

histone_profiles

2024-01-25

```
library(ggplot2)
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
```

```
library(tidyr)
library(here); i_am("R/Figure3/Fig3B_histone_profiles.Rmd")
```

```
## here() starts at /projects/imb-pkbphil/sp/rnaseq/six_donor_trans/splicing_paper
```

```
## here() starts at /projects/imb-pkbphil/sp/rnaseq/six_donor_trans/splicing_paper
```

```
k4 = read.delim(here("31_leafcutter/histone_profile", "H3K4me3", "H3K4me3_white_beige-10.profile.tab"),
to_bins = function(colname){
  return (as.double(gsub("X","", colname)))
}
```

```
k4= pivot_longer(k4, 3:ncol(k4), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(k4)
```

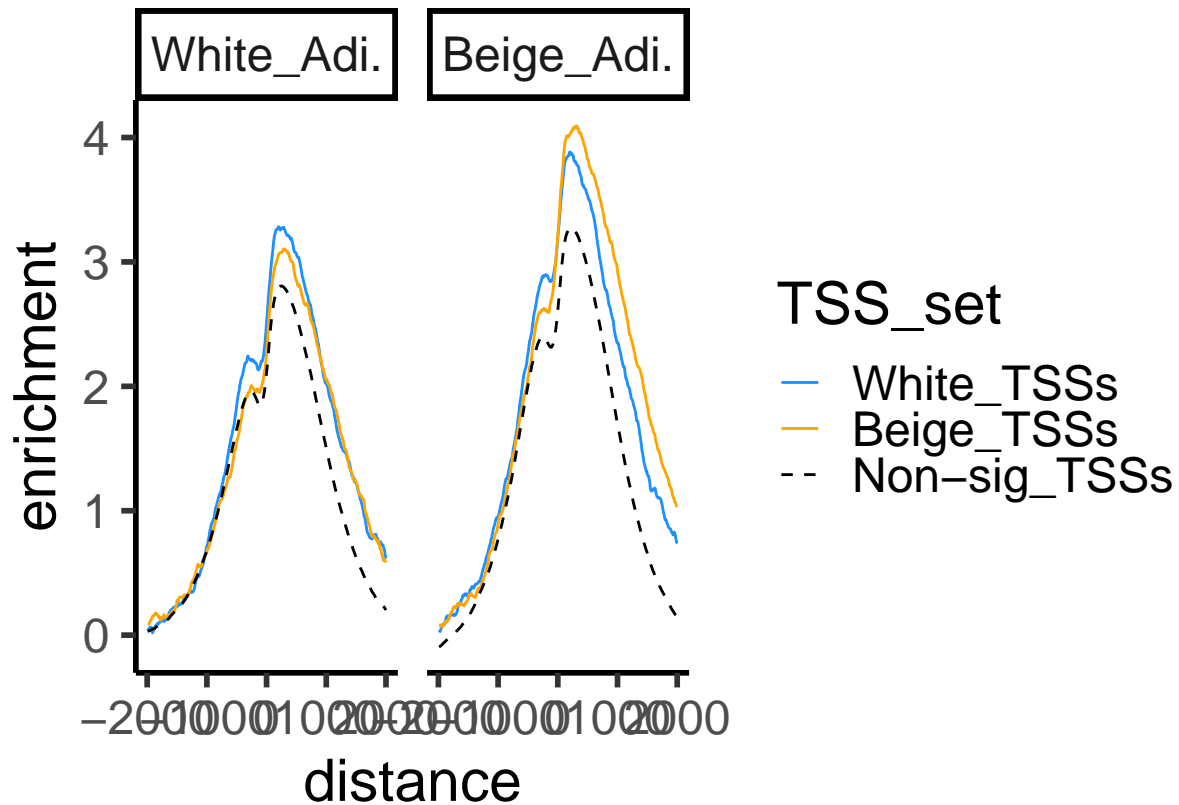
```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>      <dbl>
## 1 Beige_Adi. Non-sig_TSSs  795      NA
## 2 Beige_Adi. Non-sig_TSSs  796      NA
## 3 Beige_Adi. Non-sig_TSSs  797      NA
## 4 Beige_Adi. Non-sig_TSSs  798      NA
## 5 Beige_Adi. Non-sig_TSSs  799      NA
## 6 Beige_Adi. Non-sig_TSSs  800      NA
```

```
k4=filter(k4, !is.na(enrichment))
colnames(k4) = c("Context", "TSS_set", "bin", "enrichment")
k4 = mutate(k4, distance = bin*10-2000)
head(k4)
```

```
## # A tibble: 6 x 5
##   Context    TSS_set    bin enrichment distance
##   <chr>      <chr>    <dbl>      <dbl>    <dbl>
## 1 White_Adi. Beige_TSSs      1    0.0904    -1990
## 2 White_Adi. Beige_TSSs      2    0.0892    -1980
## 3 White_Adi. Beige_TSSs      3    0.0875    -1970
## 4 White_Adi. Beige_TSSs      4    0.104     -1960
## 5 White_Adi. Beige_TSSs      5    0.109     -1950
## 6 White_Adi. Beige_TSSs      6    0.120     -1940
```

```
k4$Context = factor(k4$Context, levels=c("White_Adi.", "Beige_Adi."))
k4$TSS_set = factor(k4$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(k4) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set), s
  facet_wrap(~Context) + theme_classic(base_size=22) +scale_color_manual(values= c("dodgerblue", "orange", "black"))
  scale_linetype_manual(values=c("solid", "solid", "dashed"))
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



