

## ChIPseq stats PPARG + MED1

```
library(tidyr)
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(ggplot2)
library(ggpubr)
library(ggrepel)
library(here); i_am("R/Figure3/Fig3G,Ipparg_stats.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
```

```
WINDOWS= "C:/Users/sarahhp/OneDrive - Universitetet i Oslo/Projects/rnaseq/beige_rnaseq_jul22/31_leafcutter"
```

```
mark = "PPARG"
window = "250:250"
tss_sets = c("beige", "white", "not_sig")
```

```
enrich_tables = list()
for (set in tss_sets){
  file = here("31_leafcutter/histone_profile", mark, paste0("window", window), paste0(set, ".", window, ".tsv"))
  annot = read.delim(file, quote="")
  colnames(annot)[grep("chr", colnames(annot))] = "chr"
  annot$tss_set = set
  enrich_tables[[set]] = annot
}
str(enrich_tables)
```

```
## List of 3
## $ beige :'data.frame': 209 obs. of 6 variables:
## ..$ chr : chr [1:209] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start : int [1:209] 6614606 14923879 23800151 45339704 55215111 87129514 113757812 1137596...
```

```
## ..$ end      : int [1:209] 6615107 14924380 23800652 45340205 55215612 87130015 113758313 1137601
## ..$ White_Adi.: num [1:209] 0.988 1 10.139 3.524 3.61 ...
## ..$ Beige_Adi.: num [1:209] 1 1.23 9.96 2.74 4.55 ...
## ..$ tss_set   : chr [1:209] "beige" "beige" "beige" "beige" ...
## $ white      : 'data.frame': 213 obs. of 6 variables:
## ..$ chr       : chr [1:213] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start     : int [1:213] 6613480 11803426 14945668 17634006 23800530 26432031 33349634 55215113
## ..$ end       : int [1:213] 6613981 11803927 14946169 17634507 23801031 26432532 33350135 55215614
## ..$ White_Adi.: num [1:213] 6.89 4.04 1.59 0.82 13.31 ...
## ..$ Beige_Adi.: num [1:213] 5.02 2.71 5.34 1.18 6.97 ...
## ..$ tss_set   : chr [1:213] "white" "white" "white" "white" ...
## $ not_sig    : 'data.frame': 23636 obs. of 6 variables:
## ..$ chr       : chr [1:23636] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start     : int [1:23636] 29119 495225 498725 498725 498924 501356 502622 502622 514162 517001
## ..$ end       : int [1:23636] 29620 495726 499226 499226 499425 501857 503123 503123 514663 517502
## ..$ White_Adi.: num [1:23636] 1.336 1.592 1.042 1.042 0.827 ...
## ..$ Beige_Adi.: num [1:23636] 1.03 1.87 1.97 1.97 1.36 ...
## ..$ tss_set   : chr [1:23636] "not_sig" "not_sig" "not_sig" "not_sig" ...
```

```
annot <- do.call(rbind, enrich_tables)
table(annot$tss_set)
```

```
##
## beige not_sig white
## 209 23636 213
```

