RNAhybrid_complete

Nikita Verheyden

2023-04-26

setup

dir

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

# work
#setwd("Z:/Personen/Nikita/Publications/miR181_paper/Figure2/RNAhybrid")
```

packages

```
#home
source("D:/Krueger_Lab/Publications/miR181_paper_v21022023/Figure_theme/theme_paper.R")
#work
#source("Z:/Personen/Nikita/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
```

```r
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(stringr)
```

data

```
#home
f2bs <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure2/MRE_bound_gene_and_bound_region/mir18
#work
#f2bs <- readRDS("Z:/Personen/Nikita/Publications/miR181_paper/Figure2/MRE_bound_gene_and_bound_region/

head(f2bs)
```

```
##   seqnames   start     end width strand scoreSum scoreMean scoreMax
## 1     chr1 6245651 6245657     7      +  9.52553  4.762765  6.00678
## 2     chr1 6248341 6248347     7      + 92.68921 23.172303 48.76900
## 3     chr1 6248857 6248863     7      + 14.07133  7.035665  7.04425
## 4     chr1 6248918 6248924     7      + 38.91451 12.971503 20.65080
## 5     chr1 7170481 7170487     7      + 66.92218 13.384436 25.84490
## 6     chr1 9899605 9899611     7      + 25.15963  6.289907  8.61019
##         geneType geneName             geneID region BS_ID          mir_IP
## 1 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     5 mmu-miR-181a-5p
## 2 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     8 mmu-miR-181a-5p
## 3 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    10 mmu-miR-181a-5p
## 4 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    11 mmu-miR-181a-5p
## 5 protein_coding   Pcmtd1 ENSMUSG00000051285   utr3    19 mmu-miR-181a-5p
## 6 protein_coding     Sgk3 ENSMUSG00000025915   utr3    23 mmu-miR-181a-5p
##   n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d               set WT KO
## 1        1         1         0         0         0 ago_bs_mir181_chi  1  1
## 2        5         5         0         0         0 ago_bs_mir181_chi  1  1
## 3        6         6         0         0         0 ago_bs_mir181_chi  1  0
## 4        6         6         0         0         0 ago_bs_mir181_chi  1  1
## 5        4         4         0         0         0 ago_bs_mir181_chi  1  1
## 6        1         1         0         0         0 ago_bs_mir181_chi NA NA
##             geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 1 ENSMUSG00000025907     Rb1cc1      cds              4              3
## 2 ENSMUSG00000025907     Rb1cc1      cds             28             32
## 3 ENSMUSG00000025907     Rb1cc1      cds             13             11
## 4 ENSMUSG00000025907     Rb1cc1      cds             15             15
## 5 ENSMUSG00000051285     Pcmtd1     utr3             12             22
## 6               <NA>       <NA>     <NA>             NA             NA
##   counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 1              3              3             10              3
## 2             27             46             41             20
## 3              4             22             13             12
## 4             10             33             20             18
## 5             14             16             20              9
## 6             NA             NA             NA             NA
##             geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 1 ENSMUSG00000025907           1609           1973           1250
## 2 ENSMUSG00000025907           1609           1973           1250
## 3 ENSMUSG00000025907           1609           1973           1250
## 4 ENSMUSG00000025907           1609           1973           1250
## 5 ENSMUSG00000051285           1355           1706           1064
## 6               <NA>             NA             NA             NA
##   counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 1           2638           2231           1352       92.10645
## 2           2638           2231           1352      281.53271
## 3           2638           2231           1352      145.51107
## 4           2638           2231           1352      186.74162
## 5           1654           1348            755      151.36245
## 6             NA             NA             NA             NA
##   resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue resBs.padj
## 1           -0.1093039   0.5923673 0.03419066    0.8533018  0.9652601
## 2            0.2749428   0.2351157 1.35874137    0.2437557  0.6729889
## 3           -0.1805519   0.3623758 0.25017050    0.6169550  0.8961239
## 4           -0.2606282   0.3062717 0.73169661    0.3923338  0.7868678
## 5            0.1466485   0.3122905 0.22052922    0.6386370  0.9013566
## 6                   NA          NA         NA           NA         NA
##   resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 1             NA                   NA          NA         NA           NA
## 2             NA                   NA          NA         NA           NA
## 3             NA                   NA          NA         NA           NA
## 4             NA                   NA          NA         NA           NA
## 5             NA                   NA          NA         NA           NA
## 6             NA                   NA          NA         NA           NA
##   resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 1         NA           133.7259           117.9980           129.8669
## 2         NA           133.7259           117.9980           129.8669
## 3         NA           133.7259           117.9980           129.8669
## 4         NA           133.7259           117.9980           129.8669
## 5         NA           248.6210           225.2505           244.0445
## 6         NA                 NA                 NA                 NA
##   tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1           139.8635           146.2855           163.5360
## 2           139.8635           146.2855           163.5360
## 3           139.8635           146.2855           163.5360
## 4           139.8635           146.2855           163.5360
## 5           193.5994           195.1330           201.6149
## 6                 NA                 NA                 NA
##                   BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported  down
## 1  ENSMUSG00000025907.bs5              3              3          TRUE FALSE
## 2  ENSMUSG00000025907.bs8              3              3          TRUE FALSE
## 3 ENSMUSG00000025907.bs10              3              3          TRUE FALSE
## 4 ENSMUSG00000025907.bs11              3              3          TRUE FALSE
## 5  ENSMUSG00000051285.bs4              3              3          TRUE FALSE
## 6                    <NA>             NA             NA            NA    NA
```

