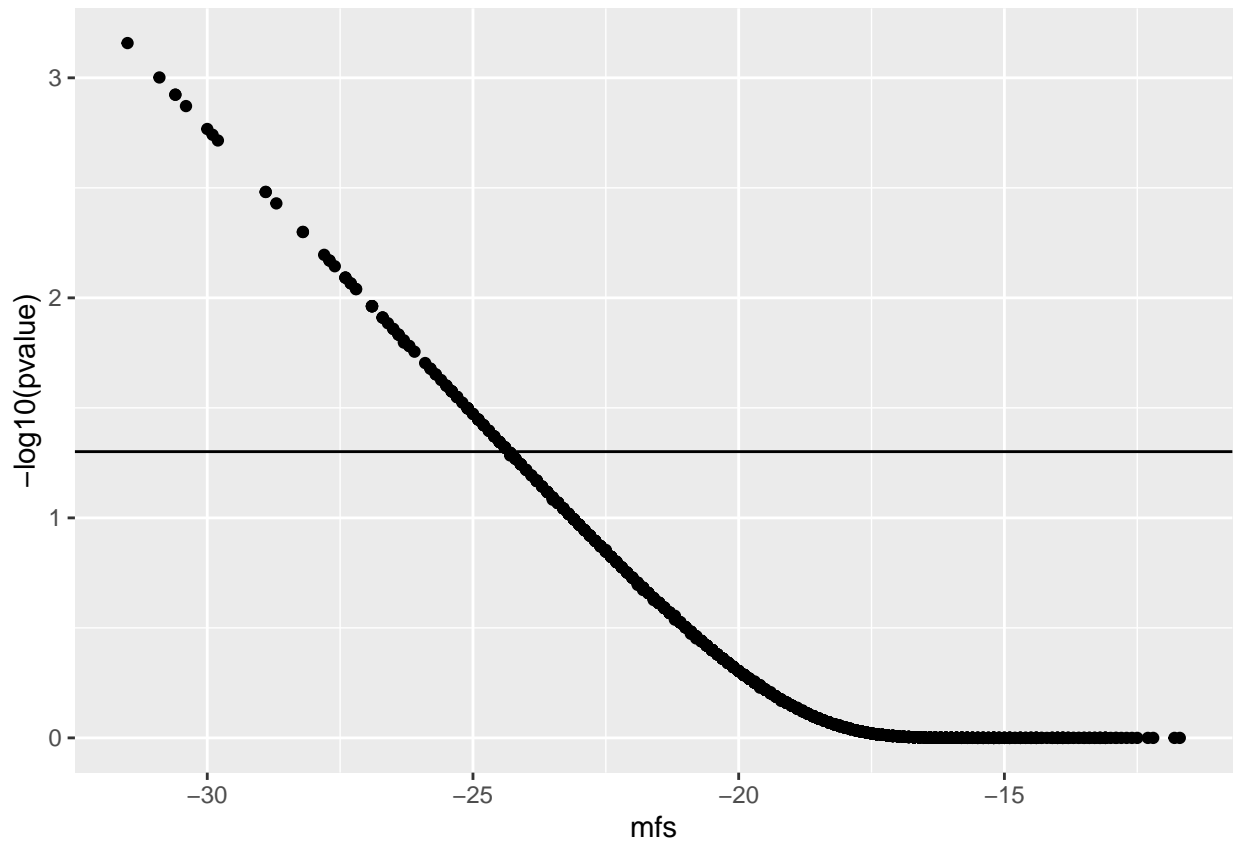
```

## 3	12531	-19.9	0.517782	57	NUUCACCNGACANUNNNUGANNGUGUUN
## 4	12532	-27.7	0.006767	29	NNUCANCCGACAGUNNUGAAUGNN
## 5	12533	-19.0	0.713812	41	NUUCACUGANNNUGUNNNUGANNNNAUGUUN
## 6	12534	-16.1	