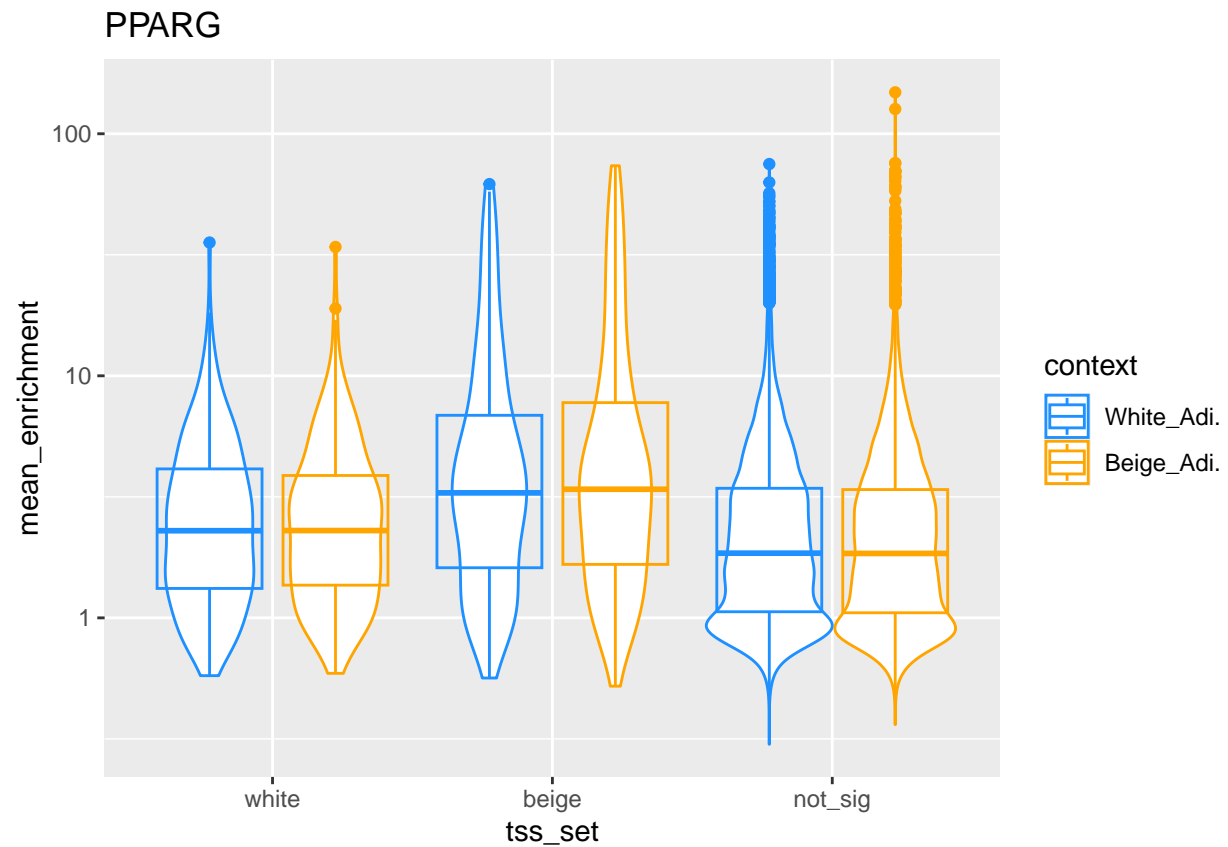
```
head(annot); nrow(annot)
```

```
##      chr      start      end White_Adi. Beige_Adi. tss_set
## beige.1 chr1 6614606 6615107 0.9879027 1.000000 beige
## beige.2 chr1 14923879 14924380 1.0000000 1.232636 beige
## beige.3 chr1 23800151 23800652 10.1387030 9.959925 beige
## beige.4 chr1 45339704 45340205 3.5236493 2.738524 beige
## beige.5 chr1 55215111 55215612 3.6102585 4.546581 beige
## beige.6 chr1 87129514 87130015 8.5680821 4.673226 beige
```

