Untitled

RAC

03/09/2022

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
library(ggplot2)
```

#Supplementary Figure 6b

3'Seq generated from pA+ and pA- ends, in HeLa cells depleted of RRP40. Below shows how the data was processed from counts files. Counts files were generated from bigwig files downloaded from GSE137612. Specifically: GSM4083151, GSM4083150, GSM4083149, GSM4083132, GSM4083131 and GSM4083130. Bigwig files were converted to bedgraph files, then intersected with a snoRNA annotation file (bed). Below is the average of 3 biological replicates.

```r
df<-read.csv("../../data/RNAi_GSE137612-snoRNAs.GRCh38andrefGene.mature.3end.100ntupdown.binN201.sense.


colnames(df)<-c("sample","chr","start","stop","info","dist", "strand", "value")




RNAi<-
df %>%
  separate(info, c("ENSG", "snoRNA", "annotation", "location", "position"), sep = ":::") %>%
  unite("snoRNA", c(ENSG, snoRNA), sep = ":::") %>%
  unite("position", c(annotation, location, position), sep = ":::") %>%
  mutate(sample = gsub("batch","", sample)) %>%
  select(-c(chr,start,stop, strand)) %>%
  mutate(snoRNAtype = case_when(
    grepl("SNORD|snoU|U3|U8|snoZ6|snoMBII-202", snoRNA) ~ "CDbox",
    grepl("SNORA|ACA", snoRNA) ~ "HACAbox",
    grepl("SCARNA", snoRNA) ~ "SCARNA",
    TRUE ~ ""
```

```r
)) %>%
  filter(grepl("CDbox|HACAbox", snoRNAtype)) %>%
  group_by(sample, snoRNAtype, dist) %>%
  summarise(sum = sum(value)) %>%
  ungroup() %>%
  spread(dist, sum, fill = 0)  %>%
  gather("dist","sum", 3:203) %>%
  separate(sample, c("GSM", "sample", "PAPgroup", "group", "rep"), sep = "_") %>%
  ungroup() %>%
  mutate(sample = gsub("si", "", sample)) %>%
  mutate(sample = gsub("GFP", "WT", sample))
```

```
## Warning: Expected 5 pieces. Missing pieces filled with `NA` in 6746 rows [1,
## 2, 3, 4, 5, 6, 197, 198, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225,
## 226, ...].
```

```
## `summarise()` has grouped output by 'sample', 'snoRNAtype'. You can override
## using the `.groups` argument.
```

```
## Warning: Expected 5 pieces. Additional pieces discarded in 2412 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```r
#make rel to siCtrl (WT)

RNAi_controls <- RNAi %>% filter(sample == "WT") %>%
                 mutate(WT = sum) %>%
                 select(-sum, -sample, -GSM)
RNAi_controls
```

```
## # A tibble: 1,206 x 6
##     PAPgroup group rep   snoRNAtype dist     WT
##     <chr>    <chr> <chr> <chr>      <chr> <dbl>
##  1 xPAP      in    1     CDbox      -100   12.3
##  2 xPAP      in    1     HACAbox    -100   18.5
##  3 xPAP      in    2     CDbox      -100   30.9
##  4 xPAP      in    2     HACAbox    -100   41.2
##  5 xPAP      in    3     CDbox      -100   11.1
##  6 xPAP      in    3     HACAbox    -100   33.3
##  7 xPAP      in    1     CDbox      -99    27.7
##  8 xPAP      in    1     HACAbox    -99    15.4
##  9 xPAP      in    2     CDbox      -99    17.2
## 10 xPAP      in    2     HACAbox    -99     6.87
## # ... with 1,196 more rows
```

```r
RNAi_RRP40 <-RNAi %>% filter(sample == "RRP40")

RNAi_rel_to_WT<-
RNAi_RRP40 %>%
  left_join(RNAi_controls) %>%
  mutate(log2_rel_WT = log2(sum/WT))
```

```
## Joining, by = c("PAPgroup", "group", "rep", "snoRNAtype", "dist")
```

```r
RNAi_rel_to_WT %>%
  ggplot() +
      geom_hline(yintercept = 0, alpha = 0.25) +
      geom_vline(xintercept = 0, linetype = "dashed") +
```

```
        geom_line(aes(x=as.numeric(dist), y=log2_rel_WT, col = group), stat = "summary", size = 1) +
        facet_grid(snoRNAtype ~ . , scales = "free_y") +
        coord_cartesian(xlim = c(0,50)) +
        scale_x_continuous(breaks = seq(0,100, 5)) +
        theme_bw() +
        scale_x_continuous(breaks = seq(0,50,5)) +
        xlab("Distance from snoRNA 3' end") +
        ylab("log2 (siRRP40/siCtrl") +
        theme(legend.position = "none") +
    theme_bw()
```

Scale for 'x' is already present. Adding another scale for 'x', which will
replace the existing scale.

Warning: Removed 59 rows containing non-finite values (stat_summary).

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