0.999191	41	NACUUNCUNACAGUNNUGGGNNNNNUGUUN

Plot pValues and mfs

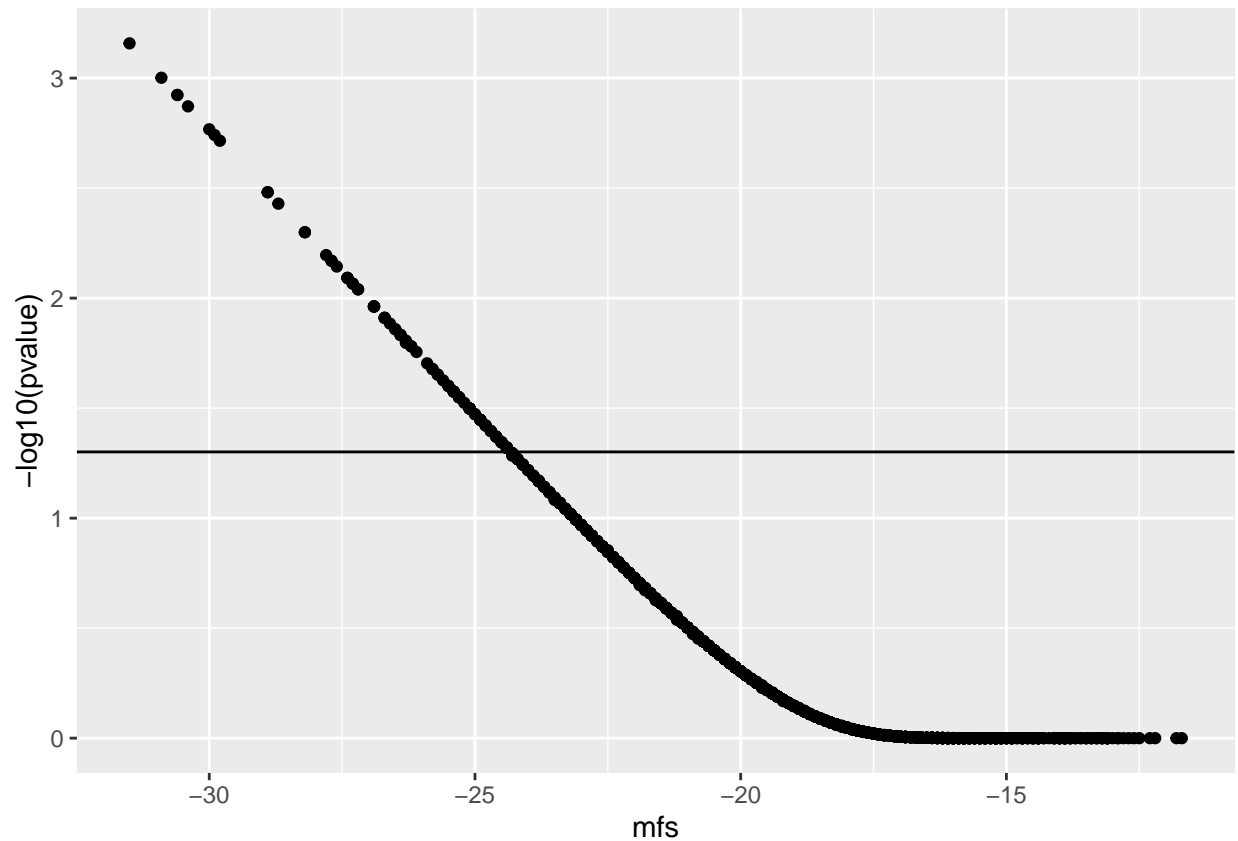
```
pmfsplotA <- ggplot(bsseqHA, aes(y=-log10(pvalue), x=mfs)) + geom_point() +
  geom_hline(yintercept=-log10(0.05))
```

pmfsplotA



