

RNAhybrid_complete_testing_c_vs_nc

Nikita Verheyden

2023-04-28

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
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)
library(gridExtra)

##
## Attaching package: 'gridExtra'

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

## The following object is masked from 'package:BiocGenerics':
##
##      combine
```

data

```
#home
f2bs <- readRDS("D:/Krueger_Lab/Publications/miR181_paper/Figure2/MRE_bound_gene_and_bound_region/mir181")
#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 Pcmt1 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	Pcmt1d1	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 <- "#008EAOFF"
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, end)))
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 Pcmt1 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 Pcmt1 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 UAAAGGACUGGACUCCUGGCCUCCUCAUUUUGUGUAUGUAUUUUUUUUUUUUAACUAGGACUAAAUUUUUUUUUUUU
## 2 CAAGAAUAGAAAGUACAACAGGCAUUACAACCACUACCUCACCAAAAAACUCCUCCUCCACUACUGUUCAGGACACCUUAUG
## 3 UACAAAAAGAACAGUGUGACUUAGCAAAUUAUUUAAAAUGUACAGCUGUAGAAAUAAGAAAUAAUUAUUGAAAAAGUAAAAUG
## 4 UAUUAUUGAAAAAGUAAAAUGUUCUCUAGAAAUAACACUAAAGGAAAAAGCAUCAGCAAGAACUCCAUCUUUAAAAAUUGAG
## 5 GGAAAAUUUCUGCUUCUCUCAUAGAGAUUUUUAAGAGCUAGUGAAUGUUAAAGUAGGAAGUGGCUACUUGACACAACUAGUU
## 6 GAAGUGUAAUAAAAUGCUACCAGAUGUUUUUUAAGGUGGUACCCACCAUAAUGUCUCUGUCACAUUUUAUAAUACAAAUG
##      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"))

#remove NAs in gene name
seed1 <- seed1[!is.na(seed1$geneName),]
seed2 <- seed2[!is.na(seed2$geneName),]

```



```
head(seed1)
```

##	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	Pcmdt1	ENSMUSG000000051285	utr3	19	mmu-miR-181a-5p					
## 7	protein_coding	Sgk3	ENSMUSG000000025915	utr3	24	mmu-miR-181a-5p					
## 22	protein_coding	Nck2	ENSMUSG000000066877	utr3	97	mmu-miR-181b-5p					
## 31	protein_coding	Cflar	ENSMUSG000000026031	utr3	130	mmu-miR-181a-5p					
## 51	protein_coding	Cab39	ENSMUSG000000036707	utr3	209	mmu-miR-181a-5p					
## 86	protein_coding	Tmem185b	ENSMUSG000000098923	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	ENSMUSG000000051285	Pcmdt1	utr3	12	22						
## 7	<NA>	<NA>	<NA>	NA	NA						
## 22	ENSMUSG000000066877	Nck2	utr3	11	15						
## 31	ENSMUSG000000026031	Cflar	utr3	6	14						
## 51	ENSMUSG000000036707	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	ENSMUSG000000051285	1355	1706	1064							
## 7	<NA>	NA	NA	NA							
## 22	ENSMUSG000000066877	5026	7988	4386							
## 31	ENSMUSG000000026031	1371	1785	1002							
## 51	ENSMUSG000000036707	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      ENSMUSG000000051285.bs4          3          3          TRUE FALSE
## 7      <NA>          NA          NA          NA    NA
## 22      ENSMUSG000000066877.bs31          3          3          TRUE FALSE
## 31      ENSMUSG000000026031.bs4          3          3          TRUE FALSE
## 51      ENSMUSG000000036707.bs4          3          3          TRUE  TRUE
## 86      <NA>          NA          NA          NA    NA
##
##                                     Sequence
## 5      GGAAAAUUUCUGCUUCUCUCAUAGAGAUUUUUAAGAGCUAGUGAAUGUUAAGUAGGAAGUGGCUACUUGACACAACUAGUU
## 7      GGCAAGUCUGGGUUGGUGUGAAUGUGUGUCACCUACACAUUCUAAACAGAAGGUAAACAUAAGUUAGCAGUGACAUUUCAGU
## 22      AUUAUUAUUUUGCUUUACAGGGAUUUUUUCAGGGUUUACAAAAGAAUAUGUGAUUAGUAGUAAACAGAAUGUUUAUGAAGAA
## 31      UGGGUGUAUAGUGUAUAGUGGUUCAAGAUUUGACACUGAAUGUAAACUUGAGACUUACCUGAGUUUGUCAUGCGACUGGGUAA
## 51      UGUUAUAUAAUUCUUAAGAAUGCUAUUUCUUUUAAAUUCGUUUAAUUGUACAGCAGAGGAAUGUUAUUGUAGUAGUAGUAAAC
## 86      UGCAUAUAUUAGUAUUUAUAUGAAUGUUUAGCAGUGUUAUCUGUGUUGAUUGUAGUUCUUGGCAGUAAUGUAUUGUGUUA
##      rownum
## 5      5
## 7      7
## 22     22
## 31     31
## 51     51
## 86     86

```

```
head(seed2)
```

```

##      seqnames      start      end width strand  scoreSum scoreMean scoreMax
## 28      chr1  58402990  58403071      7      +   38.09626  12.698753  17.08590
## 66      chr1  87792332  87792413      7      +   60.80440  15.201100  16.81080
## 84      chr1 118464515 118464596      7      +  102.33322  20.466644  40.96150
## 86      chr1 119528130 119528211      7      +   19.78365   9.891825  12.04230
## 92      chr1 127775116 127775197      7      +   11.83646   5.918230   6.29952

```

##	159	chr1	155173974	155174055	7	+	24.74386	12.371930	15.64410
##		geneType	geneName	geneID	region	BS_ID		mir_IP	
##	28	protein_coding	Bzw1	ENSMUSG000000051223	utr3	124	mmu-miR-181a-5p		
##	66	protein_coding	Atg16l1	ENSMUSG000000026289	utr3	244	mmu-miR-181a-5p		
##	84	protein_coding	Clasp1	ENSMUSG000000064302	cds	313	mmu-miR-181a-5p		
##	86	protein_coding	Tmem185b	ENSMUSG000000098923	utr3	320	mmu-miR-181a-5p		
##	92	protein_coding	Ccnt2	ENSMUSG000000026349	cds	347	mmu-miR-181a-5p		
##	159	protein_coding	Stx6	ENSMUSG000000026470	cds	568	mmu-miR-181a-5p		
##		n_mir181	n_mir181a	n_mir181b	n_mir181c	n_mir181d		set	WT KO
##	28	1	1	0	0	0	ago_bs_mir181_chi	1	1
##	66	1	1	0	0	0	ago_bs_mir181_chi	1	1
##	84	2	1	1	0	0	ago_bs_mir181_chi	1	1
##	86	7	7	0	0	0	ago_bs_mir181_chi	NA	NA
##	92	1	1	0	0	0	ago_bs_mir181_chi	1	0
##	159	3	2	1	0	0	ago_bs_mir181_chi	1	1
##		geneID.2	geneName.1	region.1	counts.bs.1_KO	counts.bs.2_KO			
##	28	ENSMUSG000000051223	Bzw1	utr3	9	8			
##	66	ENSMUSG000000026289	Atg16l1	utr3	8	9			
##	84	ENSMUSG000000064302	Clasp1	cds	25	26			
##	86	<NA>	<NA>	<NA>	NA	NA			
##	92	ENSMUSG000000026349	Ccnt2	cds	1	2			
##	159	ENSMUSG000000026470	Stx6	cds	3	9			
##		counts.bs.3_KO	counts.bs.4_WT	counts.bs.5_WT	counts.bs.6_WT				
##	28	5	17	10	5				
##	66	5	12	21	7				
##	84	15	29	25	12				
##	86	NA	NA	NA	NA				
##	92	1	7	5	5				
##	159	5	13	6	1				
##		geneID.1	counts.bg.1_KO	counts.bg.2_KO	counts.bg.3_KO				
##	28	ENSMUSG000000051223	392	541	311				
##	66	ENSMUSG000000026289	681	904	531				
##	84	ENSMUSG000000064302	2675	3791	2111				
##	86	<NA>	NA	NA	NA				
##	92	ENSMUSG000000026349	2418	3235	1941				
##	159	ENSMUSG000000026470	1807	2526	1509				
##		counts.bg.4_WT	counts.bg.5_WT	counts.bg.6_WT	resBs.baseMean				
##	28	645	475	255	72.76455				
##	66	1154	864	513	93.58266				
##	84	4231	3547	1802	249.76110				
##	86	NA	NA	NA	NA				
##	92	3490	3011	1706	112.28672				
##	159	2744	2338	1416	107.75352				
##		resBs.log2FoldChange	resBs.lfcSE	resBs.stat	resBs.pvalue	resBs.padj			
##	28	-0.2079555	0.4169909	0.25053210	0.61670066	0.8961239			
##	66	-0.4108259	0.3967434	1.09496574	0.29537357	0.7198473			
##	84	0.3513764	0.2635197	1.77460129	0.18281366	0.6019152			
##	86	NA	NA	NA	NA	NA			
##	92	-1.7732438	0.8054955	5.99595457	0.01433872	0.1776151			
##	159	0.1372395	0.4873718	0.07919813	0.77838704	0.9438527			
##		resBg.baseMean	resBg.log2FoldChange	resBg.lfcSE	resBg.stat	resBg.pvalue			
##	28	NA	NA	NA	NA	NA			
##	66	NA	NA	NA	NA	NA			
##	84	NA	NA	NA	NA	NA			

```

## 86          NA          NA          NA          NA          NA
## 92          NA          NA          NA          NA          NA
## 159         NA          NA          NA          NA          NA
##      resBg.padj tpm.counts.bg.1_K0 tpm.counts.bg.2_K0 tpm.counts.bg.3_K0
## 28          NA          118.66953          117.8524          117.69078
## 66          NA          86.35598          82.4901          84.17221
## 84          NA          149.51095          152.4723          147.49128
## 86          NA          NA          NA          NA          NA
## 92          NA          239.13690          230.2251          239.96324
## 159         NA          207.40555          208.6333          216.51119
##      tpm.counts.bg.4_WT tpm.counts.bg.5_WT tpm.counts.bg.6_WT
## 28          124.56121          113.44594          112.34934
## 66          93.35145          86.43731          94.67598
## 84          150.85586          156.40571          146.58218
## 86          NA          NA          NA
## 92          220.18395          234.93257          245.55370
## 159         200.91681          211.71368          236.53894
##      BS_ID.1 tpm_support_K0 tpm_support_WT tpm_supported down
## 28  ENSMUSG000000051223.bs3          3          3          TRUE FALSE
## 66  ENSMUSG000000026289.bs7          3          3          TRUE FALSE
## 84  ENSMUSG000000064302.bs14          3          3          TRUE FALSE
## 86          <NA>          NA          NA          NA    NA
## 92  ENSMUSG000000026349.bs2          3          3          TRUE FALSE
## 159 ENSMUSG000000026470.bs6          3          3          TRUE FALSE
##
##                                     Sequence
## 28  AAAAAAGAGGAGCUUGUAGCAGAGCAAGCCAUCAAGCACUUGAAGGUACUGGGAUUGUGAAGCUGUCACAGUUGGGGUGGGGU
## 66  CGCUGGCCAGGGCACAUUUUAUUUAUUUAUUUAUGUGUCCCAACAGAACUUGAUUGUAAAUAAGAAGAAAUUCUGUUUAU
## 84  CUUUCGUUCCAGCUCUGGGGCCAGACUCUAACACUAAGCAAGAUUGUGCCACAUUAUGUAACCUACUGGGAGAUCCCAA
## 86  UGCAUAUAUUAUGAUUUUAUUGAAUGUUUAGCAGUGUUAUCUGUGUUGAUUGUAGUUCUUGGCAGUAAUGUAUUGUGUAA
## 92  AAACACUGCGAUUGUUUAUUGCACAGGUUUUAUUGCACCAUCCUUCACCAAAUUAACAGAAUUGUAAGUACUGUUUA
## 159 CCUUGACUACAGAGAGGUACAGAAAGCAGUCAACACUGCCCAAGGAUUGUUUCAGAGAUGGACAGAGCUCCUUCAGGGCCCC
##
##      rownum
## 28      28
## 66      66
## 84      84
## 86      86
## 92      92
## 159     159

