

Figure 6 E-F

RAC

04/09/2022

#read in count data

```
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
df <- read.table("../data/RBM7_3CLIP.rRNAScaled.hg38_snoRNAs_and_host_introns.100ntupdown.binN201.18",
colnames(df) <- c("sample", "chr", "start", "stop", "info", "dist",
"strand", "value")
head(df)
```

```
##           sample chr  start  stop
## 1 RBM7_1_DMSO_3endOfRead2  1 1304176 1304177
## 2 RBM7_1_DMSO_3endOfRead2  1 1304792 1304793
## 3 RBM7_1_DMSO_3endOfRead2  1 1304792 1304793
## 4 RBM7_1_DMSO_3endOfRead2  1 8466025 8466026
## 5 RBM7_1_DMSO_3endOfRead2  1 8494917 8494918
## 6 RBM7_1_DMSO_3endOfRead2  1 8511963 8511964
##                                     info dist
## 1 SNORD167:::NR145806:::intronic:::ACAP3:::protein_coding:::intron3ss -33
## 2                               SNORD167:::NR145806:::intronic:::mature3end -64
## 3                               SNORD167:::NR145806:::intronic:::mature5end  19
## 4 SNORD128:::NR132752:::intronic:::RERE:::protein_coding:::intron3ss  -2
## 5                               SNORD128:::NR132752:::intronic:::mature5end  -5
## 6 ENSG00000221083:::SNORA77:::snoRNA:::intronic:::mature5end -45
## strand value
## 1      - 4.6454
## 2      - 4.6454
## 3      - 4.6454
## 4      - 4.6454
## 5      - 4.6454
## 6      - 4.6454
```

```
snoRNAAnno <- read.table("../data/snoRNAs.GRCh38andrefGene.mature.bed")
```

```
colnames(snoRNAAnno) <- c("chr", "start", "stop", "info", "score",
"strand")
```

```
snoRNAAnno <- separate(snoRNAAnno, info, c("ENSG", "snoRNA"),
sep = ":::")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 1102 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(snoRNAAnno)
```

```
##   chr  start  stop      ENSG  snoRNA score strand
## 1   1 1304728 1304812 SNORD167 NR145806      .    -
```

```
## 2 1 8494799 8494913 SNORD128 NR132752 . -
## 3 1 8511795 8511919 ENSG00000221083 SNORA77 . -
## 4 1 9082696 9082879 ENSG00000252404 SCARNA16 . +
## 5 1 12221148 12221271 ENSG00000252969 SNORA70 . -
## 6 1 12507246 12507397 ENSG00000239149 SNORA59A . +
```

```
snoRNA_list <- df %>%
  filter(dist %in% c(1:50) & grepl("mature3end", info) & !grepl("U8::|U3::",
    info)) %>%
  group_by(info) %>%
  summarise(sum = sum(value)) %>%
  arrange(desc(sum)) %>%
  head(n = 100) %>%
  select(info) %>%
  separate(info, c("ENSG", "snoRNA", "annotation", "location",
    "position"), sep = ":::") %>%
  unite("snoRNA", c(ENSG, snoRNA), sep = ":::")
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 8 rows [17, 29,
## 46, 54, 66, 69, 76, 84].
```

```
exclude_list <- snoRNAAnno %>%
  select(ENSG, snoRNA) %>%
  add_count(ENSG, snoRNA) %>%
  filter(n > 1) %>%
  select(snoRNA) %>%
  unique()
```

```
exclude_string <- paste0(exclude_list$snoRNA, collapse = "|")
```

```
df_1 <- df %>%
  select(-c(start, stop, chr, strand)) %>%
  filter(grepl("mature3end", info) & !grepl(exclude_string,
    info)) %>%
  separate(info, c("ENSG", "snoRNA", "annotation", "location",
    "position"), sep = ":::") %>%
  left_join(snoRNAAnno) %>%
  unite("snoRNA", c(ENSG, snoRNA), sep = ":::") %>%
  unite("position", c(annotation, location, position), sep = ":::") %>%
  mutate(sample = gsub("_trimmed_clean_fastqAligned", "", sample)) %>%
  mutate(snoRNAtype = case_when(grepl("SNORD|snoU|U3|U8|snoZ6|snoMBII-202",
    snoRNA) ~ "CDbox", grepl("SNORA|ACA", snoRNA) ~ "HACAbbox",
    grepl("SCARNA", snoRNA) ~ "SCARNA", TRUE ~ "")) %>%
  spread(dist, value, fill = 0) %>%
  gather("dist", "value", 10:210) %>%
  separate(sample, c("protein", "rep", "timepoint", "readType"),
    sep = "_") %>%
  mutate(timepoint = factor(timepoint, c("t00", "t05", "t10",
    "t15", "t20", "t40", "t60", "DMSO"))) %>%
  mutate(readType = "3CLIP")
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 8173 rows [1,
## 149, 150, 151, 152, 153, 154, 167, 469, 470, 471, 498, 499, 500, 501, 502, 503,
## 504, 505, 506, ...].
```

```
## Joining, by = c("ENSG", "snoRNA")
```