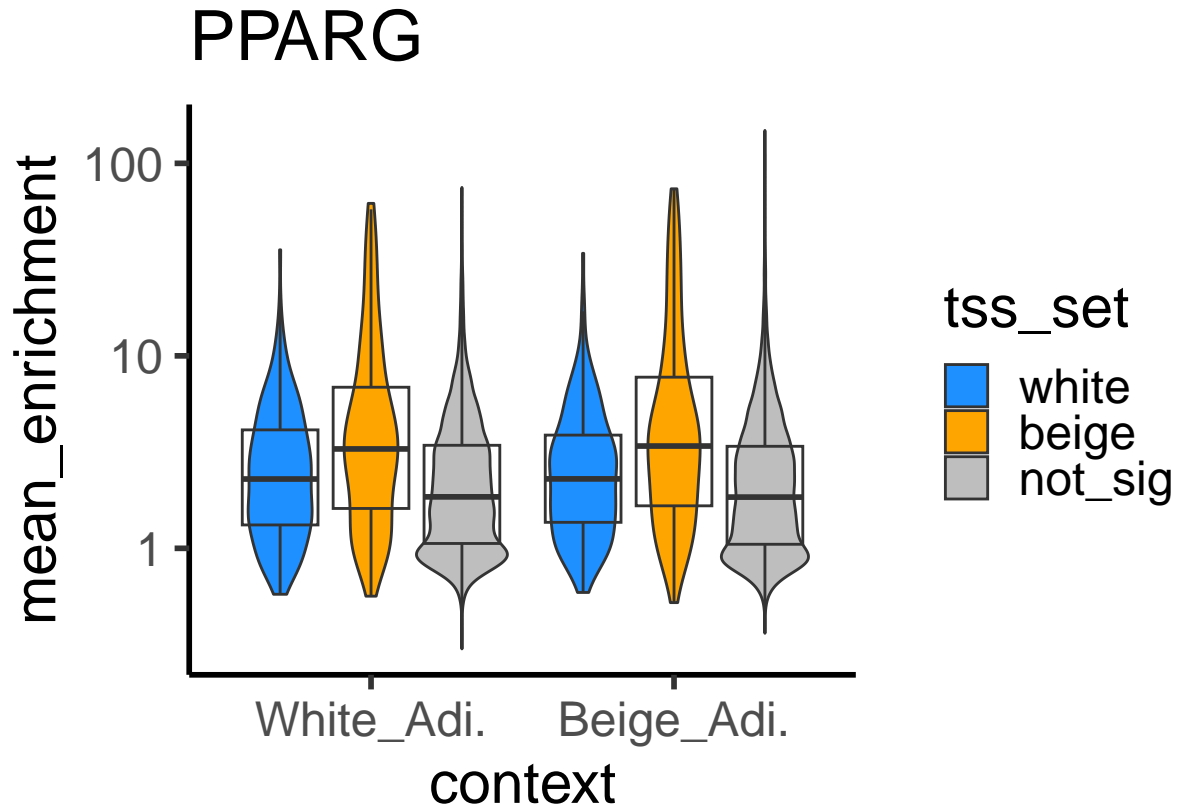
```
## [1] 24058
```

```
long = pivot_longer(annot, grep("Adi.", colnames(annot)), names_to = "context", values_to = "mean_enrichment")
long$context = factor(long$context, levels=c("White_Adi.", "Beige_Adi."))
long$tss_set = factor(long$tss_set, levels=c("white", "beige", "not_sig"))
long$group = paste0(long$context, long$tss_set)
long$group = factor(long$group, levels=paste0(rep(levels(long$context), each=3), levels(long$tss_set)))
```

```
ggplot(long,
  aes(x=tss_set, y=mean_enrichment, color=context)) +geom_violin()+
  geom_boxplot(fill=NA, position=position_dodge(0.9)) +scale_color_manual(values= c("dodgerblue", "orange"))
ggtitle(mark) + scale_y_log10()
```



```
ggplot(long,
  aes(x=context, y=mean_enrichment, fill=tss_set, group=group)) + geom_violin() +
  geom_boxplot(fill=NA, position=position_dodge(0.9), outlier.shape = NA) + scale_fill_manual(values=
  ggtitle(mark) + scale_y_log10() + theme_classic(base_size=22)
```



```
ggsave(here("31_leafcutter/histone_profile", mark, paste0(mark, "_violin_plot.pdf")))
```

```
## Saving 6.5 x 4.5 in image
```

```
long$logenrichment = log10(long$mean_enrichment)
summary(aov(logenrichment ~ tss_set + context, data=long))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## tss_set      2      34  17.210  151.266 <2e-16 ***
## context      1       0   0.122   1.074    0.3
## Residuals 48112  5474   0.114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
compare_means(logenrichment ~ tss_set, data=long, method="wilcox.test", group.by="context")
```

```
## # A tibble: 6 x 9
##   context .y.      group1 group2      p    p.adj p.format p.signif method
##   <fct>   <chr>    <chr> <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 White_Adi. logenrich~ white beige  2.63e- 5 7.90e- 5 2.6e-05 **** Wilco~
## 2 White_Adi. logenrich~ white not_s~ 1.44e- 3 1.4 e- 3 0.00144 ** Wilco~
## 3 White_Adi. logenrich~ beige not_s~ 3.81e-17 1.9 e-16 < 2e-16 **** Wilco~
## 4 Beige_Adi. logenrich~ white beige  7.59e- 6 3 e- 5 7.6e-06 **** Wilco~
## 5 Beige_Adi. logenrich~ white not_s~ 4.73e- 4 9.5 e- 4 0.00047 *** Wilco~
## 6 Beige_Adi. logenrich~ beige not_s~ 6.57e-19 3.90e-18 < 2e-16 **** Wilco~
```

```
summary(aov(logenrichment ~tss_set*context, data=long))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## tss_set        2     34  17.210  151.261 <2e-16 ***
## context        1      0   0.122    1.074  0.300
## tss_set:context 2      0   0.033    0.289  0.749
## Residuals     48110  5474   0.114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## MED1