```

Write to .fasta

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

```

candgeneNameAc <- as.list(seed1$geneName)
candnameAc <- as.list(seed1$rownum)
condgeneSeqAc <- as.list(seed1$Sequence)

candgeneNameAnc <- as.list(seed2$geneName)
candnameAnc <- as.list(seed2$rownum)
condgeneSeqAnc <- as.list(seed2$Sequence)

```

#change to output directory

```

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

for (i in 1:length(candgeneNameAc)) {
  write.fasta(condgeneSeqAc[i], candnameAc[i], paste(candnameAc[i], candgeneNameAc[i], "miR_181a", 'fastafiles_complete/Ac'),
}

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

for (i in 1:length(candgeneNameAnc)) {
  write.fasta(condgeneSeqAnc[i], candnameAnc[i], paste(candnameAnc[i], candgeneNameAnc[i], "miR_181a",
}

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
reslistAc0 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r
reslistAc1 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r
reslistAc2 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r
reslistAc3 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r
reslistAc4 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r
reslistAc5 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/r

reslistAnc0 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/
reslistAnc1 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/
reslistAnc2 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/
reslistAnc3 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/
reslistAnc4 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/
reslistAnc5 <- list.files(path = "D:/Krueger_Lab/Publications/miR181_paper_nongithub/Figure2/RNAhybrid/

#import
myfilelistAc0 <- lapply(reslistAc0, Personalized_Reader)
myfilelistAc1 <- lapply(reslistAc1, Personalized_Reader)
myfilelistAc2 <- lapply(reslistAc2, Personalized_Reader)
myfilelistAc3 <- lapply(reslistAc3, Personalized_Reader)
myfilelistAc4 <- lapply(reslistAc4, Personalized_Reader)
myfilelistAc5 <- lapply(reslistAc5, Personalized_Reader)

myfilelistAnc0 <- lapply(reslistAnc0, Personalized_Reader)
myfilelistAnc1 <- lapply(reslistAnc1, Personalized_Reader)
myfilelistAnc2 <- lapply(reslistAnc2, Personalized_Reader)
myfilelistAnc3 <- lapply(reslistAnc3, Personalized_Reader)
myfilelistAnc4 <- lapply(reslistAnc4, Personalized_Reader)
myfilelistAnc5 <- lapply(reslistAnc5, Personalized_Reader)

resframeAc0 <- bind_rows(myfilelistAc0)
resframeAc1 <- bind_rows(myfilelistAc1)
resframeAc2 <- bind_rows(myfilelistAc2)
resframeAc3 <- bind_rows(myfilelistAc3)

```

```

resframeAc4 <- bind_rows(myfilelistAc4)
resframeAc5 <- bind_rows(myfilelistAc5)

resframeAnc0 <- bind_rows(myfilelistAnc0)
resframeAnc1 <- bind_rows(myfilelistAnc1)
resframeAnc2 <- bind_rows(myfilelistAnc2)
resframeAnc3 <- bind_rows(myfilelistAnc3)
resframeAnc4 <- bind_rows(myfilelistAnc4)
resframeAnc5 <- bind_rows(myfilelistAnc5)

#colnames
colnames(resframeAc0) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAc1) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAc2) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAc3) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAc4) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAc5) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")

colnames(resframeAnc0) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAnc1) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAnc2) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAnc3) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAnc4) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")
colnames(resframeAnc5) <- c("rownumber", "mfs", "pvalue", "start_position", "binding_bases", "non_binding_bases")

resframeAc0[is.na(resframeAc0$non_binding_bases), "non_binding_bases"] <- " "
resframeAc1[is.na(resframeAc1$non_binding_bases), "non_binding_bases"] <- " "
resframeAc2[is.na(resframeAc2$non_binding_bases), "non_binding_bases"] <- " "
resframeAc3[is.na(resframeAc3$non_binding_bases), "non_binding_bases"] <- " "
resframeAc4[is.na(resframeAc4$non_binding_bases), "non_binding_bases"] <- " "
resframeAc5[is.na(resframeAc5$non_binding_bases), "non_binding_bases"] <- " "

resframeAnc0[is.na(resframeAnc0$non_binding_bases), "non_binding_bases"] <- " "
resframeAnc1[is.na(resframeAnc1$non_binding_bases), "non_binding_bases"] <- " "
resframeAnc2[is.na(resframeAnc2$non_binding_bases), "non_binding_bases"] <- " "
resframeAnc3[is.na(resframeAnc3$non_binding_bases), "non_binding_bases"] <- " "
resframeAnc4[is.na(resframeAnc4$non_binding_bases), "non_binding_bases"] <- " "
resframeAnc5[is.na(resframeAnc5$non_binding_bases), "non_binding_bases"] <- " "

```

merge with original df

```

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

seed1$rownumber <- as.character(seed1$rownum)
seed2$rownumber <- as.character(seed2$rownum)

resframeAc0$rownumber <- as.character(resframeAc0$rownumber)
resframeAc1$rownumber <- as.character(resframeAc1$rownumber)
resframeAc2$rownumber <- as.character(resframeAc2$rownumber)
resframeAc3$rownumber <- as.character(resframeAc3$rownumber)
resframeAc4$rownumber <- as.character(resframeAc4$rownumber)

```

```

resframeAc5$rownumber <- as.character(resframeAc5$rownumber)

resframeAnc0$rownumber <- as.character(resframeAnc0$rownumber)
resframeAnc1$rownumber <- as.character(resframeAnc1$rownumber)
resframeAnc2$rownumber <- as.character(resframeAnc2$rownumber)
resframeAnc3$rownumber <- as.character(resframeAnc3$rownumber)
resframeAnc4$rownumber <- as.character(resframeAnc4$rownumber)
resframeAnc5$rownumber <- as.character(resframeAnc5$rownumber)

bsseqHAc0 <- left_join(seed1, resframeAc0, by="rownumber")
bsseqHAc1 <- left_join(seed1, resframeAc1, by="rownumber")
bsseqHAc2 <- left_join(seed1, resframeAc2, by="rownumber")
bsseqHAc3 <- left_join(seed1, resframeAc3, by="rownumber")
bsseqHAc4 <- left_join(seed1, resframeAc4, by="rownumber")
bsseqHAc5 <- left_join(seed1, resframeAc5, by="rownumber")

bsseqHAnc0 <- left_join(seed2, resframeAnc0, by="rownumber")
bsseqHAnc1 <- left_join(seed2, resframeAnc1, by="rownumber")
bsseqHAnc2 <- left_join(seed2, resframeAnc2, by="rownumber")
bsseqHAnc3 <- left_join(seed2, resframeAnc3, by="rownumber")
bsseqHAnc4 <- left_join(seed2, resframeAnc4, by="rownumber")
bsseqHAnc5 <- left_join(seed2, resframeAnc5, by="rownumber")