```
ggsave(here("31_leafcutter/histone_profile", "H3K4me3", "H3K4me3_white_beige-10.profile.R.pdf"), width=10, height=10)
#
# geom_vline(aes(xintercept=-1000), linetype="dashed") + geom_vline(aes(xintercept=2000), linetype="dashed")
```

H3K27ac

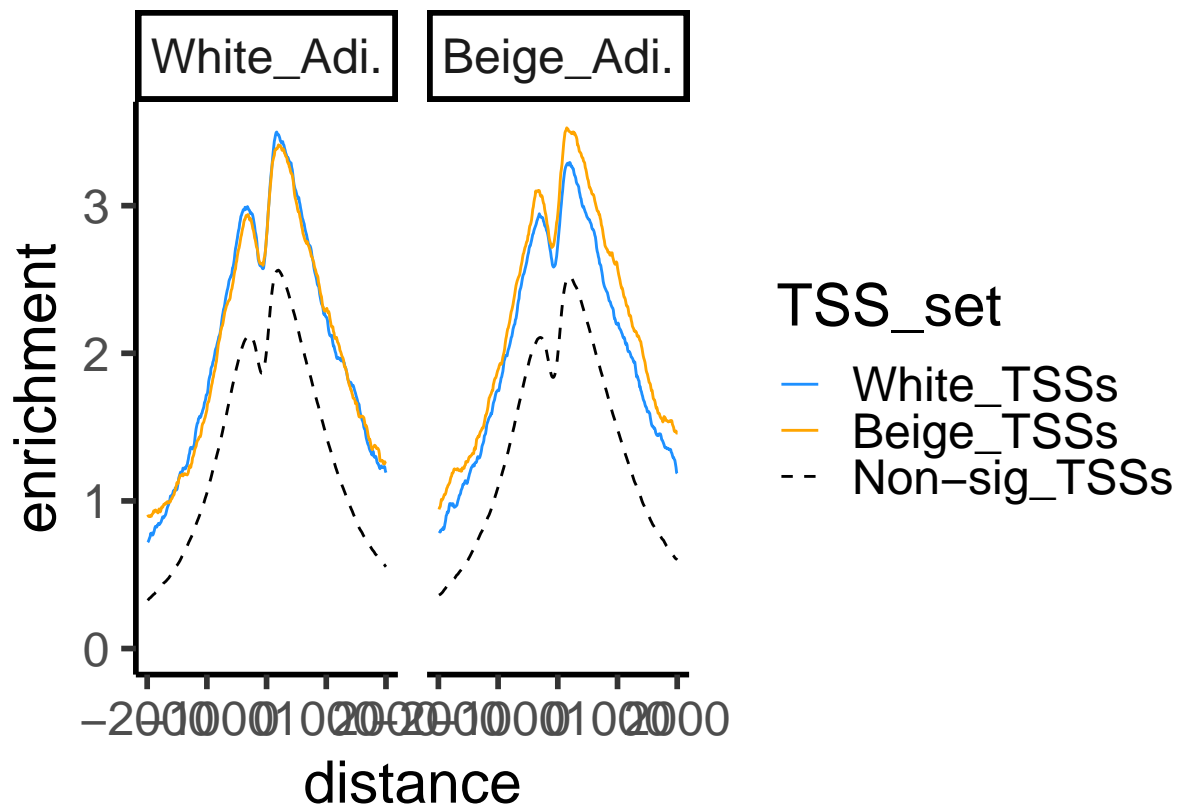
```
k27 = read.delim(here("31_leafcutter/histone_profile", "H3K27ac", "H3K27ac_white_beige-10.profile.tab"),
k27= pivot_longer(k27, 3:ncol(k27), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(k27)
```

```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>      <dbl>
## 1 Beige_Adi. Non-sig_TSSs  795      NA
## 2 Beige_Adi. Non-sig_TSSs  796      NA
## 3 Beige_Adi. Non-sig_TSSs  797      NA
## 4 Beige_Adi. Non-sig_TSSs  798      NA
## 5 Beige_Adi. Non-sig_TSSs  799      NA
## 6 Beige_Adi. Non-sig_TSSs  800      NA
```

```
k27=filter(k27, !is.na(enrichment))
colnames(k27) = c("Context", "TSS_set", "bin", "enrichment")
k27 = mutate(k27, distance = bin*10-2000)
head(k27)
```