```
mark = "MED1"
window = "500:500"
tss_sets = c("beige", "white", "not_sig")

enrich_tables = list()
for (set in tss_sets){
  file = here("31_leafcutter/histone_profile", mark, paste0("window", window), paste0(set, ".", window, ".tsv"))
  annot = read.delim(file, quote="")
  colnames(annot)[grep("chr", colnames(annot))] = "chr"
  annot$tss_set = set
  enrich_tables[[set]] = annot
}
str(enrich_tables)
```

```
## List of 3
## $ beige :'data.frame': 209 obs. of 6 variables:
## ..$ chr : chr [1:209] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start : int [1:209] 6614356 14923629 23799901 45339454 55214861 87129264 113757562 1137593...
## ..$ end : int [1:209] 6615357 14924630 23800902 45340455 55215862 87130265 113758563 1137603...
## ..$ White_Adi.: num [1:209] 1.34 1.11 2.5 1.71 2.93 ...
## ..$ Beige_Adi.: num [1:209] 1.16 1.04 2.23 1.56 2.8 ...
## ..$ tss_set : chr [1:209] "beige" "beige" "beige" "beige" ...
## $ white :'data.frame': 213 obs. of 6 variables:
## ..$ chr : chr [1:213] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start : int [1:213] 6613230 11803176 14945418 17633756 23800280 26431781 33349384 55214863...
## ..$ end : int [1:213] 6614231 11804177 14946419 17634757 23801281 26432782 33350385 55215864...
## ..$ White_Adi.: num [1:213] 3.141 1.856 1.172 0.909 2.566 ...
## ..$ Beige_Adi.: num [1:213] 2.28 1.77 1.51 1.03 2.23 ...
## ..$ tss_set : chr [1:213] "white" "white" "white" "white" ...
## $ not_sig:'data.frame': 23636 obs. of 6 variables:
## ..$ chr : chr [1:23636] "chr1" "chr1" "chr1" "chr1" ...
## ..$ start : int [1:23636] 28869 494975 498475 498475 498674 501106 502372 502372 513912 516751...
## ..$ end : int [1:23636] 29870 495976 499476 499476 499675 502107 503373 503373 514913 517752...
## ..$ White_Adi.: num [1:23636] 1.87 0.999 0.921 0.921 0.904 ...
## ..$ Beige_Adi.: num [1:23636] 1.26 1.02 1.1 1.1 1.06 ...
## ..$ tss_set : chr [1:23636] "not_sig" "not_sig" "not_sig" "not_sig" ...
```