Sup Fig. 6 A

Individual snoRNA coverage plots over selected snoRNAs

```
chosen_snoRNAs <- snoRNA_list %>%
  filter(grepl("SNORA52|SNORA80B|SNORD43|SNORD83A", snoRNA))

L = unique(chosen_snoRNAs$snoRNA)
length(L)

## [1] 4

plot_list = list()
for (i in 1:length(L)) {

  SNORNA <- as.character(L[i])
  print(SNORNA)
  saveName <- gsub(":::", "-", SNORNA)

  plot <- df_1 %>%
    mutate(timepoint = factor(timepoint, c("t00", "t05",
      "t10", "t15", "t20", "t40", "t60", "DMS0"))) %>%
    mutate(readType = "3CLIP") %>%
    filter(dist %in% c(-5:30) & timepoint != "negative" &
      grepl(SNORNA, snoRNA)) %>%
    mutate(name = gsub(".*:", "", snoRNA)) %>%
    ggplot(aes(y = value, x = as.numeric(dist), fill = timepoint)) +
    geom_bar(stat = "summary") + geom_point(aes(shape = rep),
      size = 0.5, alpha = 0.5, stat = "identity") + facet_grid(timepoint ~
      name) + theme_bw(base_size = 10, base_family = "Helvetica") +
    xlab(NULL) + ylab(NULL) + theme(text = element_text(size = 10),
      legend.position = "none") + scale_y_continuous(n.breaks = 4) +
    theme(legend.position = "none", strip.text.y.right = element_text(angle = 0))

  plot_list[[i]] = plot
}

## [1] "ENSG00000206633:::SNORA80B"
## [1] "ENSG00000199785:::SNORA52"
## [1] "ENSG00000263764:::SNORD43"
## [1] "ENSG00000209482:::SNORD83A"

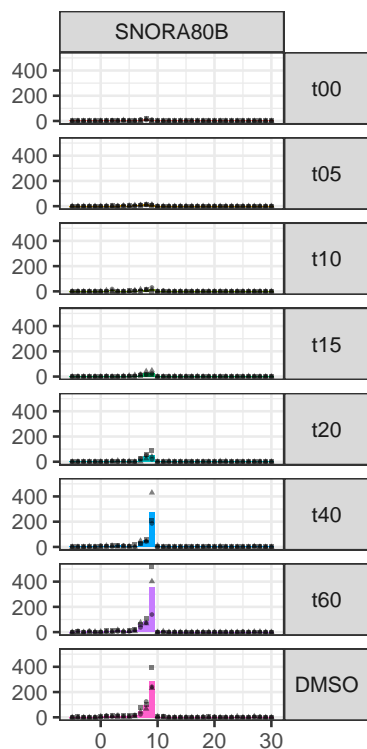
length(plot_list)

## [1] 4

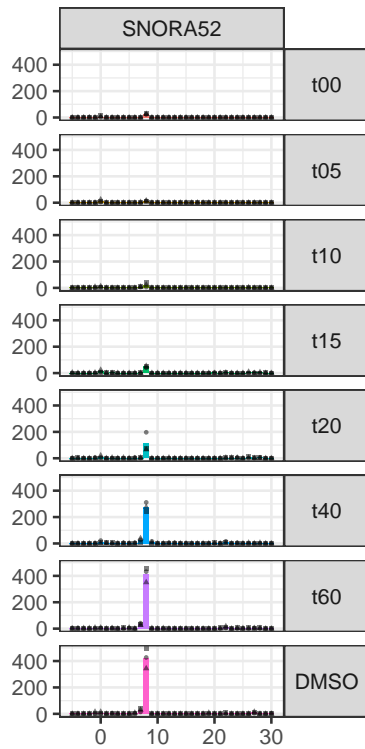
for (i in 1:length(plot_list)) {
  SNORNA <- L[i]
  saveName <- gsub(":::", "-", SNORNA)
  file_name = paste0("plots/", saveName, ".pdf")
  # pdf(file_name, height = 4.0, width = 2.0)
  print(plot_list[[i]])
  # dev.off()
}

## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```

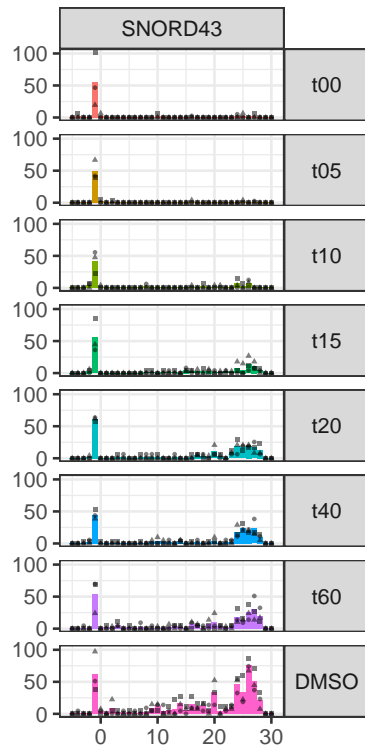
```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```