```
## # A tibble: 6 x 5
##   Context    TSS_set    bin enrichment distance
##   <chr>      <chr>    <dbl>    <dbl>    <dbl>
## 1 White_Adi. Beige_TSSs    1      0.907   -1990
## 2 White_Adi. Beige_TSSs    2      0.896   -1980
## 3 White_Adi. Beige_TSSs    3      0.895   -1970
## 4 White_Adi. Beige_TSSs    4      0.893   -1960
## 5 White_Adi. Beige_TSSs    5      0.899   -1950
## 6 White_Adi. Beige_TSSs    6      0.902   -1940
```

```
k27$Context = factor(k27$Context, levels=c("White_Adi.", "Beige_Adi."))
k27$TSS_set = factor(k27$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(k27) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set),
  facet_wrap(~Context) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange", "black")) +
  scale_linetype_manual(values=c("solid", "solid", "dashed")) + coord_cartesian(y=c(0, max(k27$enrichment)))
```



```
ggsave(here("31_leafcutter/histone_profile", "H3K27ac", "H3K27ac_white_beige-10.profile.R.pdf"), width=10, height=10)
# + geom_hline(aes(yintercept=1)) +
# geom_vline(aes(xintercept=1000)) + geom_vline(aes(xintercept=-1000))
```

If we use the 1 enrichment threshold on the highest peak then a full +2000/-2000 is a good window

H3K4me1

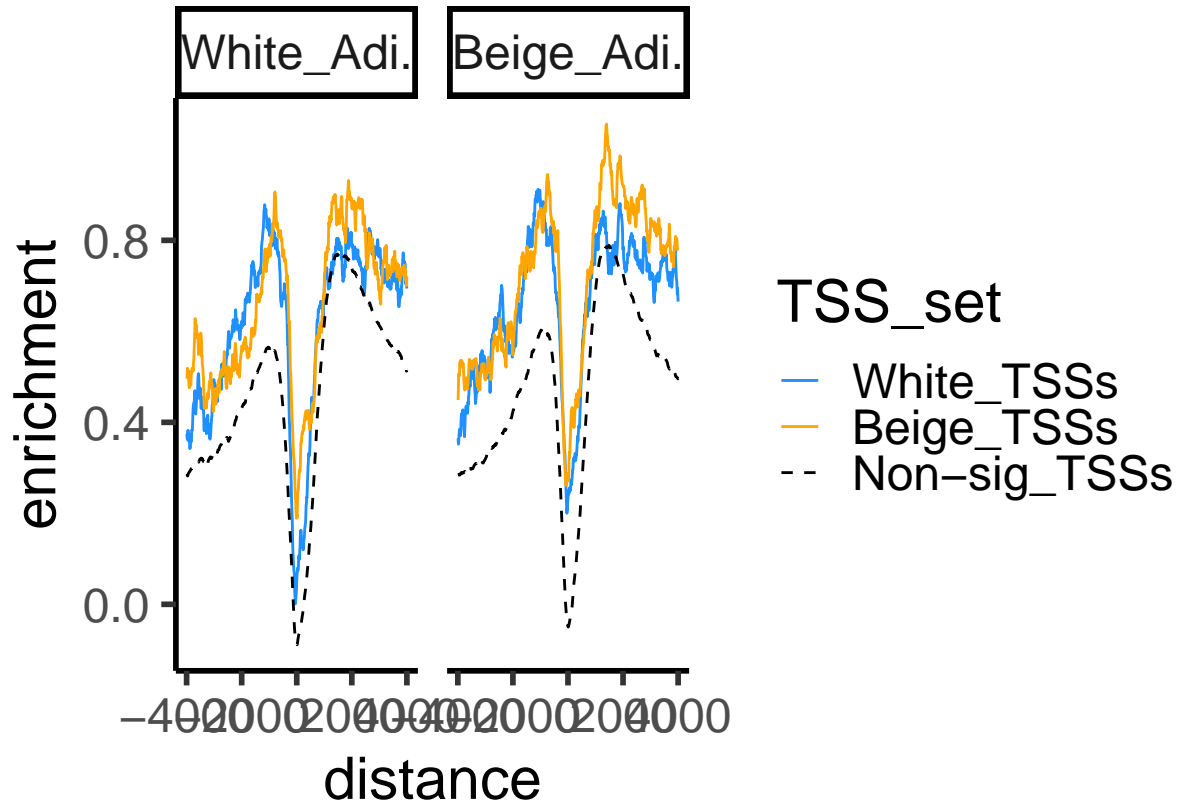
```
me1 = read.delim(here("31_leafcutter/histone_profile", "H3K4me1/H3K4me1_white_beige-10.profile.tab"), sep="\t")
me1 = pivot_longer(me1, 3:ncol(me1), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(me1)
```

```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>      <dbl>
## 1 Beige_Adi. Non-sig_TSSs 1595      NA
## 2 Beige_Adi. Non-sig_TSSs 1596      NA
## 3 Beige_Adi. Non-sig_TSSs 1597      NA
## 4 Beige_Adi. Non-sig_TSSs 1598      NA
## 5 Beige_Adi. Non-sig_TSSs 1599      NA
## 6 Beige_Adi. Non-sig_TSSs 1600      NA
```