```
#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 <- "#008EA0FF"
farbe14 <- "#8A4198FF"
farbe15 <- "#5A9599FF"
farbe16 <- "#FF6348FF"

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

Get DNA sequences

```
#resize ranges

f2bsLA <- f2bs
f2bsLA$start <- f2bsLA$start -25
f2bsLA$end <- f2bsLA$end + 50
f2bsLA$n_mir181a <- as.numeric(f2bsLA$n_mir181a)


df181A <- mutate(f2bsLA, Sequence = as.character(getSeq(BSgenome.Mmusculus.UCSC.mm10, seqnames, start, 
df181A$rownum <- rownames(df181A)
df181A <- df181A[as.numeric(df181A$n_mir181a) > 0,]

#and turn T into Us
df181A$Sequence <- gsub('T', 'U', df181A$Sequence)

head(df181A)
```

```
##   seqnames   start     end width strand scoreSum scoreMean scoreMax
## 1     chr1 6245626 6245707     7      +  9.52553  4.762765  6.00678
## 2     chr1 6248316 6248397     7      + 92.68921 23.172303 48.76900
## 3     chr1 6248832 6248913     7      + 14.07133  7.035665  7.04425
## 4     chr1 6248893 6248974     7      + 38.91451 12.971503 20.65080
## 5     chr1 7170456 7170537     7      + 66.92218 13.384436 25.84490
## 6     chr1 9899580 9899661     7      + 25.15963  6.289907  8.61019
##         geneType geneName             geneID region BS_ID          mir_IP
## 1 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     5 mmu-miR-181a-5p
## 2 protein_coding   Rb1cc1 ENSMUSG00000025907    cds     8 mmu-miR-181a-5p
## 3 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    10 mmu-miR-181a-5p
## 4 protein_coding   Rb1cc1 ENSMUSG00000025907    cds    11 mmu-miR-181a-5p
## 5 protein_coding   Pcmtd1 ENSMUSG00000051285   utr3    19 mmu-miR-181a-5p
## 6 protein_coding     Sgk3 ENSMUSG00000025915   utr3    23 mmu-miR-181a-5p
##   n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d               set WT KO
## 1        1         1         0         0         0 ago_bs_mir181_chi  1  1
## 2        5         5         0         0         0 ago_bs_mir181_chi  1  1
## 3        6         6         0         0         0 ago_bs_mir181_chi  1  0
## 4        6         6         0         0         0 ago_bs_mir181_chi  1  1
## 5        4         4         0         0         0 ago_bs_mir181_chi  1  1
## 6        1         1         0         0         0 ago_bs_mir181_chi NA NA
##             geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 1 ENSMUSG00000025907     Rb1cc1      cds              4              3
## 2 ENSMUSG00000025907     Rb1cc1      cds             28             32
## 3 ENSMUSG00000025907     Rb1cc1      cds             13             11
## 4 ENSMUSG00000025907     Rb1cc1      cds             15             15
## 5 ENSMUSG00000051285     Pcmtd1     utr3             12             22
## 6               <NA>       <NA>     <NA>             NA             NA
##   counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 1              3              3             10              3
## 2             27             46             41             20
## 3              4             22             13             12
## 4             10             33             20             18
## 5             14             16             20              9
## 6             NA             NA             NA             NA
##             geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 1 ENSMUSG00000025907           1609           1973           1250
## 2 ENSMUSG00000025907           1609           1973           1250
## 3 ENSMUSG00000025907           1609           1973           1250
## 4 ENSMUSG00000025907           1609           1973           1250
## 5 ENSMUSG00000051285           1355           1706           1064
## 6               <NA>             NA             NA             NA
##   counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 1           2638           2231           1352       92.10645
## 2           2638           2231           1352      281.53271
## 3           2638           2231           1352      145.51107
## 4           2638           2231           1352      186.74162
## 5           1654           1348            755      151.36245
## 6             NA             NA             NA             NA
##   resBs.log2FoldChange resBs.lfcSE resBs.stat resBs.pvalue resBs.padj
## 1           -0.1093039   0.5923673 0.03419066    0.8533018  0.9652601
## 2            0.2749428   0.2351157 1.35874137    0.2437557  0.6729889
## 3           -0.1805519   0.3623758 0.25017050    0.6169550  0.8961239
## 4           -0.2606282   0.3062717 0.73169661    0.3923338  0.7868678
## 5            0.1466485   0.3122905 0.22052922    0.6386370  0.9013566
## 6                   NA          NA         NA           NA         NA
##   resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 1             NA                   NA          NA         NA           NA
## 2             NA                   NA          NA         NA           NA
## 3             NA                   NA          NA         NA           NA
## 4             NA                   NA          NA         NA           NA
## 5             NA                   NA          NA         NA           NA
## 6             NA                   NA          NA         NA           NA
##   resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 1         NA           133.7259           117.9980           129.8669
## 2         NA           133.7259           117.9980           129.8669
## 3         NA           133.7259           117.9980           129.8669
## 4         NA           133.7259           117.9980           129.8669
## 5         NA           248.6210           225.2505           244.0445
## 6         NA                 NA                 NA                 NA
##   tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1           139.8635           146.2855           163.5360
## 2           139.8635           146.2855           163.5360
## 3           139.8635           146.2855           163.5360
## 4           139.8635           146.2855           163.5360
## 5           193.5994           195.1330           201.6149
## 6                 NA                 NA                 NA
##                   BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported  down
## 1  ENSMUSG00000025907.bs5              3              3          TRUE FALSE
## 2  ENSMUSG00000025907.bs8              3              3          TRUE FALSE
## 3 ENSMUSG00000025907.bs10              3              3          TRUE FALSE
## 4 ENSMUSG00000025907.bs11              3              3          TRUE FALSE
## 5  ENSMUSG00000051285.bs4              3              3          TRUE FALSE
## 6                    <NA>             NA             NA            NA    NA
##                                                                              Sequence
## 1 UAAAGGACUGGACUCCUGGCCUUCCUCAUUUUGUGUAUGUAUUUUUUUUUUUCUAACUAGGACUAAAUUUCUUUUUUUUUUU
## 2 CAAGAAUAGAAAGUACAACAGGCAUUACAACCACUACCUCACCAAAAACUCCUCCUCCACUAACUGUUCAGGACACCUUAUG
## 3 UACAAAAAGAACAGUGUGACUUAGCAAAUUAUUUAAAAUGUACAGCUGUAGAAAUAAGAAAUAUUAUUGAAAAAGUAAAAUG
## 4 UAUUAUUGAAAAAGUAAAAUGUUCUCUAGAAAUAACACUAAAGGAAAAGCAUCAGCAAGAACUCCAAUCUUUAAAAAUUGAG
## 5 GGAAAAUUUCUGCUUCUCUCAUAGAGAUUUUUAAGAGCUAGUGAAUGUUAAAGUAGGAAGUGGCUACUUGACACAACUAGUU
## 6 GAAGUGUAAUAAAAUGCUACCAGAUGUUUUUUUAAGGUGGUACCCACCAUAAUGUCUCUGUCACAUAUUUAUAUUACAAAUG
##   rownum
## 1      1
## 2      2
## 3      3
## 4      4
## 5      5
## 6      6
```