```

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 the mirna in the correct order. For that we will remove all gaps that originate in the mRNA loops.

```

#binding and non binding bases as characters in a list
Alistbbc0 <- strsplit(resframeAc0$binding_bases, "")
Alistbbc1 <- strsplit(resframeAc1$binding_bases, "")
Alistbbc2 <- strsplit(resframeAc2$binding_bases, "")
Alistbbc3 <- strsplit(resframeAc3$binding_bases, "")
Alistbbc4 <- strsplit(resframeAc4$binding_bases, "")
Alistbbc5 <- strsplit(resframeAc5$binding_bases, "")

Alistbbnc0 <- strsplit(resframeAnc0$binding_bases, "")
Alistbbnc1 <- strsplit(resframeAnc1$binding_bases, "")
Alistbbnc2 <- strsplit(resframeAnc2$binding_bases, "")
Alistbbnc3 <- strsplit(resframeAnc3$binding_bases, "")
Alistbbnc4 <- strsplit(resframeAnc4$binding_bases, "")
Alistbbnc5 <- strsplit(resframeAnc5$binding_bases, "")

Alistnbc0 <- strsplit(resframeAc0$non_binding_bases, "")
Alistnbc1 <- strsplit(resframeAc1$non_binding_bases, "")
Alistnbc2 <- strsplit(resframeAc2$non_binding_bases, "")
Alistnbc3 <- strsplit(resframeAc3$non_binding_bases, "")
Alistnbc4 <- strsplit(resframeAc4$non_binding_bases, "")
Alistnbc5 <- strsplit(resframeAc5$non_binding_bases, "")

```

```

Alistnbnc0 <- strsplit(resframeAnc0$non_binding_bases,"")
Alistnbnc1 <- strsplit(resframeAnc1$non_binding_bases,"")
Alistnbnc2 <- strsplit(resframeAnc2$non_binding_bases,"")
Alistnbnc3 <- strsplit(resframeAnc3$non_binding_bases,"")
Alistnbnc4 <- strsplit(resframeAnc4$non_binding_bases,"")
Alistnbnc5 <- strsplit(resframeAnc5$non_binding_bases,"")

#combine the two lists
Alistc0 <- Map(cbind, Alistbbc0, Alistnbnc0)

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

## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
Alistc1 <- Map(cbind, Alistbbc1, Alistnbnc1)

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

## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
Alistc2 <- Map(cbind, Alistbbc2, Alistnbnc2)

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

## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
Alistc3 <- Map(cbind, Alistbbc3, Alistnbnc3)

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

## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
Alistc4 <- Map(cbind, Alistbbc4, Alistnbnc4)

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

## Warning in cbind(...): number of rows of result is not a multiple of vector
## length (arg 2)
Alistc5 <- Map(cbind, Alistbbc5, Alistnbnc5)

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

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

```



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

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

```
Alistnc0 <- Map(cbind, Alistbbnc0, Alistnbnc0)
Alistnc1 <- Map(cbind, Alistbbnc1, Alistnbnc1)
Alistnc2 <- Map(cbind, Alistbbnc2, Alistnbnc2)
Alistnc3 <- Map(cbind, Alistbbnc3, Alistnbnc3)
Alistnc4 <- Map(cbind, Alistbbnc4, Alistnbnc4)
Alistnc5 <- Map(cbind, Alistbbnc5, Alistnbnc5)
```

```
Alistc0 <- lapply(Alistc0, as.data.frame)
Alistc1 <- lapply(Alistc1, as.data.frame)
Alistc2 <- lapply(Alistc2, as.data.frame)
Alistc3 <- lapply(Alistc3, as.data.frame)
Alistc4 <- lapply(Alistc4, as.data.frame)
Alistc5 <- lapply(Alistc5, as.data.frame)
```

```
Alistnc0 <- lapply(Alistnc0, as.data.frame)
Alistnc1 <- lapply(Alistnc1, as.data.frame)
Alistnc2 <- lapply(Alistnc2, as.data.frame)
Alistnc3 <- lapply(Alistnc3, as.data.frame)
Alistnc4 <- lapply(Alistnc4, as.data.frame)
Alistnc5 <- lapply(Alistnc5, as.data.frame)
```

```
#remove all empty rows (mRNA loops)
```

```
Alist0c0 <- lapply(Alistc0, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0c1 <- lapply(Alistc1, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0c2 <- lapply(Alistc2, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0c3 <- lapply(Alistc3, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0c4 <- lapply(Alistc4, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0c5 <- lapply(Alistc5, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0nc0 <- lapply(Alistnc0, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0nc1 <- lapply(Alistnc1, function(x){
  x[!(x[,1]== " " & x[,2] == " "),]
})
```

```
Alist0nc2 <- lapply(Alistnc2, function(x){
```

```

  x[!(x[,1]== " " & x[,2] == " " ),]
})
Alist0nc3 <- lapply(Alistnc3, function(x){
  x[!(x[,1]== " " & x[,2] == " " ),]
})
Alist0nc4 <- lapply(Alistnc4, function(x){
  x[!(x[,1]== " " & x[,2] == " " ),]
})
Alist0nc5 <- lapply(Alistnc5, function(x){
  x[!(x[,1]== " " & x[,2] == " " ),]
})

#rewrite as characters
AlistFc0 <- lapply(Alist0c0, function(x){
  paste(x[,1], collapse = '')
})
AlistFc1 <- lapply(Alist0c1, function(x){
  paste(x[,1], collapse = '')
})
AlistFc2 <- lapply(Alist0c2, function(x){
  paste(x[,1], collapse = '')
})
AlistFc3 <- lapply(Alist0c3, function(x){
  paste(x[,1], collapse = '')
})
AlistFc4 <- lapply(Alist0c4, function(x){
  paste(x[,1], collapse = '')
})
AlistFc5 <- lapply(Alist0c5, function(x){
  paste(x[,1], collapse = '')
})

AlistFnc0 <- lapply(Alist0nc0, function(x){
  paste(x[,1], collapse = '')
})
AlistFnc1 <- lapply(Alist0nc1, function(x){
  paste(x[,1], collapse = '')
})
AlistFnc2 <- lapply(Alist0nc2, function(x){
  paste(x[,1], collapse = '')
})
AlistFnc3 <- lapply(Alist0nc3, function(x){
  paste(x[,1], collapse = '')
})
AlistFnc4 <- lapply(Alist0nc4, function(x){
  paste(x[,1], collapse = '')
})
AlistFnc5 <- lapply(Alist0nc5, function(x){
  paste(x[,1], collapse = '')
})

#Attach lists back onto original data.frame as new column

```

```

resframeAc0$binding_nospace <-unlist(AlistFc0)
resframeAc1$binding_nospace <-unlist(AlistFc1)
resframeAc2$binding_nospace <-unlist(AlistFc2)
resframeAc3$binding_nospace <-unlist(AlistFc3)
resframeAc4$binding_nospace <-unlist(AlistFc4)
resframeAc5$binding_nospace <-unlist(AlistFc5)

resframeAnc0$binding_nospace <-unlist(AlistFnc0)
resframeAnc1$binding_nospace <-unlist(AlistFnc1)
resframeAnc2$binding_nospace <-unlist(AlistFnc2)
resframeAnc3$binding_nospace <-unlist(AlistFnc3)
resframeAnc4$binding_nospace <-unlist(AlistFnc4)
resframeAnc5$binding_nospace <-unlist(AlistFnc5)

```

Transform into Numbers

add 0s

replace all gaps with 0 and all letters with 1

#0

```

resframeAc0$binding_nospace <- chartr(" ", "0", resframeAc0$binding_nospace)
resframeAc1$binding_nospace <- chartr(" ", "0", resframeAc1$binding_nospace)
resframeAc2$binding_nospace <- chartr(" ", "0", resframeAc2$binding_nospace)
resframeAc3$binding_nospace <- chartr(" ", "0", resframeAc3$binding_nospace)
resframeAc4$binding_nospace <- chartr(" ", "0", resframeAc4$binding_nospace)
resframeAc5$binding_nospace <- chartr(" ", "0", resframeAc5$binding_nospace)

resframeAnc0$binding_nospace <- chartr(" ", "0", resframeAnc0$binding_nospace)
resframeAnc1$binding_nospace <- chartr(" ", "0", resframeAnc1$binding_nospace)
resframeAnc2$binding_nospace <- chartr(" ", "0", resframeAnc2$binding_nospace)
resframeAnc3$binding_nospace <- chartr(" ", "0", resframeAnc3$binding_nospace)
resframeAnc4$binding_nospace <- chartr(" ", "0", resframeAnc4$binding_nospace)
resframeAnc5$binding_nospace <- chartr(" ", "0", resframeAnc5$binding_nospace)

```

#1

```

resframeAc0$binding_nospace <- mgsub::mgsub(resframeAc0$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAc1$binding_nospace <- mgsub::mgsub(resframeAc1$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAc2$binding_nospace <- mgsub::mgsub(resframeAc2$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAc3$binding_nospace <- mgsub::mgsub(resframeAc3$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAc4$binding_nospace <- mgsub::mgsub(resframeAc4$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAc5$binding_nospace <- mgsub::mgsub(resframeAc5$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))

resframeAnc0$binding_nospace <- mgsub::mgsub(resframeAnc0$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAnc1$binding_nospace <- mgsub::mgsub(resframeAnc1$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAnc2$binding_nospace <- mgsub::mgsub(resframeAnc2$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAnc3$binding_nospace <- mgsub::mgsub(resframeAnc3$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAnc4$binding_nospace <- mgsub::mgsub(resframeAnc4$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))
resframeAnc5$binding_nospace <- mgsub::mgsub(resframeAnc5$binding_nospace, c("A", "U", "C", "G"), c(rep("1", 10), rep("0", 10)))