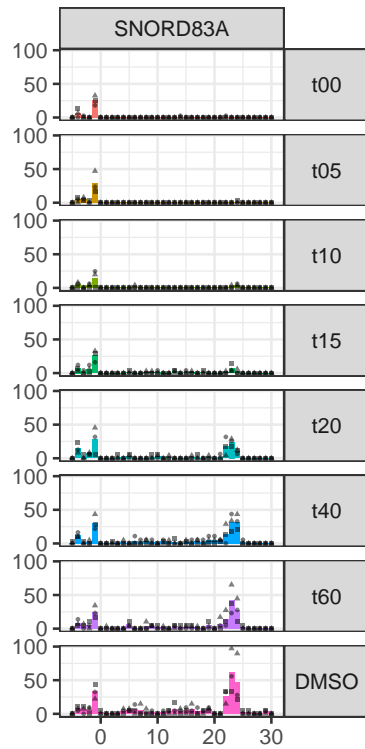
```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```



```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```



```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```



#Fig 6 e-f #make heatmaps depicting RBM7 3'CLIP signals downstream of

CD- and HACA-box snoRNAs

```
SNORNATYPE = c("HACAbbox", "CDbbox")
plot_list2 = list()
for (i in 1:length(SNORNATYPE)) {
  df2 <- df_1 %>%
    # separate(sample, c('protein', 'rep', 'timepoint',
    # 'readType'), sep = '_') %>%
    mutate(timepoint = factor(timepoint, c("negative", "t00",
      "t05", "t10", "t15", "t20", "t40", "t60", "DMSO"))) %>%
    mutate(readType = "3CLIP") %>%
    filter(dist %in% c(-5:30) & grepl("mature3end", position) &
      !grepl("nonintronic", position) & protein == "RBM7" &
      !grepl("negative", timepoint)) %>%
    spread(dist, value, fill = 0) %>%
    mutate(row_wise_sum = rowSums(.[13:48])) %>%
    filter(snoRNAtype == SNORNATYPE[i] & !grepl("U|u", snoRNA)) %>%

  arrange(desc(row_wise_sum))

  totals <- df2 %>%
    filter(timepoint == "DMSO") %>%
    group_by(snoRNA) %>%
    summarise(totals = sum(row_wise_sum)) %>%
    ungroup() %>%
    arrange(desc(totals)) %>%
    head(n = 50)

  plot <- df2 %>%
    filter(timepoint == "DMSO") %>%
    gather("dist", "value", 13:48) %>%

```

```

mutate(pct = value/row_wise_sum) %>%
group_by(protein, timepoint, readType, snoRNA, position,
chr, start, stop, score, strand, snoRNAtype, dist) %>%
summarise(pct_avg = sum(pct)) %>%
ungroup() %>%
spread(dist, pct_avg, fill = 0) %>%
mutate(row_wise_sum = rowSums(.[12:47])) %>%
gather("dist", "value", 12:47) %>%
mutate(pct = value/row_wise_sum) %>%
right_join(totals) %>%
arrange(desc(totals)) %>%
mutate(name = gsub(":::NR[0-9]*", "", snoRNA)) %>%
mutate(name = gsub(".*:", "", name)) %>%
ggplot(aes(x = as.numeric(dist), y = reorder(name, totals),
fill = pct)) + geom_raster() + scale_fill_distiller(palette = "Spectral",
direction = -1) + facet_grid(. ~ snoRNAtype) + theme_bw() +
theme(text = element_text(size = 8), legend.position = "none") +
xlab("") + ylab("")

plot_list2[[i]] = plot
# ggsave(paste0('top50_', SNORNATYPE, 'DMSO_3CLIP.pdf'),
# height = 8, width = 2)
}