find seed

```
#find both seeds
seed1 <- df181A %>% filter(str_detect(Sequence, "GAAUGU"))
seed2 <- df181A %>% filter(str_detect(Sequence, "GAUUGU"))

#combine
seedm <- rbind(seed1, seed2)
#remove duplicates
seedm <- seedm[!duplicated(seedm$rownum),]
#remove NAs in gene name
seedm <- seedm[!is.na(seedm$geneName),]

head(seedm)
```

```
##    seqnames     start       end width strand scoreSum scoreMean scoreMax
## 5      chr1   7170456   7170537     7      + 66.92218 13.384436  25.8449
## 7      chr1   9899846   9899927     7      + 20.67430  6.891433  11.2987
## 22     chr1  43570279  43570360     7      + 55.09962 13.774905  21.7788
## 31     chr1  58754246  58754327     7      + 21.46580 10.732900  15.0257
## 51     chr1  85849941  85850022     7      + 56.80659 18.935530  31.2761
## 86     chr1 119528130 119528211     7      + 19.78365  9.891825  12.0423
##          geneType geneName             geneID region BS_ID          mir_IP
## 5  protein_coding   Pcmtd1 ENSMUSG00000051285   utr3    19 mmu-miR-181a-5p
## 7  protein_coding     Sgk3 ENSMUSG00000025915   utr3    24 mmu-miR-181a-5p
## 22 protein_coding     Nck2 ENSMUSG00000066877   utr3    97 mmu-miR-181b-5p
## 31 protein_coding    Cflar ENSMUSG00000026031   utr3   130 mmu-miR-181a-5p
## 51 protein_coding    Cab39 ENSMUSG00000036707   utr3   209 mmu-miR-181a-5p
## 86 protein_coding Tmem185b ENSMUSG00000098923   utr3   320 mmu-miR-181a-5p
##    n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d               set WT KO
## 5         4         4         0         0         0 ago_bs_mir181_chi  1  1
## 7         1         1         0         0         0 ago_bs_mir181_chi NA NA
## 22        2         1         1         0         0 ago_bs_mir181_chi  1  1
## 31      170       163         5         2         0 ago_bs_mir181_chi  0  1
## 51      326       297        29         0         0 ago_bs_mir181_chi  1  0
## 86        7         7         0         0         0 ago_bs_mir181_chi NA NA
##              geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 5  ENSMUSG00000051285     Pcmtd1     utr3             12             22
## 7                <NA>       <NA>     <NA>             NA             NA
## 22 ENSMUSG00000066877       Nck2     utr3             11             15
## 31 ENSMUSG00000026031      Cflar     utr3              6             14
## 51 ENSMUSG00000036707      Cab39     utr3              2              2
## 86               <NA>       <NA>     <NA>             NA             NA
##    counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 5              14             16             20              9
## 7              NA             NA             NA             NA
## 22             10             20             21             12
## 31             11             24             23             10
## 51              3             85             50             32
## 86             NA             NA             NA             NA
##              geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 5  ENSMUSG00000051285           1355           1706           1064
## 7                <NA>             NA             NA             NA
## 22 ENSMUSG00000066877           5026           7988           4386
## 31 ENSMUSG00000026031           1371           1785           1002
## 51 ENSMUSG00000036707           1606           2091           1226
## 86               <NA>             NA             NA             NA
##    counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 5            1654           1348            755       151.3625
## 7              NA             NA             NA             NA
## 22           8478           6584           3537       300.8802
## 31           2308           1960           1040       148.2842
## 51           2802           2422           1406       234.7085
## 86             NA             NA             NA             NA
##    resBs.log2FoldChange resBs.lfcSE  resBs.stat resBs.pvalue   resBs.padj
## 5             0.1466485   0.3122905   0.2205292 6.386370e-01 9.013566e-01
## 7                    NA          NA          NA           NA           NA
## 22           -0.2719827   0.3328411   0.6723828 4.122221e-01 8.000076e-01
## 31           -0.3288789   0.3315817   1.0018062 3.168738e-01 7.352453e-01
## 51           -3.9344782   0.5635927 107.5785787 3.324592e-25 1.174121e-21
## 86                   NA          NA          NA           NA           NA
##    resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 5              NA                   NA          NA         NA           NA
## 7              NA                   NA          NA         NA           NA
## 22             NA                   NA          NA         NA           NA
## 31             NA                   NA          NA         NA           NA
## 51             NA                   NA          NA         NA           NA
## 86             NA                   NA          NA         NA           NA
##    resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 5          NA           248.6210           225.2505           244.0445
## 7          NA                 NA                 NA                 NA
## 22         NA          1377.3876          1575.2874          1502.5578
## 31         NA           132.1224           123.7842           120.7079
## 51         NA           277.0625           259.5818           264.3940
## 86         NA                 NA                 NA                 NA
##    tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 5            193.5994           195.1330           201.6149
## 7                  NA                 NA                 NA
## 22          1482.1666          1423.5269          1410.7365
## 31           141.8877           149.0175           145.8645
## 51           308.3687           329.6465           353.0162
## 86                 NA                 NA                 NA
##                    BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported  down
## 5   ENSMUSG00000051285.bs4              3              3          TRUE FALSE
## 7                     <NA>             NA             NA            NA    NA
## 22 ENSMUSG00000066877.bs31              3              3          TRUE FALSE
## 31  ENSMUSG00000026031.bs4              3              3          TRUE FALSE
## 51  ENSMUSG00000036707.bs4              3              3          TRUE  TRUE
## 86                    <NA>             NA             NA            NA    NA
##                                                                               Sequence
## 5  GGAAAAUUUCUGCUUCUCUCAUAGAGAUUUUUAAGAGCUAGUGAAUGUUAAAGUAGGAAGUGGCUACUUGACACAACUAGUU
## 7  GGCAAGUCUGGGUUGGUGUGAAUGUGUGUCACCUACACAUUCUAACAGAAGGUAACAAUAAGUUAGCAGUGACAUAUUCAGU
## 22 AUAUAUUAUUUGCUUUACAGGGAAAUUUUUCAGGGUUUACAAAAGAAUAUGUGAUUAGUAGUAACAGAAUGUUUAUGAAGAA
## 31 UGGGUGUAUAGUGUAUAGUGGUUCAAGAUUUGACACUGAAUGUAACUUGAGACUUACCUGAGUUUGUCAUGCGACUGGGUAA
## 51 UGUAUAUAAUUCUUAGAAUGCUCAUUUCUUUUAAAUCGUUUAAUUUGUACAGCAGAGGAAUGUUAUUGUAGUAGUAUGUAAC
## 86 UGCAUAUAUUAGUAUUUAUAUGAAUGUUUUAGCAGUGUUAUCUGUGUUGAUUGUAGUUCUUGGCAGUAAUGUAUUGUGUUAA
##    rownum
## 5       5
## 7       7
## 22     22
## 31     31
## 51     51
## 86     86
```