```
me1 = filter(me1, !is.na(enrichment))
colnames(me1) = c("Context", "TSS_set", "bin", "enrichment")
me1 = mutate(me1, distance = bin*10-4000)
head(me1)
```

```
## # A tibble: 6 x 5
##   Context    TSS_set      bin enrichment distance
##   <chr>      <chr>    <dbl>      <dbl>      <dbl>
## 1 White_Adi. Beige_TSSs      1      0.496     -3990
## 2 White_Adi. Beige_TSSs      2      0.503     -3980
## 3 White_Adi. Beige_TSSs      3      0.519     -3970
## 4 White_Adi. Beige_TSSs      4      0.513     -3960
## 5 White_Adi. Beige_TSSs      5      0.512     -3950
## 6 White_Adi. Beige_TSSs      6      0.505     -3940
```

```
me1$Context = factor(me1$Context, levels=c("White_Adi.", "Beige_Adi."))
me1$TSS_set = factor(me1$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(me1) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set),
  facet_wrap(~Context) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange", "black"))
  scale_linetype_manual(values=c("solid", "solid", "dashed"))
```



```
ggsave(here("31_leafcutter/histone_profile", "H3K4me1", "H3K4me1_white_beige-10.profile.R.pdf"), width=10, height=10)
# + geom_hline(aes(yintercept=0.25)) + geom_vline(aes(xintercept=-250)) +
#   geom_vline(aes(xintercept=500)) + geom_vline(aes(xintercept=-2000))
```

Should probably do the upstream and downstream windows... I wouldn't do stats on depletion levels, cos that depends on the noise/input and stuff. But the most obvious one is +500 to 4000... or is it? upstream is the promoter, makes more sense

H3K27me3

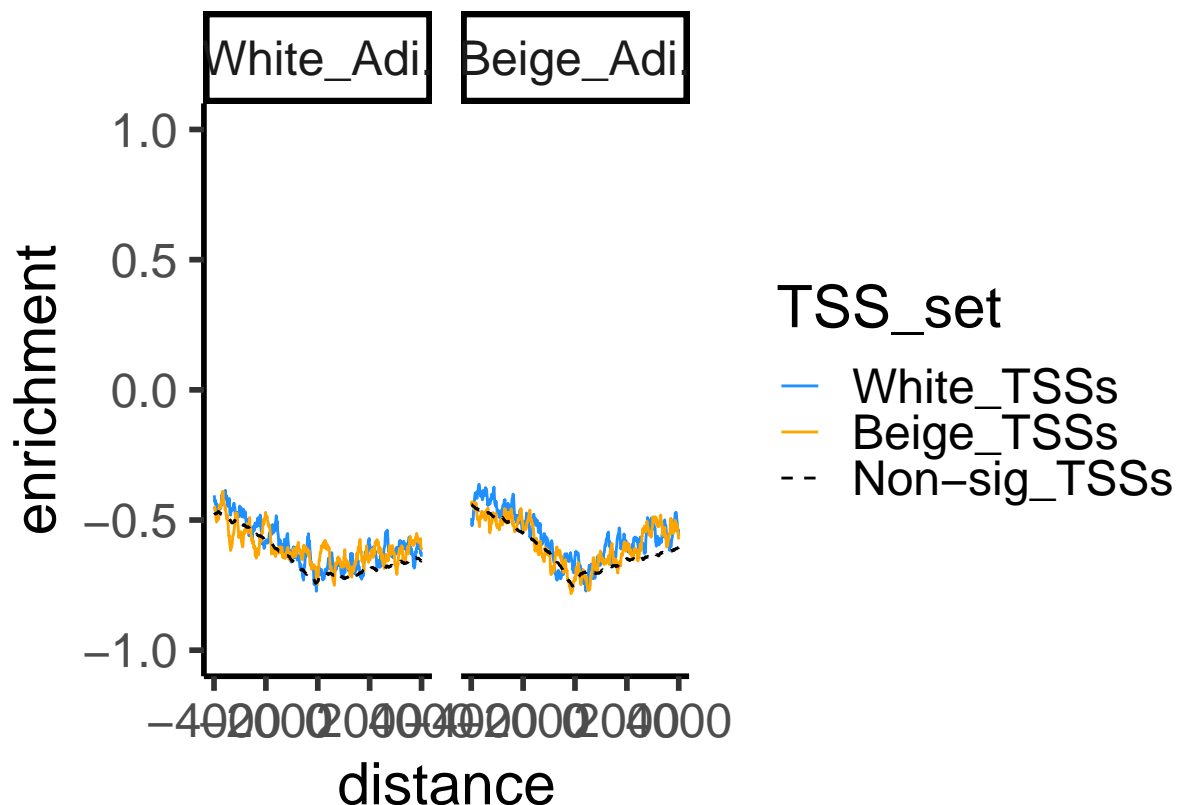
```
pc = read.delim(here("31_leafcutter/histone_profile", "H3K27me3/H3K27me3_white_beige-10.profile.tab"), as.is=TRUE)
pc = pivot_longer(pc, 3:ncol(pc), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(pc)
```

