ChIPseq stats PPARg + MED1

```r
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
```

```r
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
```

```r
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, "
  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': 232 obs. of  6 variables:
##   ..$ chr      : chr [1:232] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start    : int [1:232] 6614606 14923879 23800151 45339704 55215111 87129514 113757812 14695960
##   ..$ end      : int [1:232] 6615107 14924380 23800652 45340205 55215612 87130015 113758313 14696010
##   ..$ White_Adi.: num [1:232] 0.988 1 10.139 3.524 3.61 ...
```

```
##   ..$ Beige_Adi.: num [1:232] 1 1.23 9.96 2.74 4.55 ...
##   ..$ tss_set   : chr [1:232] "beige" "beige" "beige" "beige" ...
## $ white  :'data.frame': 239 obs. of  6 variables:
##   ..$ chr       : chr [1:239] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start     : int [1:239] 6613480 11803426 14945668 23800530 33349634 55215113 89632821 93613846
##   ..$ end       : int [1:239] 6613981 11803927 14946169 23801031 33350135 55215614 89633322 93614347
##   ..$ White_Adi.: num [1:239] 6.89 4.04 1.59 13.31 1.69 ...
##   ..$ Beige_Adi.: num [1:239] 5.02 2.71 5.34 6.97 1.14 ...
##   ..$ tss_set   : chr [1:239] "white" "white" "white" "white" ...
## $ not_sig:'data.frame': 23529 obs. of  6 variables:
##   ..$ chr       : chr [1:23529] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start     : int [1:23529] 29119 502622 502622 514162 517001 522677 946002 950907 961198 1220041
##   ..$ end       : int [1:23529] 29620 503123 503123 514663 517502 523178 946503 951408 961699 1220543
##   ..$ White_Adi.: num [1:23529] 1.34 1.33 1.33 1.03 1.05 ...
##   ..$ Beige_Adi.: num [1:23529] 1.03 1.805 1.805 0.881 1.084 ...
##   ..$ tss_set   : chr [1:23529] "not_sig" "not_sig" "not_sig" "not_sig" ...
```