Write to .fasta

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

```r
candgeneNameA <- as.list(seedm$geneName)
candrnameA <- as.list(seedm$rownum)
condgeneSeqA <- as.list(seedm$Sequence)


#change to output directory

setwd("D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/fastafiles_complete/A")

for (i in 1:length(candgeneNameA)) {
  write.fasta(condgeneSeqA[i],candrnameA[i],paste(candrnameA[i], candgeneNameA[i], "miR_181a", 'fasta',
}
```

```
Personalized_Reader <- function(lambda){
 read.table(lambda, sep = ":") %>% select(V1, V5, V6, V7, V10, V11)}

#remove NA file...I just dont get it....where is it coming from?

#File lists
reslistA <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/resu



#import
myfilelistA <- lapply(reslistA, Personalized_Reader)


resframeA <- bind_rows(myfilelistA)




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

resframeA[is.na(resframeA$non_binding_bases),"non_binding_bases"] <- "                       "




head(resframeA)
```

```
##   rownumber   mfs   pvalue start_position               binding_bases
## 1      1004 -24.7 0.024036             66         UGGCUGUC   ACUUACA
## 2      1005 -25.4 0.015632             46         UGGCUGUC   ACUUACA
## 3      1035 -11.0 1.000000              1                     CUUACA
## 4      1043 -18.4 0.703212             46     UG C  UCG AA    CUUACA
## 5      1050 -14.6 0.999997             60          GCUG CG AAC UACAA
## 6      1054 -17.8 0.828455             25       GUGG UGUCG AACU  ACA
##             non_binding_bases
## 1     UGAG        GCA       A
## 2     UGAG        GCA       A
## 3     UGAGUGGCUGUCGCAA      A
## 4 UGAG  G UG   C            A
## 5    UGAGUG    U  C   U
## 6    UGA    C     C    U    A
```

merge with original df

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

seedm$rownumber <- as.character(seedm$rownum)
resframeA$rownumber <- as.character(resframeA$rownumber)
```

```
bsseqHA <- left_join(seedm, resframeA, by="rownumber")