```

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

```
heatframeAc0 <- do.call(rbind.data.frame, strsplit(resframeAc0$binding_nospace,""))
heatframeAc1 <- do.call(rbind.data.frame, strsplit(resframeAc1$binding_nospace,""))
heatframeAc2 <- do.call(rbind.data.frame, strsplit(resframeAc2$binding_nospace,""))
heatframeAc3 <- do.call(rbind.data.frame, strsplit(resframeAc3$binding_nospace,""))
heatframeAc4 <- do.call(rbind.data.frame, strsplit(resframeAc4$binding_nospace,""))
heatframeAc5 <- do.call(rbind.data.frame, strsplit(resframeAc5$binding_nospace,""))
```

```
heatframeAnc0 <- do.call(rbind.data.frame, strsplit(resframeAnc0$binding_nospace,""))
heatframeAnc1 <- do.call(rbind.data.frame, strsplit(resframeAnc1$binding_nospace,""))
heatframeAnc2 <- do.call(rbind.data.frame, strsplit(resframeAnc2$binding_nospace,""))
heatframeAnc3 <- do.call(rbind.data.frame, strsplit(resframeAnc3$binding_nospace,""))
heatframeAnc4 <- do.call(rbind.data.frame, strsplit(resframeAnc4$binding_nospace,""))
heatframeAnc5 <- do.call(rbind.data.frame, strsplit(resframeAnc5$binding_nospace,""))
```

```
heatframeAc0 <- sapply( heatframeAc0, as.numeric )
heatframeAc1 <- sapply( heatframeAc1, as.numeric )
heatframeAc2 <- sapply( heatframeAc2, as.numeric )
heatframeAc3 <- sapply( heatframeAc3, as.numeric )
heatframeAc4 <- sapply( heatframeAc4, as.numeric )
heatframeAc5 <- sapply( heatframeAc5, as.numeric )
```

```
heatframeAnc0 <- sapply( heatframeAnc0, as.numeric )
heatframeAnc1 <- sapply( heatframeAnc1, as.numeric )
heatframeAnc2 <- sapply( heatframeAnc2, as.numeric )
heatframeAnc3 <- sapply( heatframeAnc3, as.numeric )
heatframeAnc4 <- sapply( heatframeAnc4, as.numeric )
heatframeAnc5 <- sapply( heatframeAnc5, as.numeric )
```

```
colnames(heatframeAc0) <- c(23:1)
colnames(heatframeAc1) <- c(23:1)
colnames(heatframeAc2) <- c(23:1)
colnames(heatframeAc3) <- c(23:1)
colnames(heatframeAc4) <- c(23:1)
colnames(heatframeAc5) <- c(23:1)
```

```
colnames(heatframeAnc0) <- c(23:1)
colnames(heatframeAnc1) <- c(23:1)
colnames(heatframeAnc2) <- c(23:1)
colnames(heatframeAnc3) <- c(23:1)
colnames(heatframeAnc4) <- c(23:1)
colnames(heatframeAnc5) <- c(23:1)
```

```
rownames(heatframeAc0) <- resframeAc0[,1]
rownames(heatframeAc1) <- resframeAc1[,1]
rownames(heatframeAc2) <- resframeAc2[,1]
rownames(heatframeAc3) <- resframeAc3[,1]
rownames(heatframeAc4) <- resframeAc4[,1]
rownames(heatframeAc5) <- resframeAc5[,1]
```

```

rownames(heatframeAnc0) <- resframeAnc0[,1]
rownames(heatframeAnc1) <- resframeAnc1[,1]
rownames(heatframeAnc2) <- resframeAnc2[,1]
rownames(heatframeAnc3) <- resframeAnc3[,1]
rownames(heatframeAnc4) <- resframeAnc4[,1]
rownames(heatframeAnc5) <- resframeAnc5[,1]

#reverse column order
heatframeAc0 <-heatframeAnc0[,23:1]
heatframeAc1 <-heatframeAnc1[,23:1]
heatframeAc2 <-heatframeAnc2[,23:1]
heatframeAc3 <-heatframeAnc3[,23:1]
heatframeAc4 <-heatframeAnc4[,23:1]
heatframeAc5 <-heatframeAnc5[,23:1]

heatframeAnc0 <-heatframeAnc0[,23:1]
heatframeAnc1 <-heatframeAnc1[,23:1]
heatframeAnc2 <-heatframeAnc2[,23:1]
heatframeAnc3 <-heatframeAnc3[,23:1]
heatframeAnc4 <-heatframeAnc4[,23:1]
heatframeAnc5 <-heatframeAnc5[,23:1]

```

Heatmap

Colours

```
hmcols1 <- c("white", "black")
```

Heatmap of all the single reads

make heatmap without column clustering but with row clustering

```

set.seed(123)
HMAc0 <- Heatmap(heatframeAc0, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAc1 <- Heatmap(heatframeAc1, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAc2 <- Heatmap(heatframeAc2, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAc3 <- Heatmap(heatframeAc3, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAc4 <- Heatmap(heatframeAc4, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAc5 <- Heatmap(heatframeAc5, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)

set.seed(123)
HMAnc0 <- Heatmap(heatframeAnc0, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAnc1 <- Heatmap(heatframeAnc1, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAnc2 <- Heatmap(heatframeAnc2, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAnc3 <- Heatmap(heatframeAnc3, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)

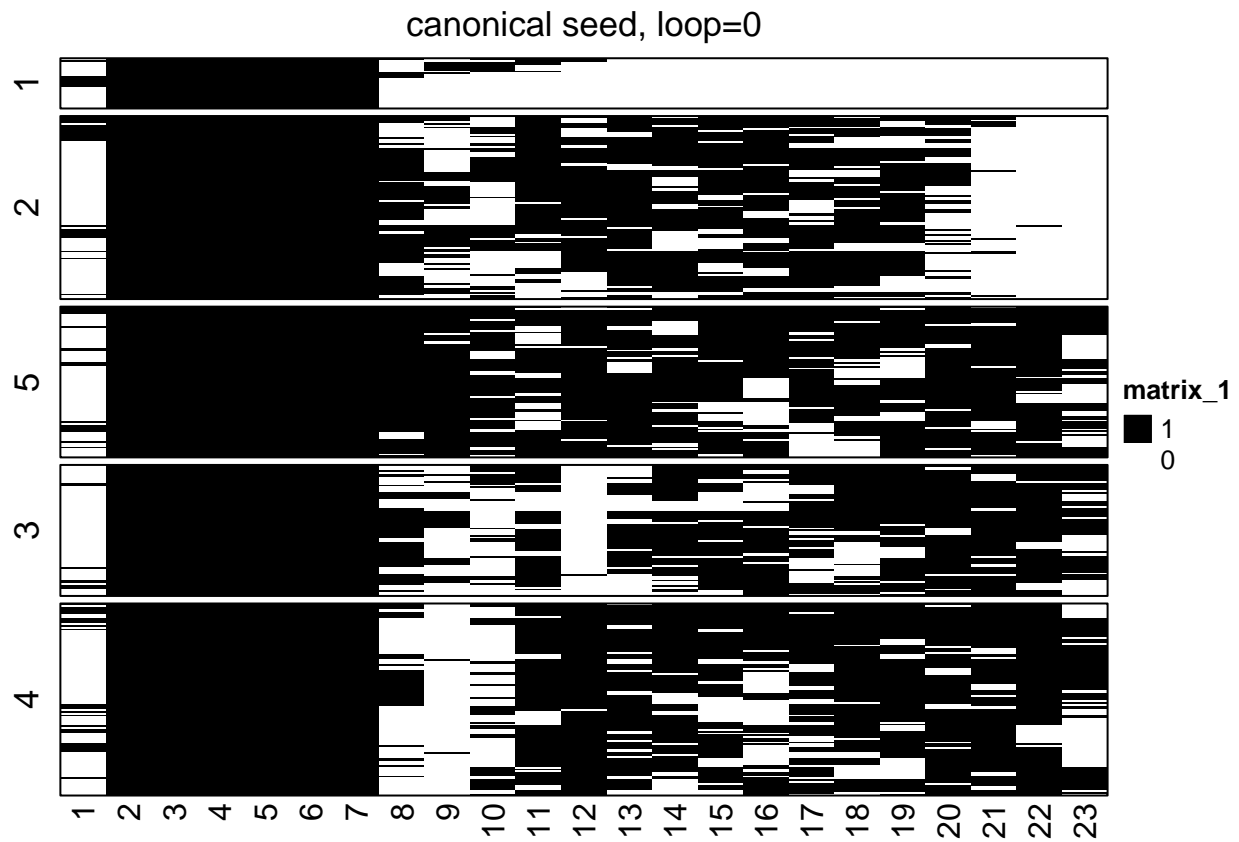
```

```

set.seed(123)
HMAnc4 <- Heatmap(heatframeAnc4, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)
set.seed(123)
HMAnc5 <- Heatmap(heatframeAnc5, cluster_columns = F, col = hmcols1, row_km = 5, show_row_names = F, show_col_names = F)

# canonical seed, loop=0
HMAc0

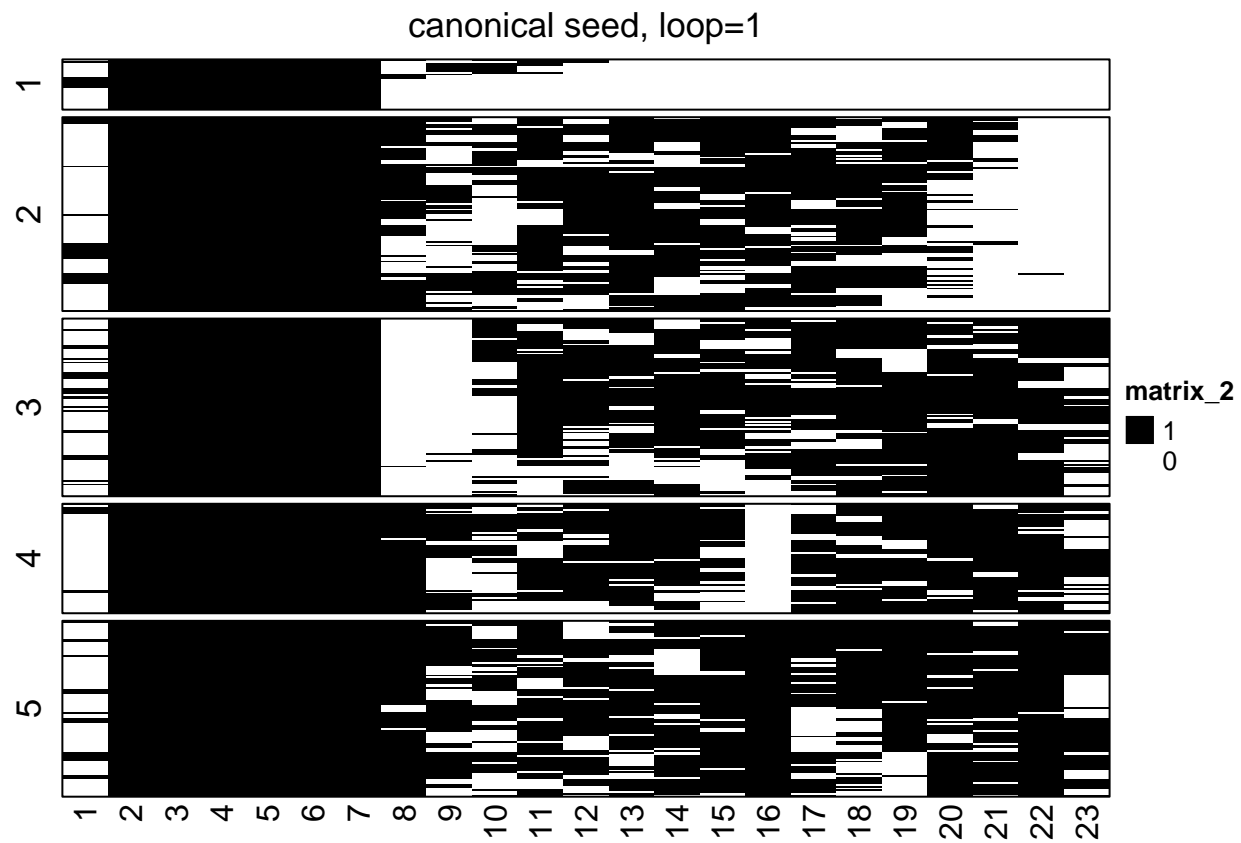
```