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

```
##
##   beige not_sig   white
##     232   23529     239
```
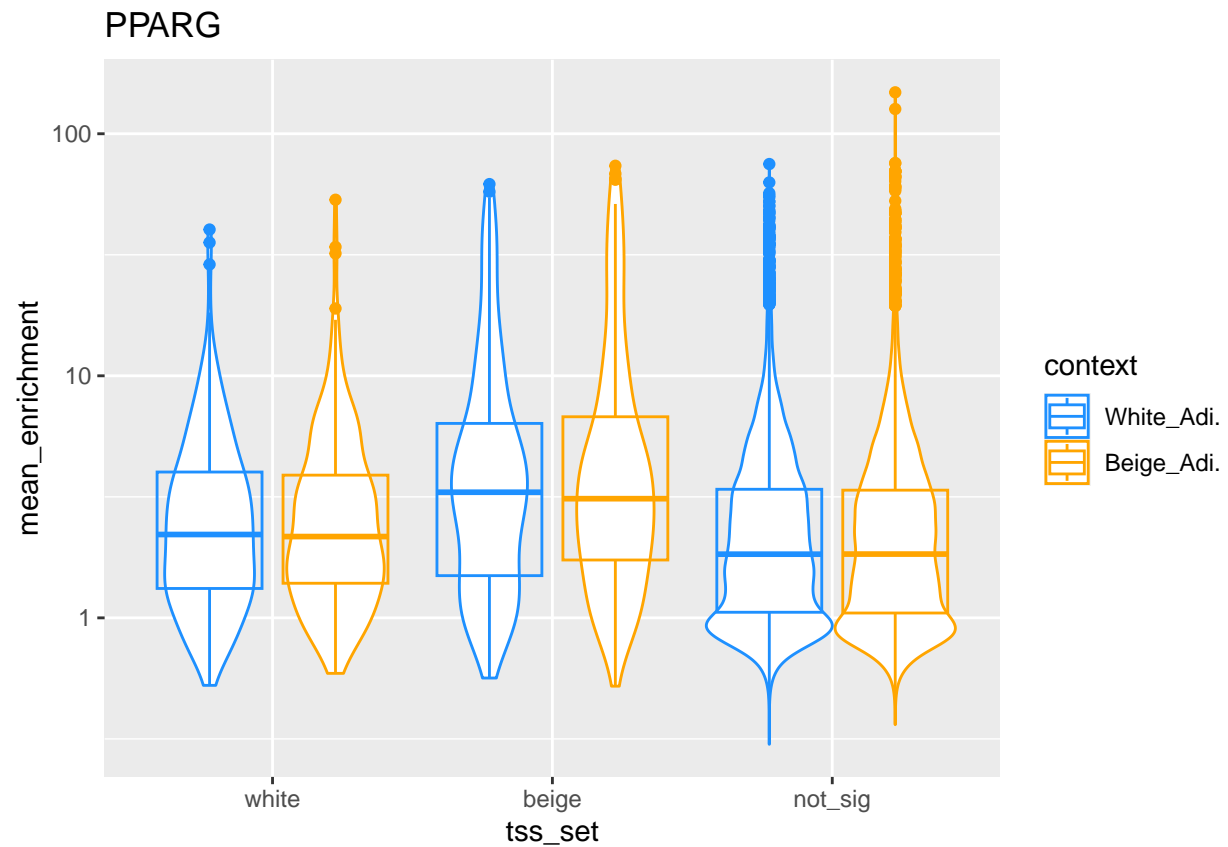
```
head(annot); nrow(annot)
```

```
##           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
```
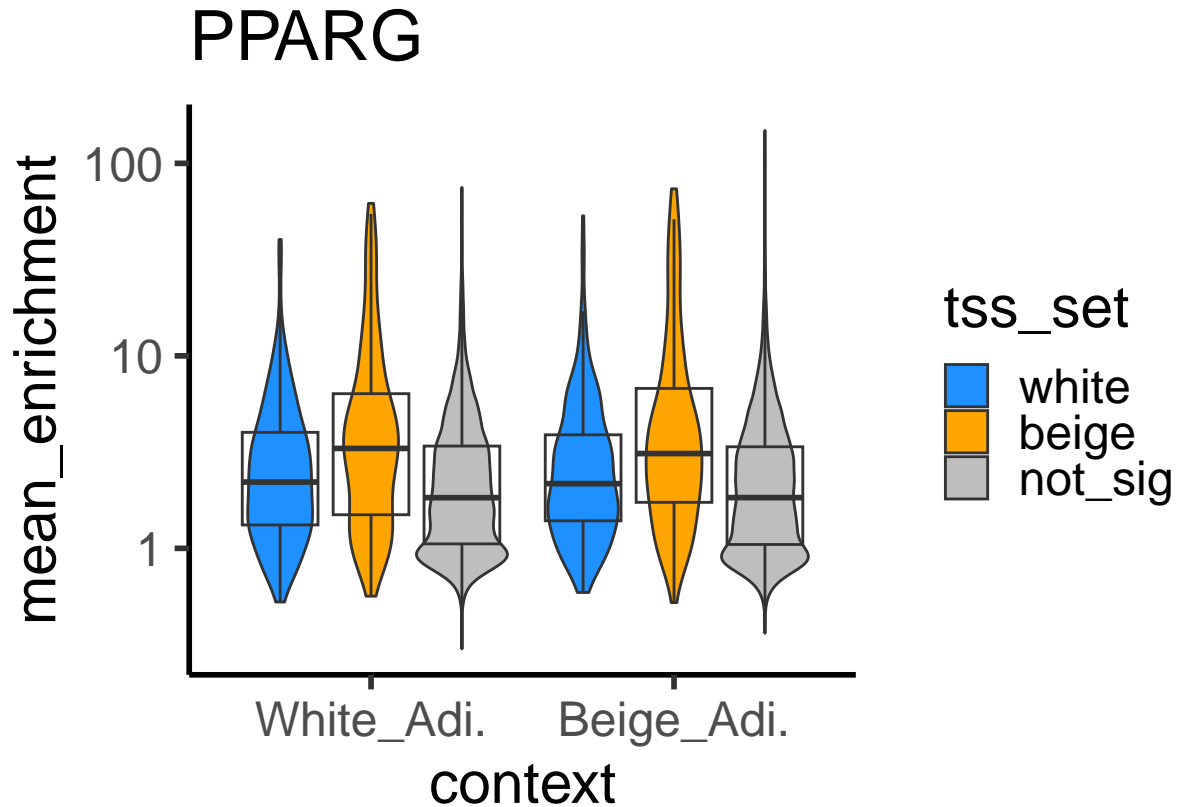
```
## [1] 24000
```

```
long = pivot_longer(annot, grep("Adi.",colnames(annot)), names_to = "context", values_to = "mean_enrich
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","orar
  ggtitle(mark) + scale_y_log10()
```