```
annot <- do.call(rbind, enrich_tables)
table(annot$tss_set)
```

```
##
## beige not_sig white
## 209 23636 213
```

```
head(annot); nrow(annot)
```

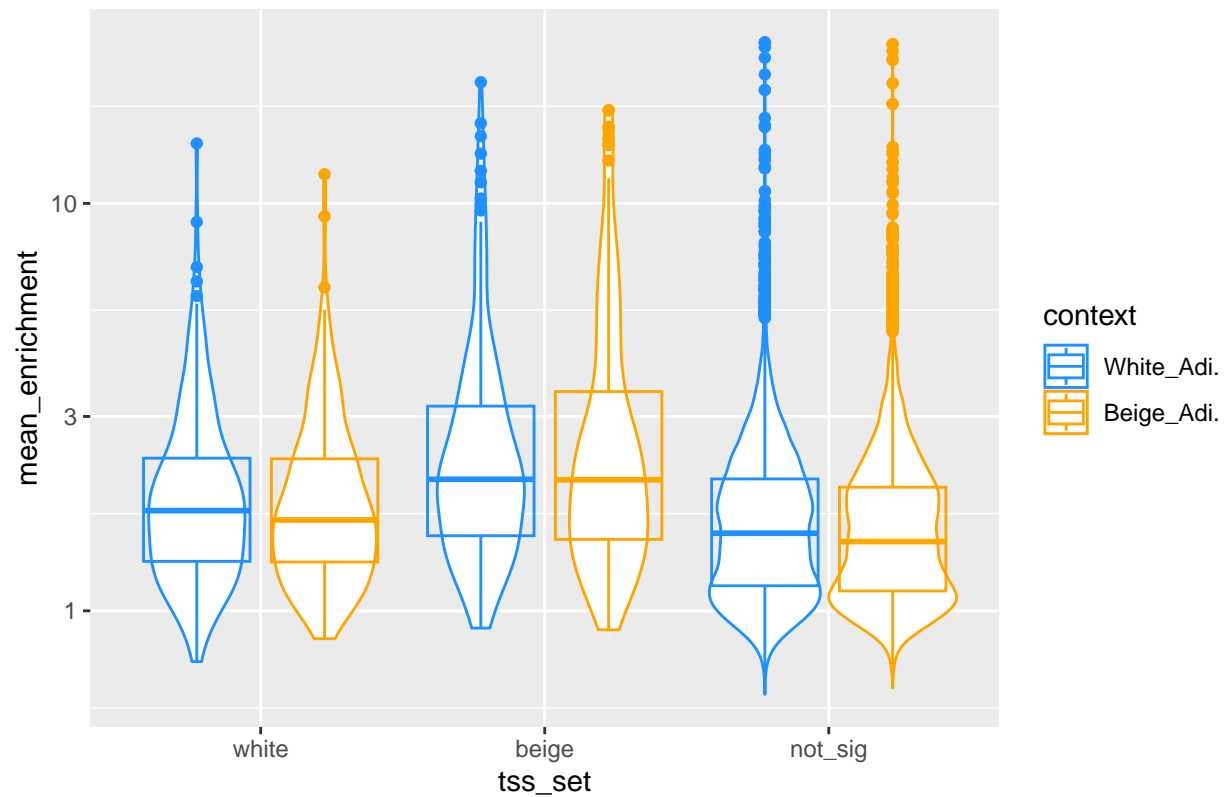
```
##      chr      start      end White_Adi. Beige_Adi. tss_set
## beige.1 chr1 6614356 6615357 1.338763 1.164247 beige
## beige.2 chr1 14923629 14924630 1.110319 1.043586 beige
## beige.3 chr1 23799901 23800902 2.502486 2.227372 beige
## beige.4 chr1 45339454 45340455 1.707652 1.556093 beige
## beige.5 chr1 55214861 55215862 2.932704 2.800650 beige
## beige.6 chr1 87129264 87130265 3.746468 1.930515 beige
```

```
## [1] 24058
```

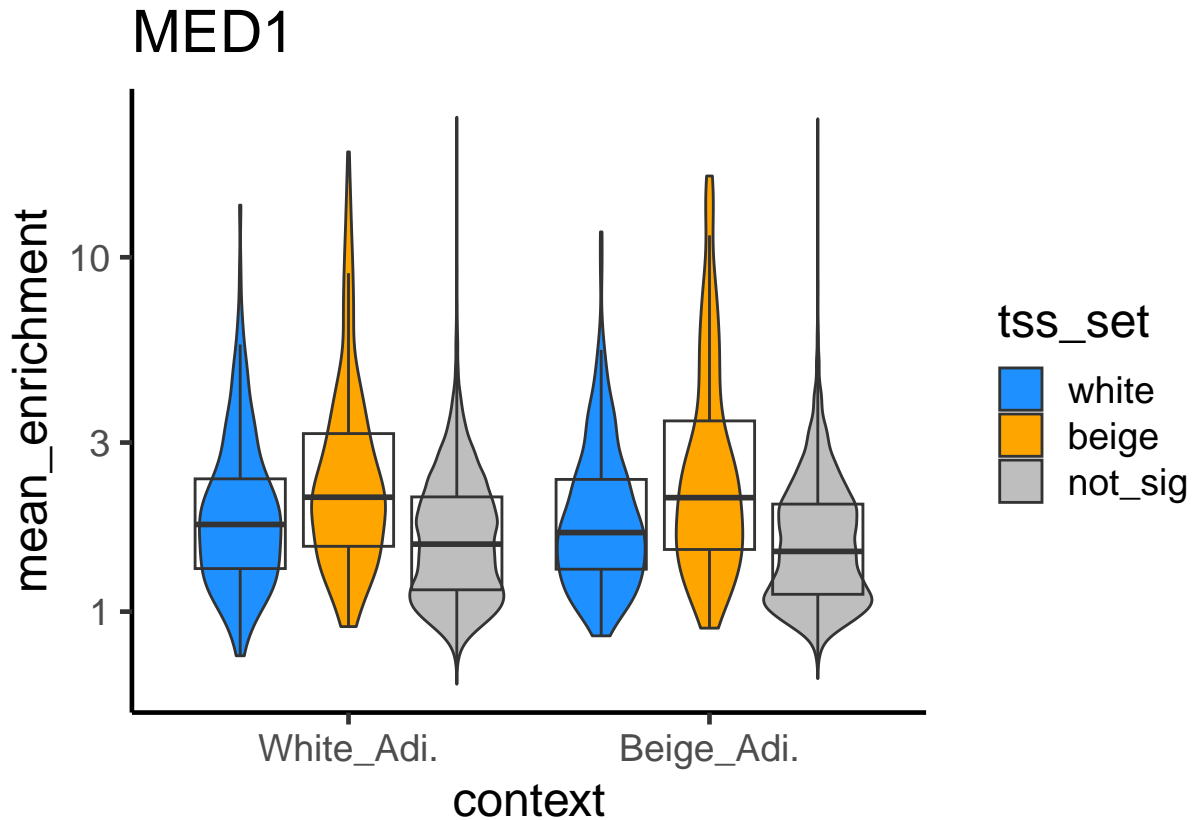
```
long = pivot_longer(annot, grep("Adi.", colnames(annot)), names_to = "context", values_to = "mean_enrichment")
long$context = factor(long$context, levels=c("White_Adi.", "Beige_Adi."))
long$tss_set = factor(long$tss_set, levels=c("white", "beige", "not_sig"))
long$group = paste0(long$context, long$tss_set)
long$group = factor(long$group, levels=paste0(rep(levels(long$context), each=3), levels(long$tss_set)))
```