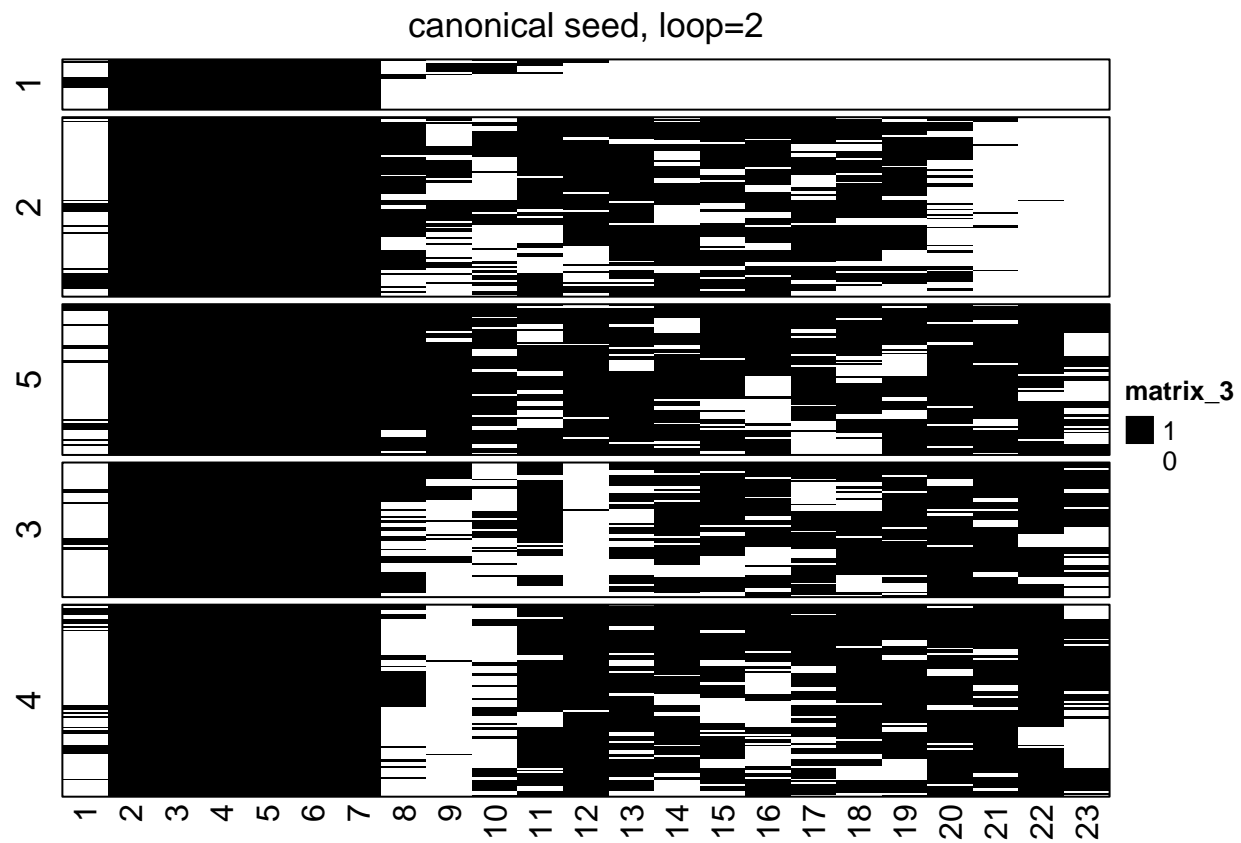
```

# canonical seed, loop=1
HMAc1

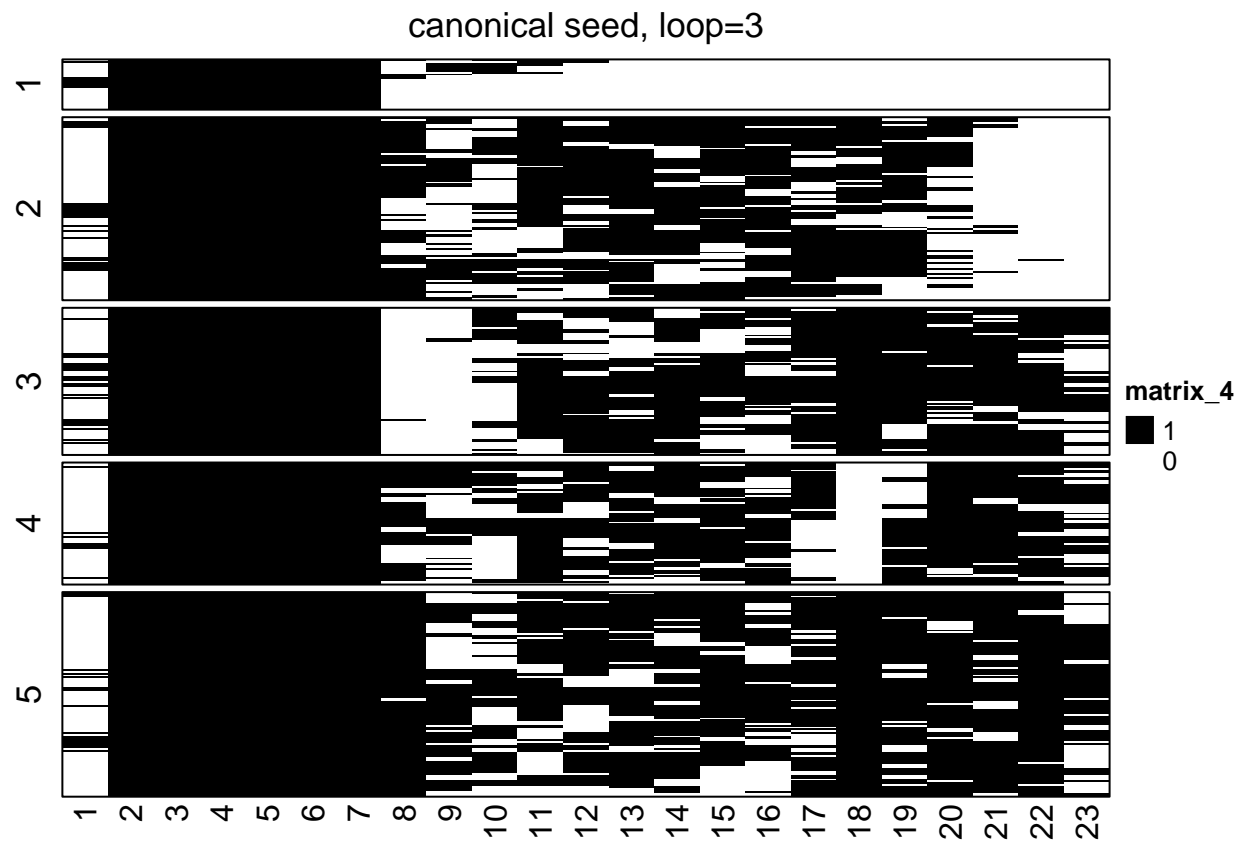
```