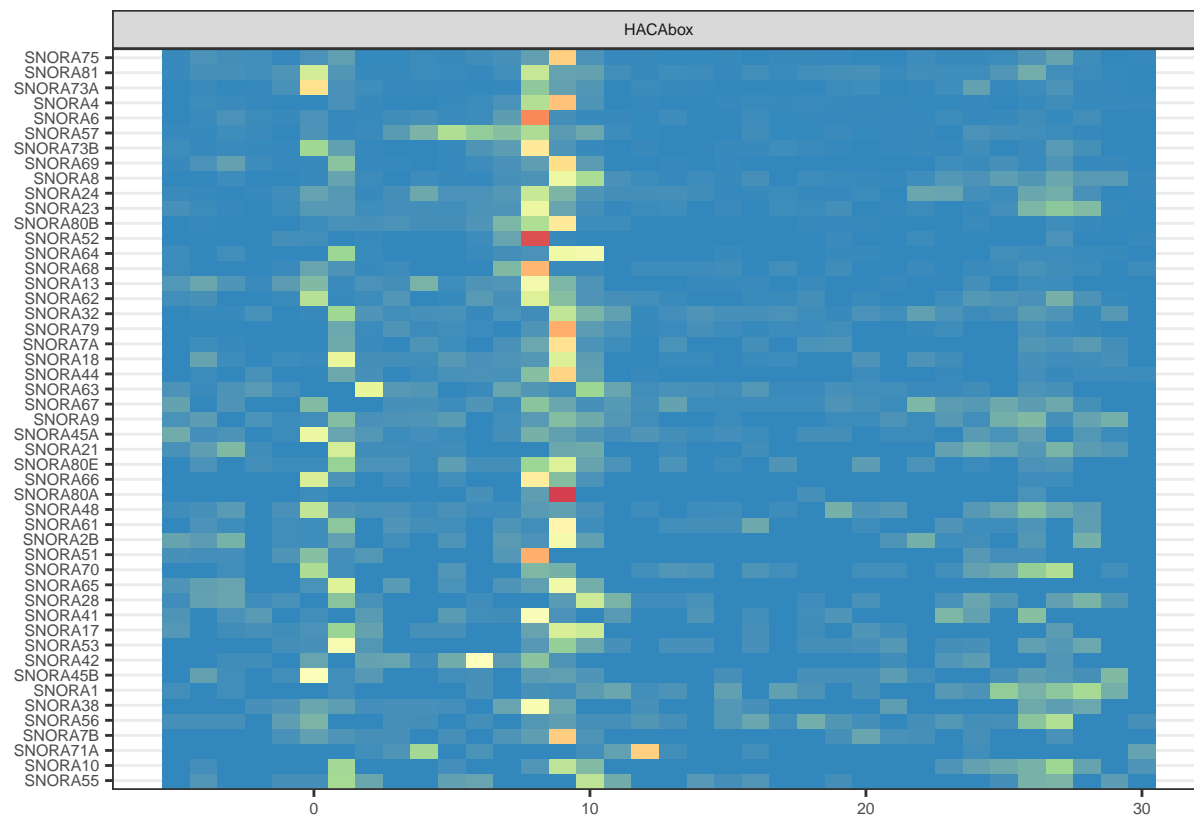
## `summarise()` has grouped output by 'protein', 'timepoint', 'readType',
## 'snoRNA', 'position', 'chr', 'start', 'stop', 'score', 'strand', 'snoRNAtype'.
## You can override using the `.groups` argument.
## Joining, by = "snoRNA"
## `summarise()` has grouped output by 'protein', 'timepoint', 'readType',
## 'snoRNA', 'position', 'chr', 'start', 'stop', 'score', 'strand', 'snoRNAtype'.
## You can override using the `.groups` argument.
## Joining, by = "snoRNA"

for (i in 1:length(plot_list2)) {

file_name = paste0("top50_", SNORNATYPE[i], "DMSO_3CLIP.pdf")
# pdf(file_name, height = 5.5, width = 1.5)
print(plot_list2[i])
# dev.off()
}

## [[1]]

```

##

[[1]]