```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>    <dbl>
## 1 Beige_Adi. Non-sig_TSSs 1595      NA
## 2 Beige_Adi. Non-sig_TSSs 1596      NA
## 3 Beige_Adi. Non-sig_TSSs 1597      NA
## 4 Beige_Adi. Non-sig_TSSs 1598      NA
## 5 Beige_Adi. Non-sig_TSSs 1599      NA
## 6 Beige_Adi. Non-sig_TSSs 1600      NA
```

```
pc=filter(pc, !is.na(enrichment))
colnames(pc) = c("Context", "TSS_set", "bin", "enrichment")
pc = mutate(pc, distance = bin*10-4000)
head(pc)
```

```
## # A tibble: 6 x 5
##   Context    TSS_set    bin enrichment distance
##   <chr>      <chr>    <dbl>      <dbl>    <dbl>
## 1 White_Adi. Beige_TSSs     1    -0.459    -3990
## 2 White_Adi. Beige_TSSs     2    -0.451    -3980
## 3 White_Adi. Beige_TSSs     3    -0.453    -3970
## 4 White_Adi. Beige_TSSs     4    -0.458    -3960
## 5 White_Adi. Beige_TSSs     5    -0.468    -3950
## 6 White_Adi. Beige_TSSs     6    -0.474    -3940
```

```
pc$Context = factor(pc$Context, levels=c("White_Adi.", "Beige_Adi."))
pc$TSS_set = factor(pc$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(pc) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set), s
  facet_wrap(~Context) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange", "black"))
  scale_linetype_manual(values=c("solid", "solid", "dashed"))+ coord_cartesian(y=c(-1,1))
```



```
ggsave(here("31_leafcutter/histone_profile", "H3K27me3", "H3K27me3_white_beige-10.profile.R.pdf"), width=
  #+ geom_hline(aes(yintercept=0))
```

PPARG

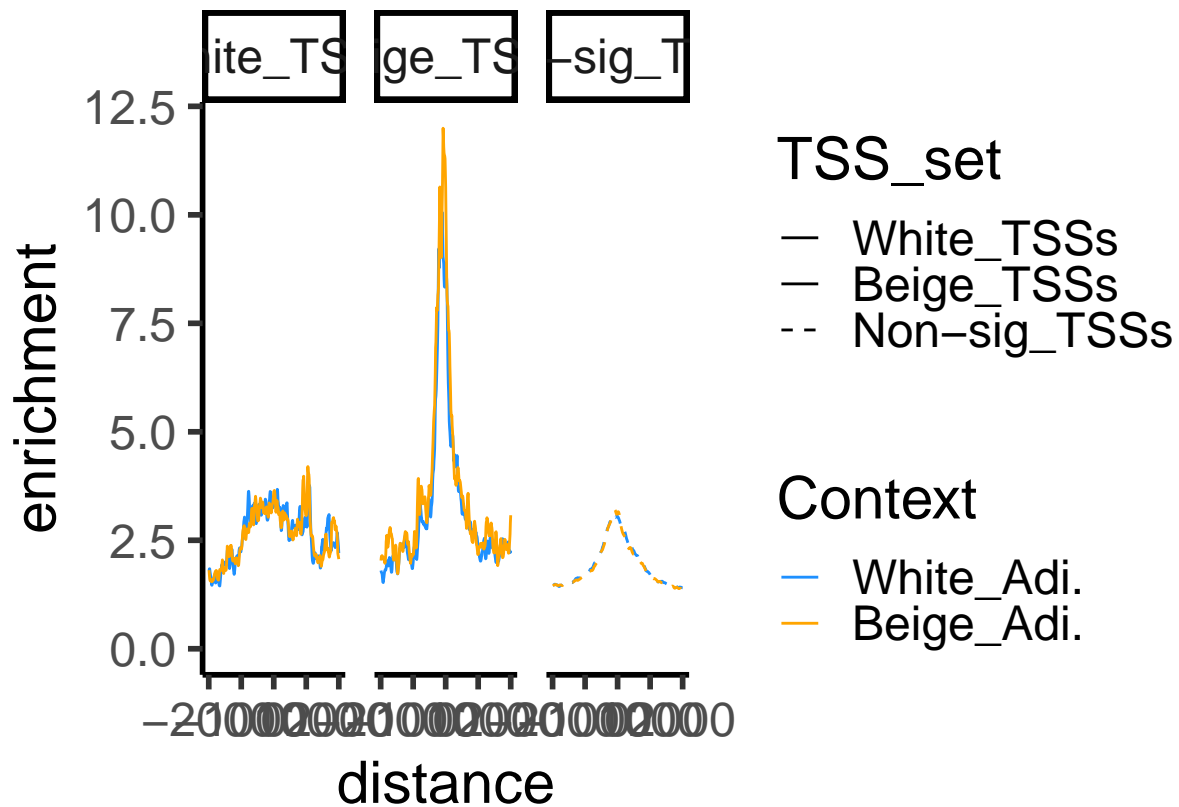
```
pparg = read.delim(here("31_leafcutter/histone_profile", "PPARG/PPARG_white_beige-10.profile.tab"), skip=1)
pparg = pivot_longer(pparg, 3:ncol(pparg), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(pparg)
```