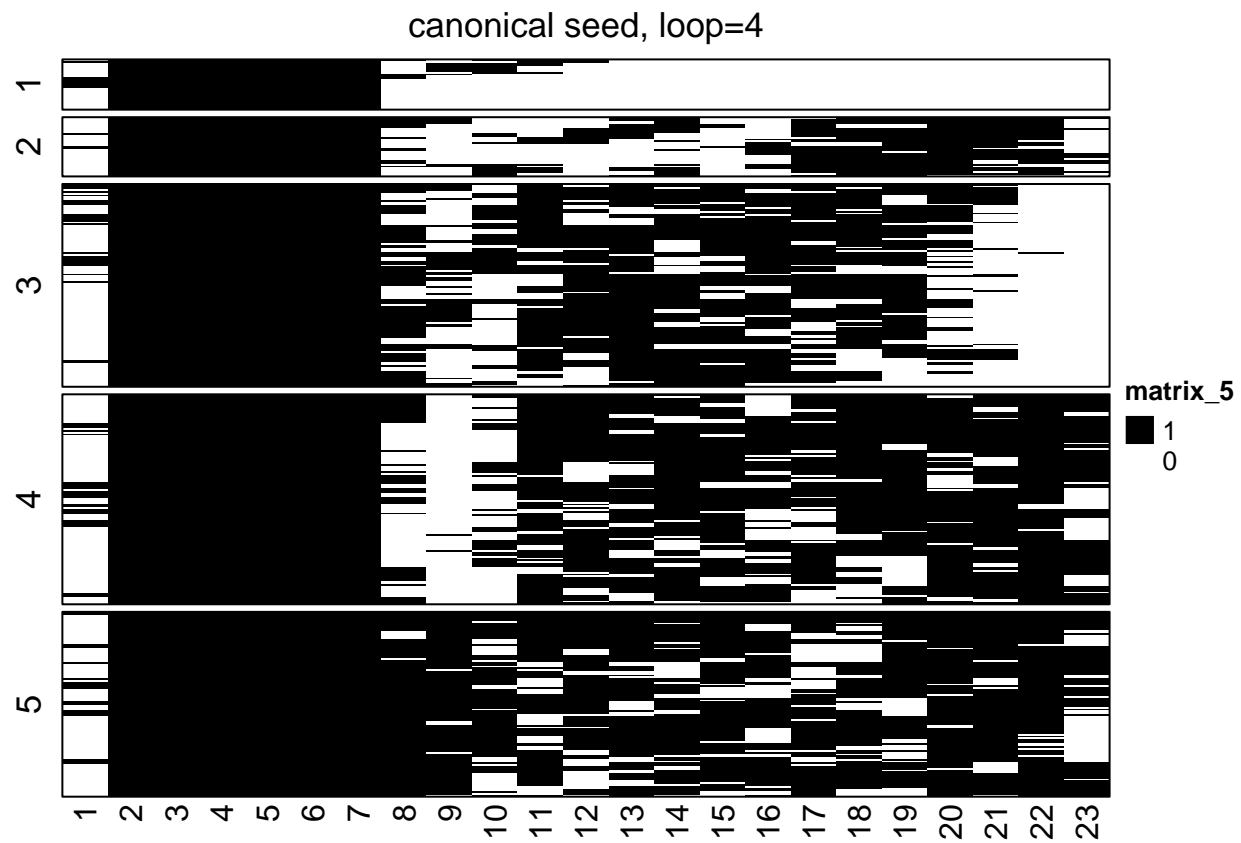
```
# canonical seed, loop=2
HMAC2
```



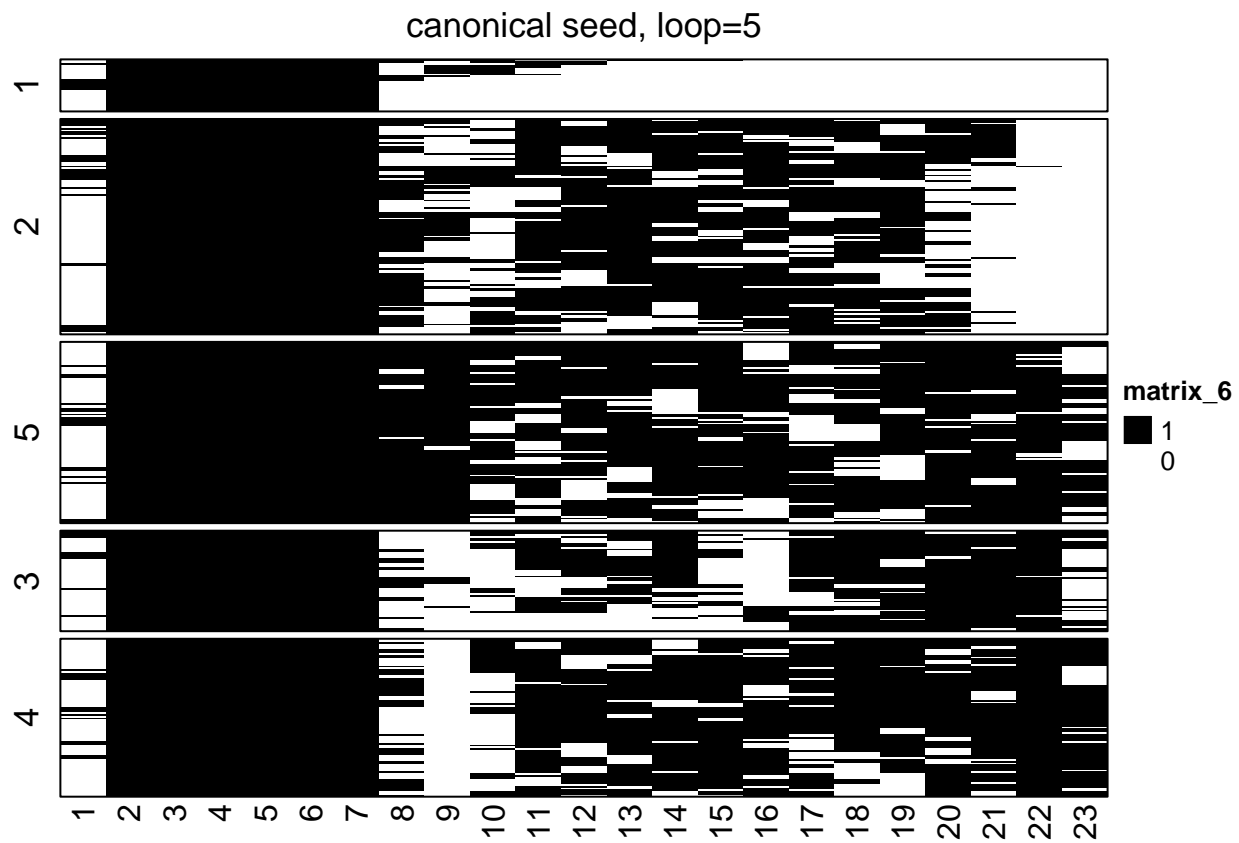
```
# canonical seed, loop=3
HMAC3
```

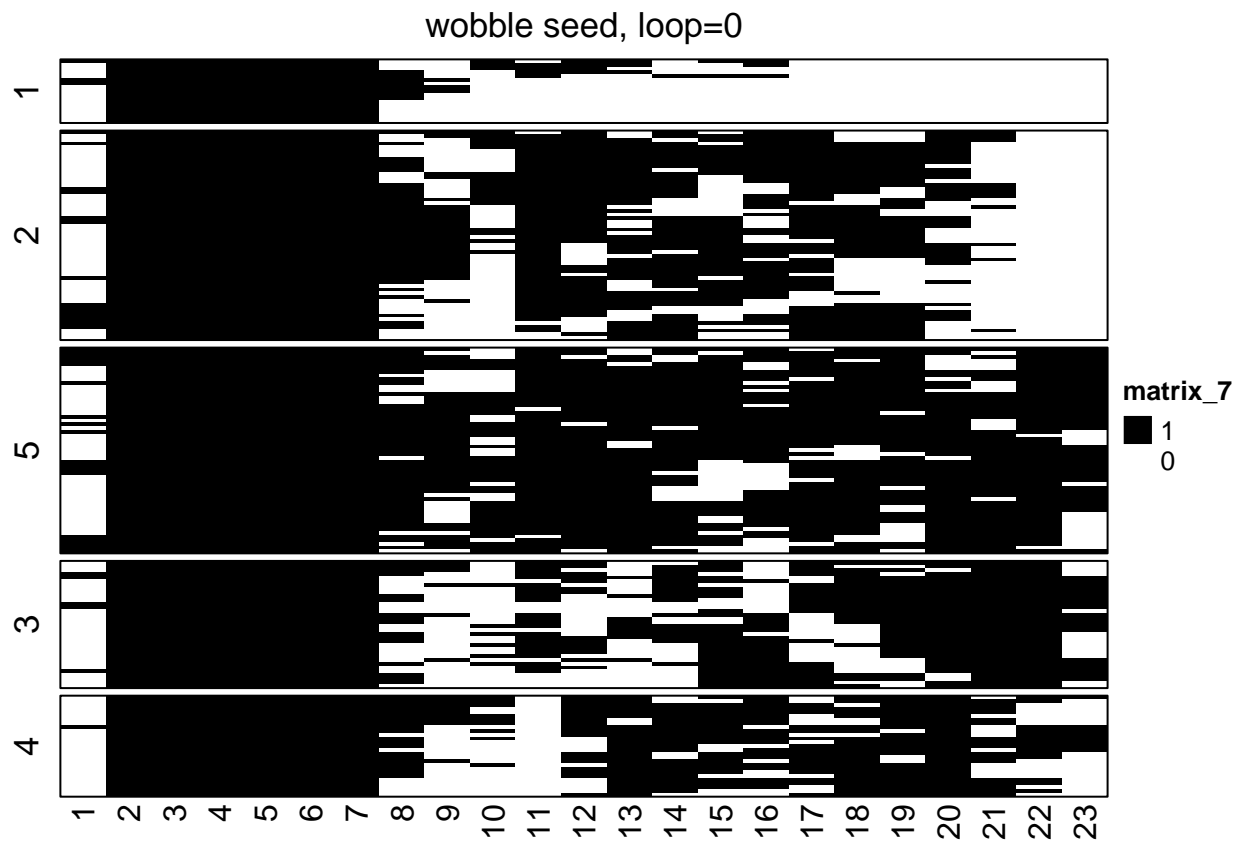
```
# canonical seed, loop=4
HMAC4
```



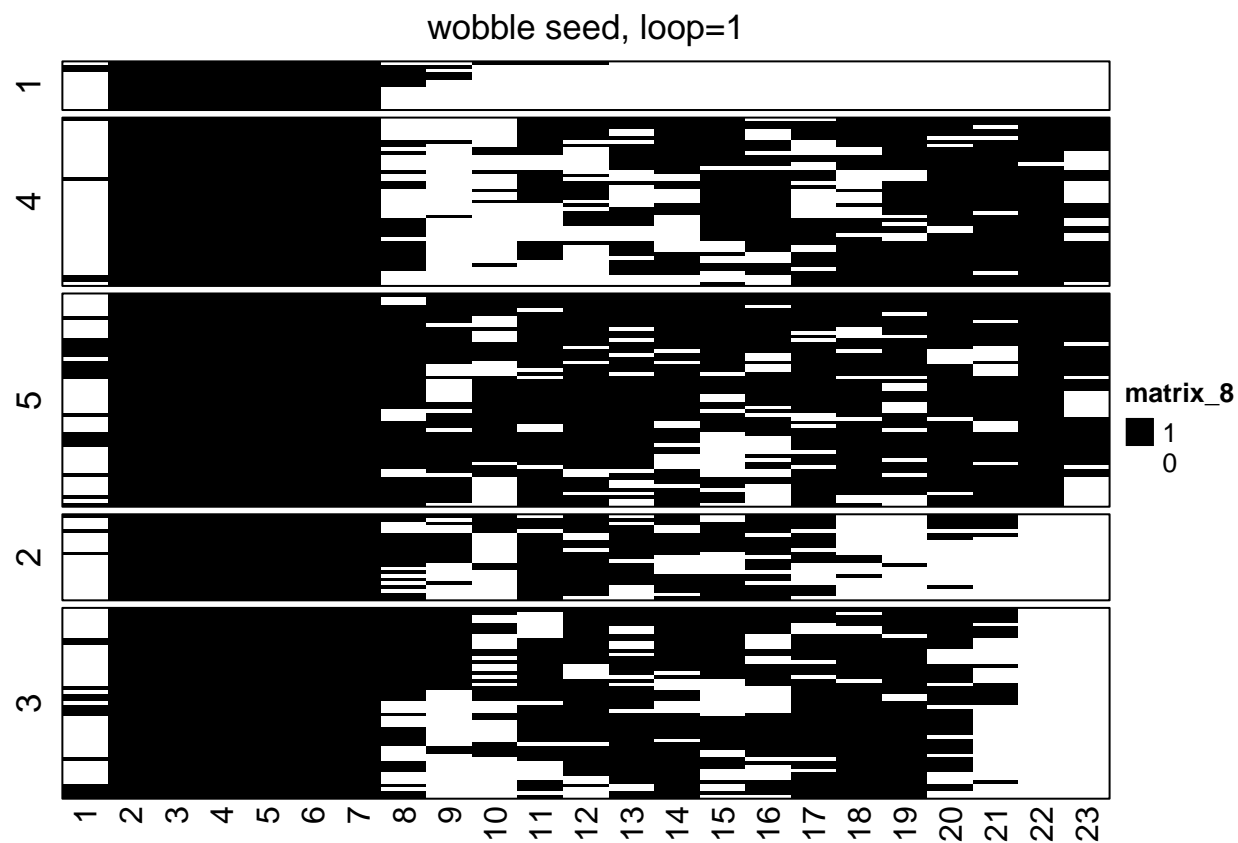
```
# canonical seed, loop=5
HMAC5