```
pmfsplotB <- ggplot(bsseqHA, aes(y=-log10(pvalue), x=mfs)) + geom_point() +
  geom_hline(yintercept=-log10(0.05))
```

pmfsplotB



find motifs

make DNASTringSet

```
#make a DNA sting set as well because the motif finder requires it
b_baseSetDNAA <- DNASTringSet(chartr("U", "T", bsseqHA[bsseqHA$pvalue <= 0.05, "binding_bases"]))
b_baseSetDNAB <- DNASTringSet(chartr("U", "T", bsseqHB[bsseqHB$pvalue <= 0.05, "binding_bases"]))

head(b_baseSetDNAA)
```

```
## DNASTringSet object of length 6:
##      width seq
## [1]    24 NNTCANCCGACAGTNNTGAATGNN
## [2]    38 NACTCNNNNNNNNNNNACTGATNGTGTNGAATGTTN
## [3]    25 NCTCACTNNGATGGTNTTGAATGNN
## [4]    25 NGCTNNCCGACAGCNNNTGAATGNN
## [5]    25 NNNCACCGNCAGCNNNTTGGATGTN
## [6]    24 NNNNGCNGACAGCGNNNGAATGNN
```

```
head(b_baseSetDNAB)
```

```
## DNASTringSet object of length 6:
##      width seq
## [1]    24 NNTCANCCGACAGTNNTGAATGNN
## [2]    38 NACTCNNNNNNNNNNNACTGATNGTGTNGAATGTTN
## [3]    25 NCTCACTNNGATGGTNTTGAATGNN
```

```
## [4] 25 NGCTNNCCGACAGCANNNTGAATGNN
## [5] 25 NNNCACCGNCAGCANNNTTGGATGTN
## [6] 24 NNNNGCNGACAGCGNNGAATGNN
```

Run rGadem

```
enrichmotA <- GADEM(b_baseSetDNAA, seed = 123)
```

```
## top 3 4, 5-mers: 4 30 44
```

```
enrichmotB <- GADEM(b_baseSetDNAB, seed = 123)
```

```
## top 3 4, 5-mers: 4 30 44
```

miR181a

```
rGADEM::plot(enrichmotA)
```

miR181b

```
rGADEM::plot(enrichmotB)
```

Session info

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8 LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## attached base packages:
## [1] grid      stats4    stats      graphics  grDevices  utils      datasets
## [8] methods   base
##
## other attached packages:
## [1] rGADEM_2.46.0          seqLogo_1.64.0
## [3] seqinr_4.2-30          ComplexHeatmap_2.15.2
## [5] circlize_0.4.15        ggplot2_3.4.2
## [7] dplyr_1.1.1            BSgenome.Mmusculus.UCSC.mm10_1.4.3
## [9] BSgenome_1.66.3        rtracklayer_1.58.0
## [11] Biostrings_2.66.0      XVector_0.38.0
## [13] GenomicRanges_1.50.2   GenomeInfoDb_1.34.9
## [15] IRanges_2.32.0         S4Vectors_0.36.2
## [17] BiocGenerics_0.44.0
##
```

```

## loaded via a namespace (and not attached):
## [1] MatrixGenerics_1.10.0      Biobase_2.58.0
## [3] foreach_1.5.2              highr_0.10
## [5] GenomeInfoDbData_1.2.9     Rsamtools_2.14.0
## [7] yaml_2.3.7                 pillar_1.9.0
## [9] lattice_0.20-45            glue_1.6.2
## [11] digest_0.6.31              RColorBrewer_1.1-3
## [13] colorspace_2.1-0           htmltools_0.5.4
## [15] Matrix_1.5-3               XML_3.99-0.14
## [17] pkgconfig_2.0.3            GetoptLong_1.0.5
## [19] zlibbioc_1.44.0            scales_1.2.1
## [21] BiocParallel_1.32.6        tibble_3.2.1
## [23] farver_2.1.1               generics_0.1.3
## [25] withr_2.5.0                SummarizedExperiment_1.28.0
## [27] cli_3.6.0                  magrittr_2.0.3
## [29] crayon_1.5.2               evaluate_0.20
## [31] fansi_1.0.4                doParallel_1.0.17
## [33] MASS_7.3-58.2              tools_4.2.3
## [35] GlobalOptions_0.1.2        BiocIO_1.8.0
## [37] lifecycle_1.0.3            matrixStats_0.63.0
## [39] munsell_0.5.0              cluster_2.1.4
## [41] DelayedArray_0.23.2        ade4_1.7-22
## [43] compiler_4.2.3             rlang_1.1.0
## [45] RCurl_1.98-1.12            iterators_1.0.14
## [47] rstudioapi_0.14            rjson_0.2.21
## [49] bitops_1.0-7               labeling_0.4.2
## [51] rmarkdown_2.21             restfulr_0.0.15
## [53] gtable_0.3.3               codetools_0.2-19
## [55] R6_2.5.1                   GenomicAlignments_1.34.1
## [57] knitr_1.42                 fastmap_1.1.1
## [59] utf8_1.2.3                 clue_0.3-64
## [61] shape_1.4.6                parallel_4.2.3
## [63] Rcpp_1.0.10                vctrs_0.6.1
## [65] png_0.1-8                  tidyselect_1.2.0
## [67] xfun_0.37

```