head(bsseqHA)
```

```
##   seqnames     start       end width strand scoreSum scoreMean scoreMax
## 1     chr1   7170456   7170537     7      + 66.92218 13.384436  25.8449
## 2     chr1   9899846   9899927     7      + 20.67430  6.891433  11.2987
## 3     chr1  43570279  43570360     7      + 55.09962 13.774905  21.7788
## 4     chr1  58754246  58754327     7      + 21.46580 10.732900  15.0257
## 5     chr1  85849941  85850022     7      + 56.80659 18.935530  31.2761
## 6     chr1 119528130 119528211     7      + 19.78365  9.891825  12.0423
##         geneType geneName             geneID region BS_ID          mir_IP
## 1 protein_coding   Pcmtd1 ENSMUSG00000051285   utr3    19 mmu-miR-181a-5p
## 2 protein_coding     Sgk3 ENSMUSG00000025915   utr3    24 mmu-miR-181a-5p
## 3 protein_coding     Nck2 ENSMUSG00000066877   utr3    97 mmu-miR-181b-5p
## 4 protein_coding    Cflar ENSMUSG00000026031   utr3   130 mmu-miR-181a-5p
## 5 protein_coding    Cab39 ENSMUSG00000036707   utr3   209 mmu-miR-181a-5p
## 6 protein_coding Tmem185b ENSMUSG00000098923   utr3   320 mmu-miR-181a-5p
##   n_mir181 n_mir181a n_mir181b n_mir181c n_mir181d               set WT KO
## 1        4         4         0         0         0 ago_bs_mir181_chi  1  1
## 2        1         1         0         0         0 ago_bs_mir181_chi NA NA
## 3        2         1         1         0         0 ago_bs_mir181_chi  1  1
## 4      170       163         5         2         0 ago_bs_mir181_chi  0  1
## 5      326       297        29         0         0 ago_bs_mir181_chi  1  0
## 6        7         7         0         0         0 ago_bs_mir181_chi NA NA
##           geneID.2 geneName.1 region.1 counts.bs.1_KO counts.bs.2_KO
## 1 ENSMUSG00000051285     Pcmtd1     utr3             12             22
## 2               <NA>       <NA>     <NA>             NA             NA
## 3 ENSMUSG00000066877       Nck2     utr3             11             15
## 4 ENSMUSG00000026031      Cflar     utr3              6             14
## 5 ENSMUSG00000036707      Cab39     utr3              2              2
## 6               <NA>       <NA>     <NA>             NA             NA
##   counts.bs.3_KO counts.bs.4_WT counts.bs.5_WT counts.bs.6_WT
## 1             14             16             20              9
## 2             NA             NA             NA             NA
## 3             10             20             21             12
## 4             11             24             23             10
## 5              3             85             50             32
## 6             NA             NA             NA             NA
##           geneID.1 counts.bg.1_KO counts.bg.2_KO counts.bg.3_KO
## 1 ENSMUSG00000051285           1355           1706           1064
## 2               <NA>             NA             NA             NA
## 3 ENSMUSG00000066877           5026           7988           4386
## 4 ENSMUSG00000026031           1371           1785           1002
## 5 ENSMUSG00000036707           1606           2091           1226
## 6               <NA>             NA             NA             NA
##   counts.bg.4_WT counts.bg.5_WT counts.bg.6_WT resBs.baseMean
## 1           1654           1348            755       151.3625
## 2             NA             NA             NA             NA
## 3           8478           6584           3537       300.8802
## 4           2308           1960           1040       148.2842
## 5           2802           2422           1406       234.7085
## 6             NA             NA             NA             NA
##   resBs.log2FoldChange resBs.lfcSE  resBs.stat resBs.pvalue   resBs.padj
## 1            0.1466485   0.3122905   0.2205292 6.386370e-01 9.013566e-01
## 2                   NA          NA          NA           NA           NA
## 3           -0.2719827   0.3328411   0.6723828 4.122221e-01 8.000076e-01
## 4           -0.3288789   0.3315817   1.0018062 3.168738e-01 7.352453e-01
## 5           -3.9344782   0.5635927 107.5785787 3.324592e-25 1.174121e-21
## 6                   NA          NA          