```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>      <dbl>
## 1 Beige_Adi. Non-sig_TSSs  795      NA
## 2 Beige_Adi. Non-sig_TSSs  796      NA
## 3 Beige_Adi. Non-sig_TSSs  797      NA
## 4 Beige_Adi. Non-sig_TSSs  798      NA
## 5 Beige_Adi. Non-sig_TSSs  799      NA
## 6 Beige_Adi. Non-sig_TSSs  800      NA
```

```
pparg = filter(pparg, !is.na(enrichment))
colnames(pparg) = c("Context", "TSS_set", "bin", "enrichment")
pparg = mutate(pparg, distance = bin*10-2000)
head(pparg)
```

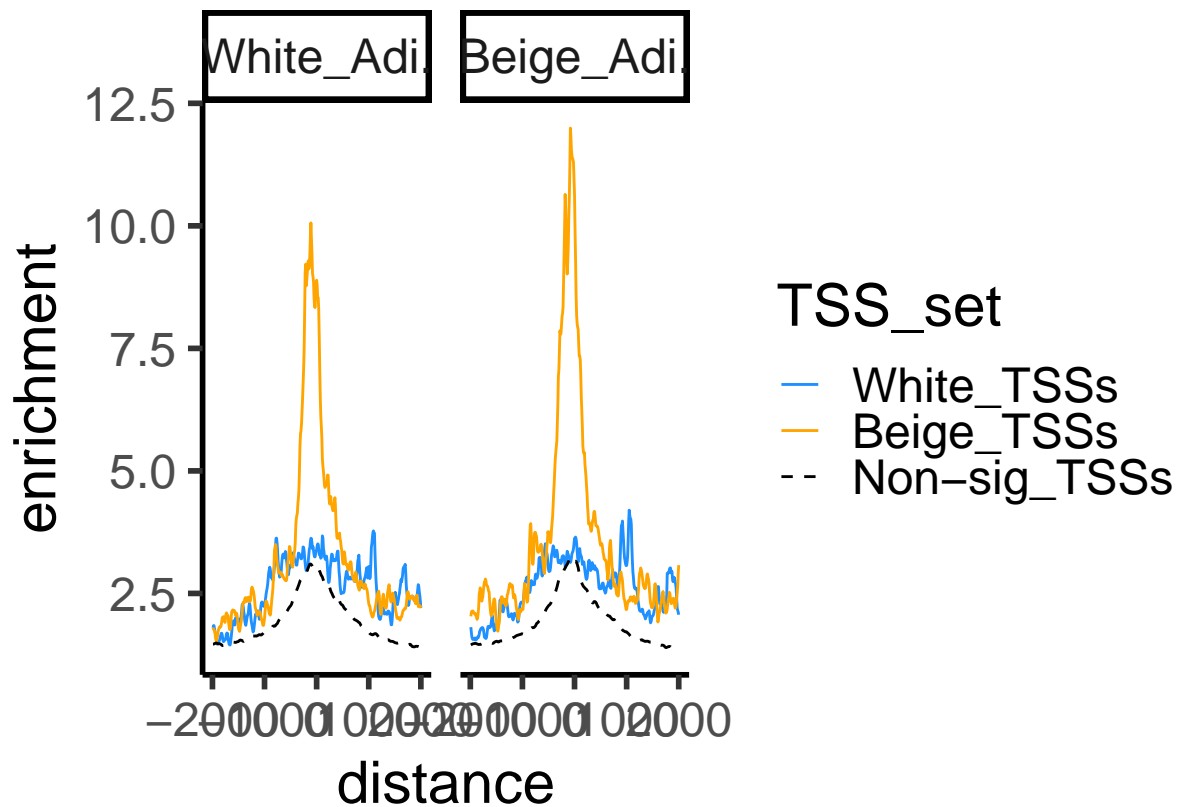
```
## # A tibble: 6 x 5
##   Context      TSS_set      bin enrichment distance
##   <chr>      <chr>    <dbl>      <dbl>    <dbl>
## 1 White_Adi. Beige_TSSs      1      1.80    -1990
## 2 White_Adi. Beige_TSSs      2      1.77    -1980
## 3 White_Adi. Beige_TSSs      3      1.75    -1970
## 4 White_Adi. Beige_TSSs      4      1.75    -1960
## 5 White_Adi. Beige_TSSs      5      1.68    -1950
## 6 White_Adi. Beige_TSSs      6      1.59    -1940
```