```
ggplot(long,
  aes(x=tss_set, y=mean_enrichment, color=context)) +geom_violin()+
  geom_boxplot(fill=NA, position=position_dodge(0.9)) +scale_color_manual(values= c("dodgerblue", "orange", "green"))
ggtitle(mark) + scale_y_log10()
```

## MED1



```
ggplot(long,
  aes(x=context, y=mean_enrichment, fill=tss_set, group=group)) + geom_violin() +
  geom_boxplot(fill=NA, position=position_dodge(0.9), outlier.shape = NA) + scale_fill_manual(values=
  ggtitle(mark) + scale_y_log10() + theme_classic(base_size=18)
```



```
ggsave(here("31_leafcutter/histone_profile", mark, paste0(mark, "_violin_plot.pdf")))
```

```
## Saving 6.5 x 4.5 in image
```

```
long$logenrichment = log10(long$mean_enrichment)
summary(aov(logenrichment ~ tss_set*context, data=long))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## tss_set      2   15.4   7.715  249.668 <2e-16 ***
## context      1    2.5   2.549   82.494 <2e-16 ***
## tss_set:context 2    0.1   0.046    1.494  0.224
## Residuals 48110 1486.7   0.031
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
compare_means(logenrichment ~ tss_set, data=long, method="wilcox.test", group.by="context")
```

```
## # A tibble: 6 x 9
##   context .y.      group1 group2      p    p.adj p.format p.signif method
##   <fct>   <chr>    <chr> <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 White_Adi. logenrich~ white beige  7.29e- 5 7.3 e- 5 7.3e-05 **** Wilco~
## 2 White_Adi. logenrich~ white not_s~ 7.17e- 6 2.20e- 5 7.2e-06 **** Wilco~
## 3 White_Adi. logenrich~ beige not_s~ 3.19e-22 1.6 e-21 < 2e-16 **** Wilco~
## 4 Beige_Adi. logenrich~ white beige  1.07e- 5 2.20e- 5 1.1e-05 **** Wilco~
```



```
## 5 Beige_Adi. logenrich~ white not_s~ 7.01e- 7 2.8 e- 6 7.0e-07 **** Wilco~
## 6 Beige_Adi. logenrich~ beige not_s~ 4.44e-25 2.70e-24 < 2e-16 **** Wilco~
```

```
compare_means(logenrichment ~ context, data=long, method="wilcox.test", group.by="tss_set")
```

```
## # A tibble: 3 x 9
##   tss_set .y.          group1 group2      p    p.adj p.format p.signif method
##   <fct>   <chr>        <chr> <chr>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 beige   logenrichment White~ Beige~ 9.95e- 1 1    e+ 0 1.00     ns      Wilco~
## 2 white   logenrichment White~ Beige~ 4.79e- 1 9.6 e- 1 0.48     ns      Wilco~
## 3 not_sig logenrichment White~ Beige~ 7.31e-22 2.20e-21 <2e-16 ****    Wilco~
```

Huh interesting Not significant TSSes have more med1 signal in white adipocytes than beige. Its highly significant even though the difference in means is so small; i guess thats the number of points helps it look more significant.