```



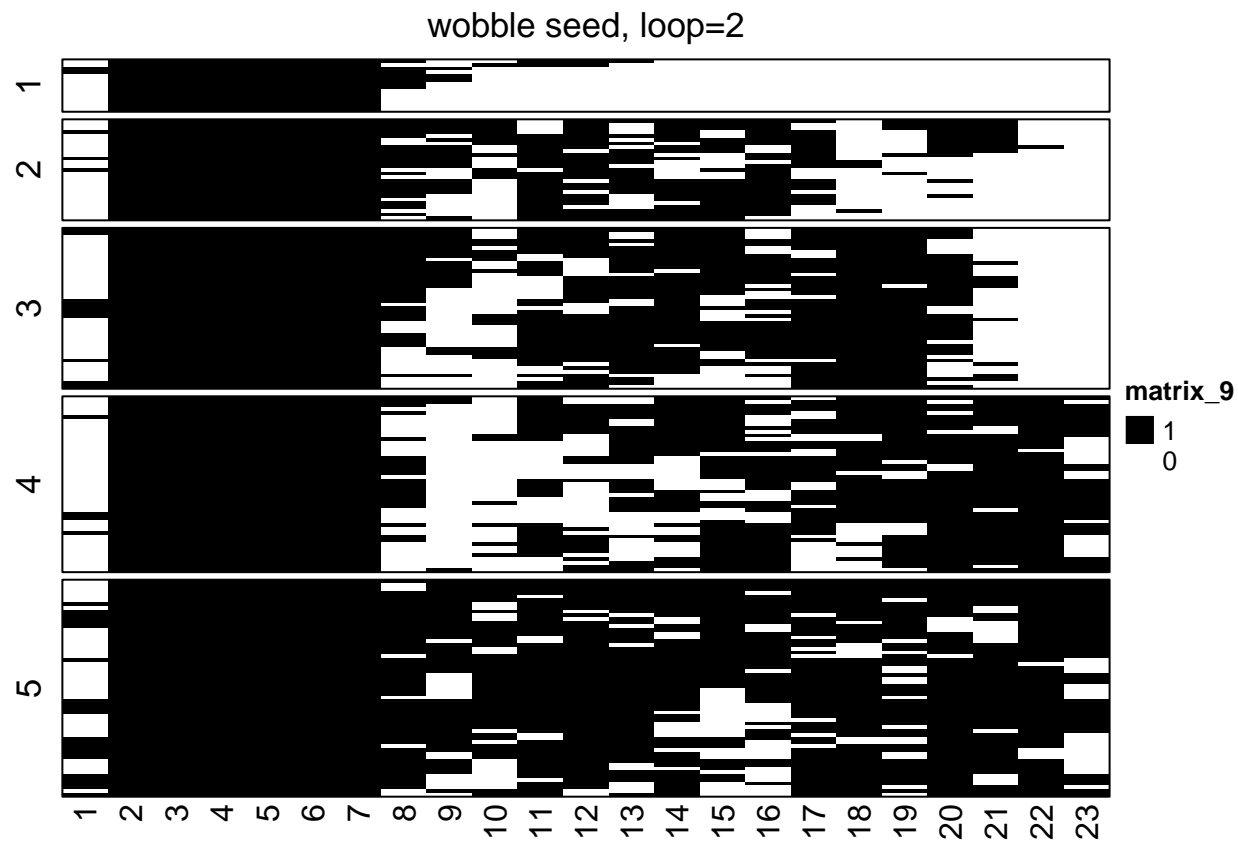
```
# wobble seed, loop=0
HMAnc0
```



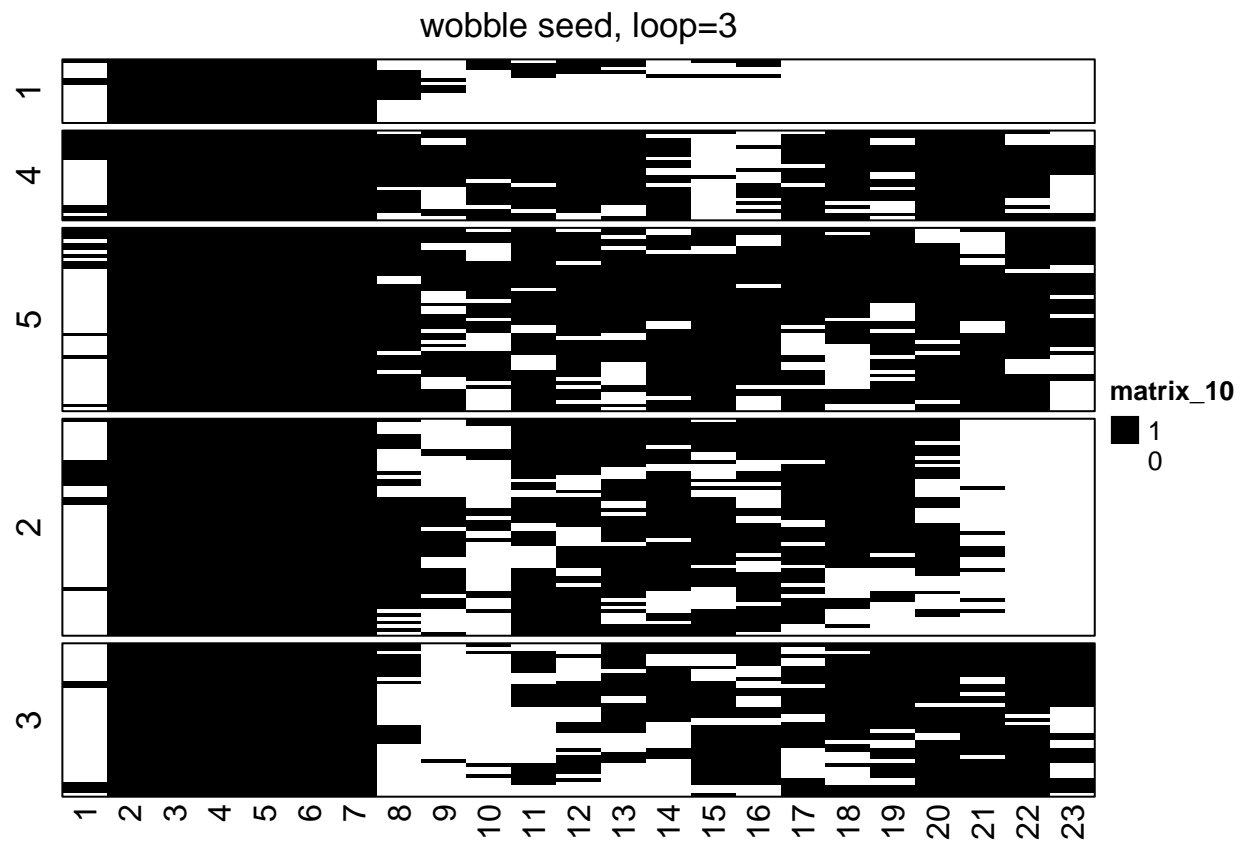
```
# wobble seed, loop=1
HMAnc1
```



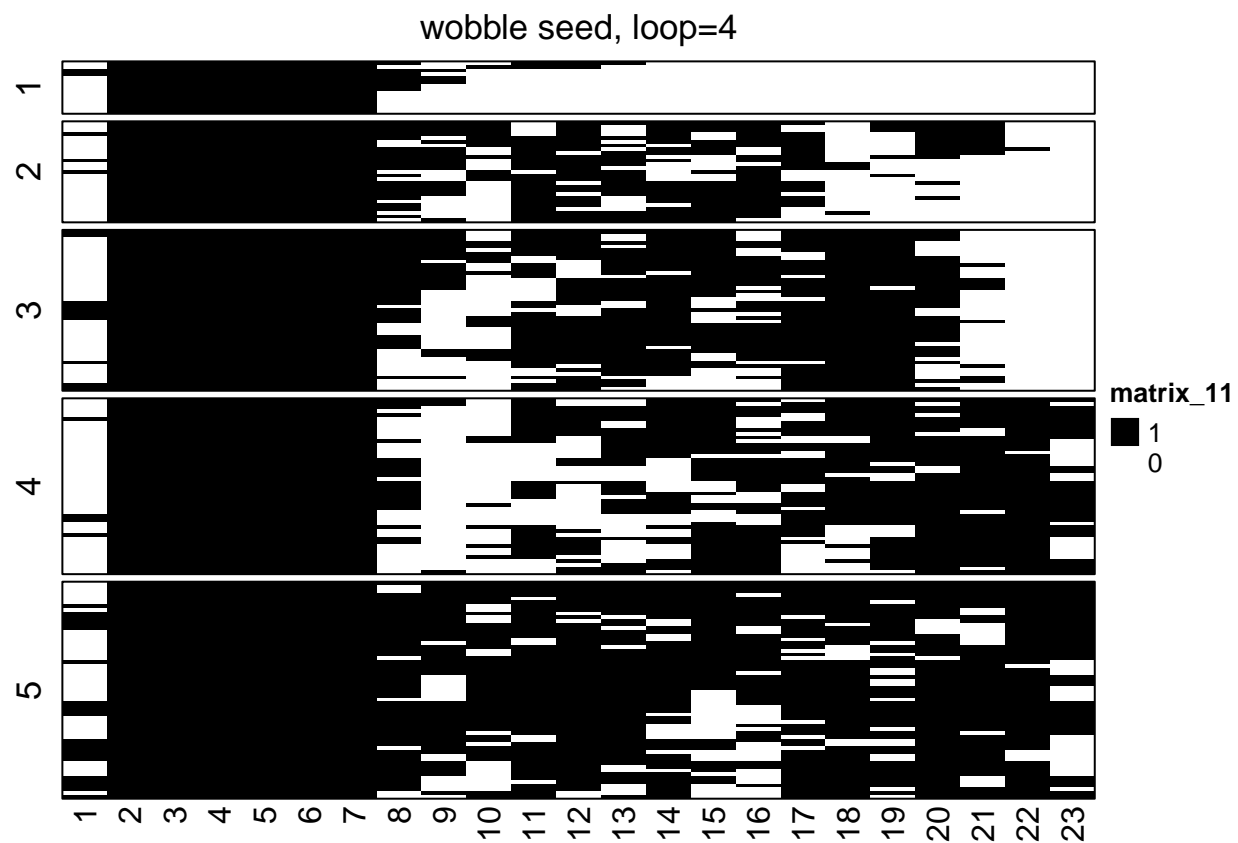
```
# wobble seed, loop=2
HMAnc2
```



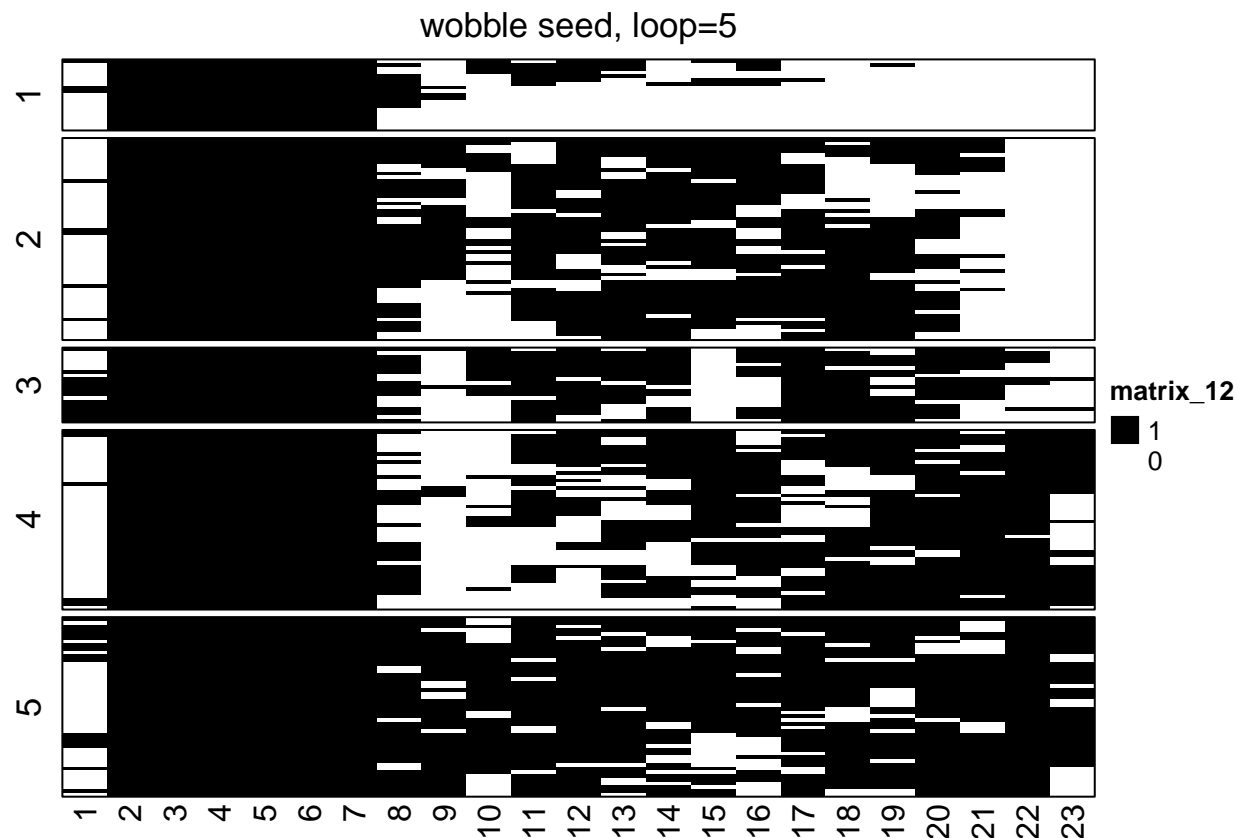
```
# wobble seed, loop=3
HMAnc3
```



```
# wobble seed, loop=4
HMAnc4
```



```
# wobble seed, loop=5
HMAnc5
```

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] gridExtra_2.3          stringr_1.5.0
## [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] 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

```