```
pparg$Context = factor(pparg$Context, levels=c("White_Adi.", "Beige_Adi."))
pparg$TSS_set = factor(pparg$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(pparg) + geom_line(aes(x=distance, y=enrichment, group=Context, color=Context, linetype=TSS_set))
  facet_wrap(~TSS_set) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange", "black"))
  scale_linetype_manual(values=c("solid", "solid", "dashed")) + coord_cartesian(y=c(0,12))
```

```
ggsave(here("31_leafcutter/histone_profile", "PPARG", "PPARG_white_beige-10.profile.R.pdf"), width=15, height=10)

ggplot(pparg) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set))
  facet_wrap(~Context) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange", "orange"))
  scale_linetype_manual(values=c("solid", "solid", "dashed")) #+ coord_cartesian(y=c(-1,1))
```



```
#+ geom_hline(aes(yintercept=0))
```

```
##MED1
```

```
med1 = read.delim(here("31_leafcutter/histone_profile", "MED1/MED1_white_beige-10.profile.tab"), skip=1)
med1= pivot_longer(med1, 3:ncol(med1), names_to = "bin", names_transform = to_bins, values_to = "enrichment")
tail(med1)
```

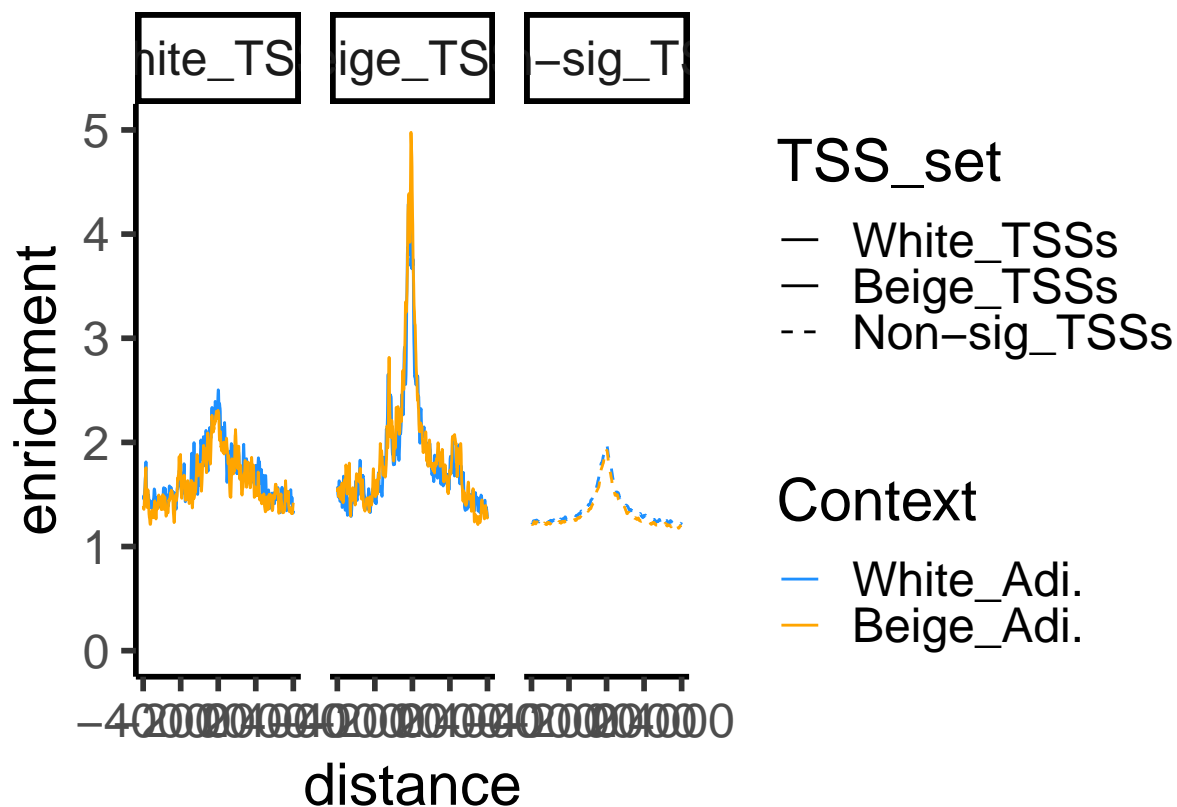
```
## # A tibble: 6 x 4
##   bins      X      bin enrichment
##   <chr>    <chr>    <dbl>      <dbl>
## 1 Beige_Adi. Non-sig_TSSs 1595      NA
## 2 Beige_Adi. Non-sig_TSSs 1596      NA
## 3 Beige_Adi. Non-sig_TSSs 1597      NA
## 4 Beige_Adi. Non-sig_TSSs 1598      NA
## 5 Beige_Adi. Non-sig_TSSs 1599      NA
## 6 Beige_Adi. Non-sig_TSSs 1600      NA
```