NA           NA           NA
##   resBg.baseMean resBg.log2FoldChange resBg.lfcSE resBg.stat resBg.pvalue
## 1             NA                   NA          NA         NA           NA
## 2             NA                   NA          NA         NA           NA
## 3             NA                   NA          NA         NA           NA
## 4             NA                   NA          NA         NA           NA
## 5             NA                   NA          NA         NA           NA
## 6             NA                   NA          NA         NA           NA
##   resBg.padj tpm.counts.bg.1_KO tpm.counts.bg.2_KO tpm.counts.bg.3_KO
## 1         NA           248.6210           225.2505           244.0445
## 2         NA                 NA                 NA                 NA
## 3         NA          1377.3876          1575.2874          1502.5578
## 4         NA           132.1224           123.7842           120.7079
## 5         NA           277.0625           259.5818           264.3940
## 6         NA                 NA                 NA                 NA
##   tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 1           193.5994           195.1330           201.6149
## 2                 NA                 NA                 NA
## 3          1482.1666          1423.5269          1410.7365
## 4           141.8877           149.0175           145.8645
## 5           308.3687           329.6465           353.0162
## 6                 NA                 NA                 NA
##                   BS_ID.1 tpm_support_KO tpm_support_WT tpm_supported  down
## 1  ENSMUSG00000051285.bs4              3              3          TRUE FALSE
## 2                    <NA>             NA             NA            NA    NA
## 3 ENSMUSG00000066877.bs31              3              3          TRUE FALSE
## 4  ENSMUSG00000026031.bs4              3              3          TRUE FALSE
## 5  ENSMUSG00000036707.bs4              3              3          TRUE  TRUE
## 6                    <NA>             NA             NA            NA    NA
##                                                                             Sequence
## 1 GGAAAAUUUCUGCUUCUCUCAUAGAGAUUUUUAAGAGCUAGUGAAUGUUAAAGUAGGAAGUGGCUACUUGACACAACUAGUU
## 2 GGCAAGUCUGGGUUGGUGUGAAUGUGUGUCACCUACACAUUCUAACAGAAGGUAACAAUAAGUUAGCAGUGACAUAUUCAGU
## 3 AUAUAUUAUUUGCUUUACAGGGAAAUUUUUCAGGGUUUACAAAAGAAUAUGUGAUUAGUAGUAACAGAAUGUUUAUGAAGAA
## 4 UGGGUGUAUAGUGUAUAGUGGUUCAAGAUUUGACACUGAAUGUAACUUGAGACUUACCUGAGUUUGUCAUGCGACUGGGUAA
## 5 UGUAUAUAAUUCUUAGAAUGCUCAUUUCUUUUAAAUCGUUUAAUUUGUACAGCAGAGGAAUGUUAUUGUAGUAGUAUGUAAC
## 6 UGCAUAUAUUAGUAUUUAUAUGAAUGUUUUAGCAGUGUUAUCUGUGUUGAUUGUAGUUCUUGGCAGUAAUGUAUUGUGUUAA
##   rownum rownumber   mfs   pvalue start_position
## 1      5         5 -23.3 0.056025             17
## 2      7         7 -18.8 0.612351              7
## 3     22        22 -13.2 1.000000             13
## 4     31        31 -18.3 0.725425             21
## 5     51        51 -21.9 0.129199             19
## 6     86        86 -18.4 0.703212             56
##                                     binding_bases
## 1              GAGUG   CUG       UCG   ACUUACAA
## 2                            GGCU  GUCGC ACUUACA
## 3        GAGUG   CU       GUC CAA       CU  UACA
## 4                      GAGU    GGCUGU G  ACUUACA
## 5  UGAGU           GGC          UGUCG C  CUUACAA
## 6                         GAG GGCUGUCG  AC   ACA
##                                  non_binding_bases
## 1              U     G               CA
## 2                       UGAGU           A       A
## 3       U     G              G                  A
## 4                     U              C CA       A
## 5                                      AA
## 6                        U   U        CA  UU    A
```

——————————————————————————-continue here

Process data (remove gaps)

Due to the loops in the mRNA there are additional spaces in the mirna. We only want the binding and non binding bases of hte mirna in te correct order. For that we will remove all gaps that origin in the mRNA loops.