```
med1=filter(med1, !is.na(enrichment))
colnames(med1) = c("Context", "TSS_set", "bin", "enrichment")
med1 = mutate(med1, distance = bin*10-4000)
head(med1)
```

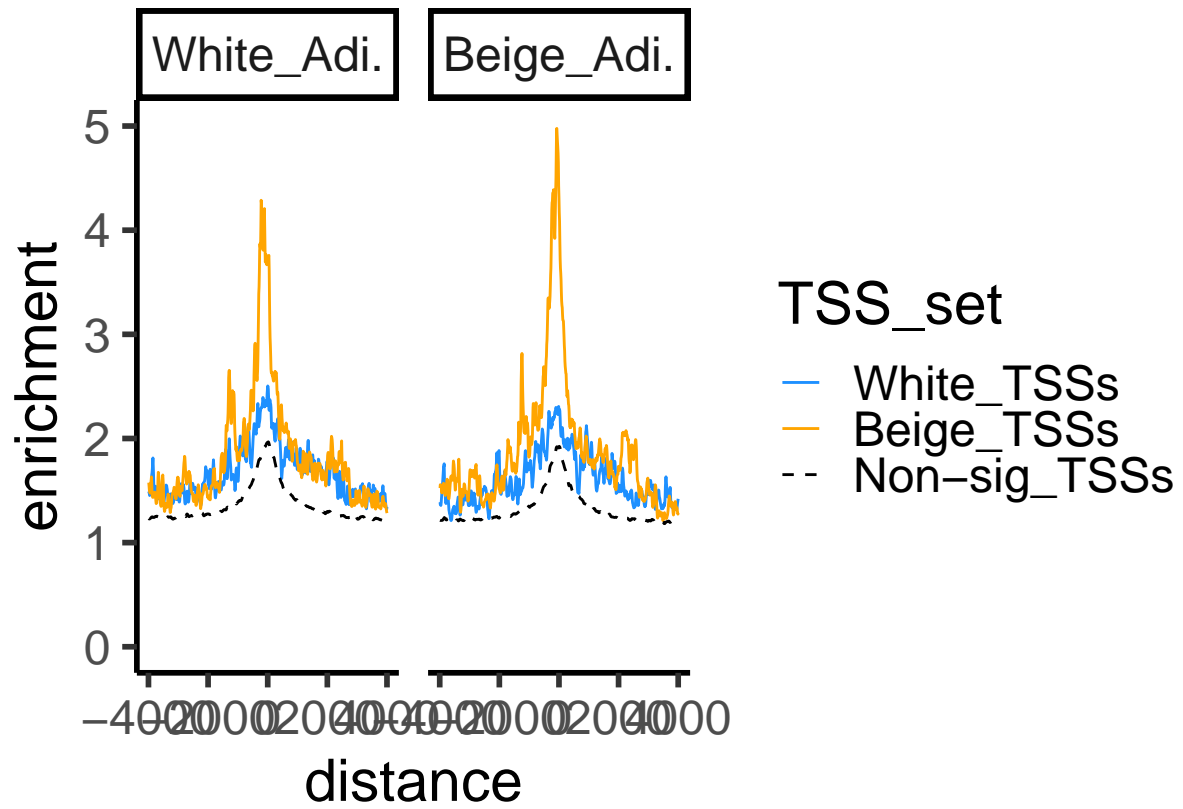
```
## # A tibble: 6 x 5
```

```
## Context TSS_set bin enrichment distance
## <chr> <chr> <dbl> <dbl> <dbl>
## 1 White_Adi. Beige_TSSs 1 1.58 -3990
## 2 White_Adi. Beige_TSSs 2 1.54 -3980
## 3 White_Adi. Beige_TSSs 3 1.50 -3970
## 4 White_Adi. Beige_TSSs 4 1.49 -3960
## 5 White_Adi. Beige_TSSs 5 1.48 -3950
## 6 White_Adi. Beige_TSSs 6 1.48 -3940
```

```
med1$Context = factor(med1$Context, levels=c("White_Adi.", "Beige_Adi."))
med1$TSS_set = factor(med1$TSS_set, levels = c("White_TSSs", "Beige_TSSs", "Non-sig_TSSs"))
ggplot(med1) + geom_line(aes(x=distance, y=enrichment, group=Context, color=Context, linetype=TSS_set),
  facet_wrap(~TSS_set) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange"))
  scale_linetype_manual(values=c("solid", "solid", "dashed"))+ coord_cartesian(y=c(0,5))
```



```
ggsave(here("31_leafcutter/histone_profile", "MED1", "MED1_white_beige-10.profile.R.pdf"), width=15, height=10)
ggplot(med1) + geom_line(aes(x=distance, y=enrichment, group=TSS_set, color=TSS_set, linetype=TSS_set),
  facet_wrap(~Context) + theme_classic(base_size=22) + scale_color_manual(values= c("dodgerblue", "orange"))
  scale_linetype_manual(values=c("solid", "solid", "dashed"))+ coord_cartesian(y=c(0,5))
```



```
#+ geom_hline(aes(yintercept=0))
```