```r
#binding and non binding bases as characters in a list
Alistbb <- strsplit(resframeA$binding_bases,"")
Alistnb <- strsplit(resframeA$non_binding_bases,"")

#combine the two lists
Alist <- Map(cbind, Alistbb, Alistnb)
```

```
## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
```

```r
Alist <- lapply(Alist, as.data.frame)

#remove all empty rows (mRNA loops)
Alist0 <- lapply(Alist, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})

#rewrite as characters
AlistF <- lapply(Alist0, function(x){
  paste(x[,1],  collapse = '')
})


#Attach lists back onto original data.frame as new column
resframeA$binding_nospace <-unlist(AlistF)
head(resframeA$binding_nospace)
```

```
## [1] "    UGGCUGUC   ACUUACA " "    UGGCUGUC   ACUUACA "
## [3] "                CUUACA " "    UG C  UCG AACUUACA "
## [5] "      GCUG CG AAC UACAA" "   GUGG UGUCG AACU ACA "
```

Transform into Numbers

add 0s

replace all gaps with 0 and all letters with 1

```
#0
resframeA$binding_nospace <- chartr(" ", "0", resframeA$binding_nospace)

#1
resframeA$binding_nospace <- mgsub::mgsub(resframeA$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 4

head(resframeA)
```

```
##   rownumber   mfs   pvalue start_position               binding_bases
## 1      1004 -24.7 0.024036             66         UGGCUGUC   ACUUACA
## 2      1005 -25.4 0.015632             46         UGGCUGUC   ACUUACA
## 3      1035 -11.0 1.000000              1                     CUUACA
## 4      1043 -18.4 0.703212             46     UG C  UCG AA    CUUACA
## 5      1050 -14.6 0.999997             60          GCUG CG AAC UACAA
## 6      1054 -17.8 0.828455             25       GUGG UGUCG AACU  ACA
##              non_binding_bases         binding_nospace
## 1     UGAG        GCA       A 00001111111100011111110
## 2     UGAG        GCA       A 00001111111100011111110
## 3     UGAGUGGCUGUCGCAA      A 00000000000000001111110
## 4 UGAG  G UG   C            A 00001101001110111111110
## 5    UGAGUG    U  C   U       00000011110110111011111
## 6    UGA    C     C    U    A 00011110111110111101110
```

seperate into columns

for each base make 1 column so it can be added and also put into a heatmap

```
#for the heatmap with every binding site
heatframeA <- do.call(rbind.data.frame, strsplit(resframeA$binding_nospace,""))
heatframeA <- sapply( heatframeA, as.numeric )
colnames(heatframeA) <- c(23:1)
rownames(heatframeA) <- resframeA[,1]
head(heatframeA)
```

```
##      23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
## 1004  0  0  0  0  1  1  1  1  1  1  1  1  0  0 0 1 1 1 1 1 1 1 0
## 1005  0  0  0  0  1  1  1  1  1  1  1  1  0  0 0 1 1 1 1 1 1 1 0
## 1035  0  0  0  0  0  0  0  0  0  0  0  0  0  0 0 0 1 1 1 1 1 1 0
## 1043  0  0  0  0  1  1  0  1  0  0  1  1  1  0 1 1 1 1 1 1 1 1 0
## 1050  0  0  0  0  0  0  1  1  1  1  0  1  1  0 1 1 1 0 1 1 1 1 1
## 1054  0  0  0  1  1  1  1  0  1  1  1  1  1  0 1 1 1 1 0 1 1 1 0
```

```
#reverse column order
heatframeA <-heatframeA[,23:1]
```

Heatmap

Colours

```
hmcols1 <- c("white", "black")
hmcols2 <- colorRamp2(c(-2, 2), c("white", "red"))
```

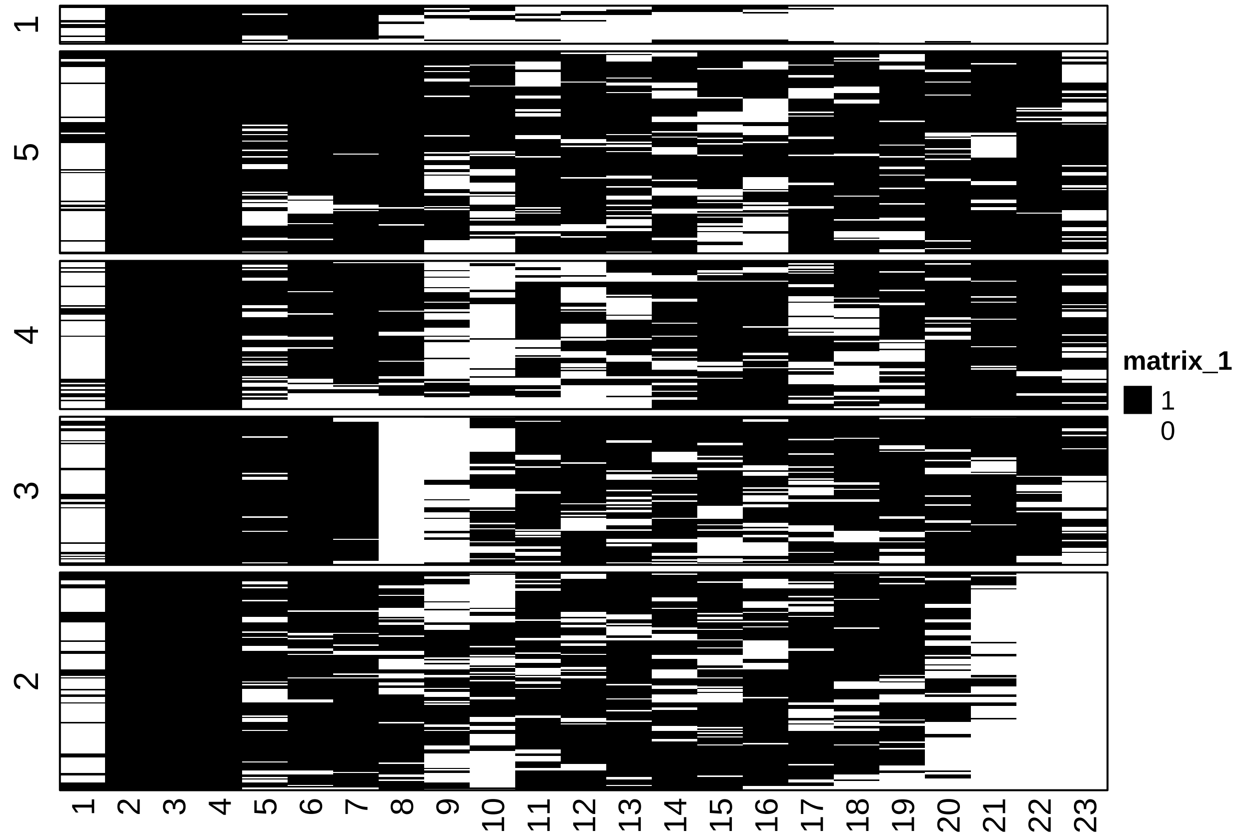
Heatmap of all the single reads

make heatmap without column clustering but with row clustering

```
set.seed(123)

HMA <- Heatmap(heatframeA, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_row
```

```
HMA
```



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] stringr_1.5.0                      seqinr_4.2-30
##  [3] ComplexHeatmap_2.15.2              circlize_0.4.15
##  [5] ggplot2_3.4.2                      dplyr_1.1.1
##  [7] BSgenome.Mmusculus.UCSC.mm10_1.4.3 BSgenome_1.66.3
##  [9] rtracklayer_1.58.0                 Biostrings_2.66.0
## [11] XVector_0.38.0                     GenomicRanges_1.50.2
## [13] GenomeInfoDb_1.34.9                IRanges_2.32.0
## [15] S4Vectors_0.36.2                   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] magick_2.7.4                zlibbioc_1.44.0
## [21] scales_1.2.1                BiocParallel_1.32.6
## [23] tibble_3.2.1                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               Cairo_1.6-0
## [35] tools_4.2.3                 GlobalOptions_0.1.2
## [37] BiocIO_1.8.0                lifecycle_1.0.3
## [39] matrixStats_0.63.0          mgsub_1.7.3
## [41] munsell_0.5.0               cluster_2.1.4
## [43] DelayedArray_0.23.2         ade4_1.7-22
## [45] compiler_4.2.3              rlang_1.1.0
## [47] RCurl_1.98-1.12             iterators_1.0.14
## [49] rstudioapi_0.14             rjson_0.2.21
## [51] bitops_1.0-7                rmarkdown_2.21
## [53] restfulr_0.0.15             gtable_0.3.3
## [55] codetools_0.2-19            R6_2.5.1
## [57] GenomicAlignments_1.34.1    knitr_1.42
## [59] fastmap_1.1.1               utf8_1.2.3
## [61] clue_0.3-64                 shape_1.4.6
## [63] stringi_1.7.12              parallel_4.2.3
## [65] Rcpp_1.0.10                 vctrs_0.6.1
## [67] png_0.1-8                   tidyselect_1.2.0
## [69] xfun_0.37
```