PPARG


```
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)
```

PPARG


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

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

```r
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  17.127 151.481 <2e-16 ***
## context        1      0   0.091   0.801  0.371
## Residuals  47996   5427   0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
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   3.13e- 5 9.40e- 5 3.1e-05  ****     Wilco~
## 2 White_Adi. logenrich~ white  not_s~  1.07e- 3 1.1 e- 3 0.00107  **       Wilco~
## 3 White_Adi. logenrich~ beige  not_s~  1.49e-17 7.5 e-17 < 2e-16  ****     Wilco~
## 4 Beige_Adi. logenrich~ white  beige   1.89e- 5 7.6 e- 5 1.9e-05  ****     Wilco~
## 5 Beige_Adi. logenrich~ white  not_s~  2.08e- 4 4.2 e- 4 0.00021  ***      Wilco~
## 6 Beige_Adi. logenrich~ beige  not_s~  2.44e-19 1.5 e-18 < 2e-16  ****     Wilco~
```

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

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## tss_set          2     34  17.127 151.476 <2e-16 ***
## context          1      0   0.091   0.801  0.371
## tss_set:context  2      0   0.030   0.268  0.765
## Residuals    47994   5427   0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MED1

```r
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, "
  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': 232 obs. of  6 variables:
##   ..$ chr       : chr [1:232] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start     : int [1:232] 6614356 14923629 23799901 45339454 55214861 87129264 113757562 1469593!
##   ..$ end       : int [1:232] 6615357 14924630 23800902 45340455 55215862 87130265 113758563 1469603!
##   ..$ White_Adi.: num [1:232] 1.34 1.11 2.5 1.71 2.93 ...
##   ..$ Beige_Adi.: num [1:232] 1.16 1.04 2.23 1.56 2.8 ...
##   ..$ tss_set   : chr [1:232] "beige" "beige" "beige" "beige" ...
##  $ white :'data.frame': 239 obs. of  6 variables:
##   ..$ chr       : chr [1:239] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start     : int [1:239] 6613230 11803176 14945418 23800280 33349384 55214863 89632571 93613596
##   ..$ end       : int [1:239] 6614231 11804177 14946419 23801281 33350385 55215864 89633572 93614597
##   ..$ White_Adi.: num [1:239] 3.14 1.86 1.17 2.57 1.41 ...
##   ..$ Beige_Adi.: num [1:239] 2.28 1.77 1.51 2.23 1.61 ...
##   ..$ tss_set   : chr [1:239] "white" "white" "white" "white" ...
##  $ not_sig:'data.frame': 23529 obs. of  6 variables:
##   ..$ chr       : chr [1:23529] "chr1" "chr1" "chr1" "chr1" ...
##   ..$ start     : int [1:23529] 28869 502372 502372 513912 516751 522427 945752 950657 960948 1219792
##   ..$ end       : int [1:23529] 29870 503373 503373 514913 517752 523428 946753 951658 961949 1220793
##   ..$ White_Adi.: num [1:23529] 1.87 0.971 0.971 1.017 0.919 ...
##   ..$ Beige_Adi.: num [1:23529] 1.26 1.02 1.02 1.93 1.09 ...
##   ..$ tss_set   : chr [1:23529] "not_sig" "not_sig" "not_sig" "not_sig" ...
```

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

```
##
##  beige not_sig  white
##    232  23529     239
```

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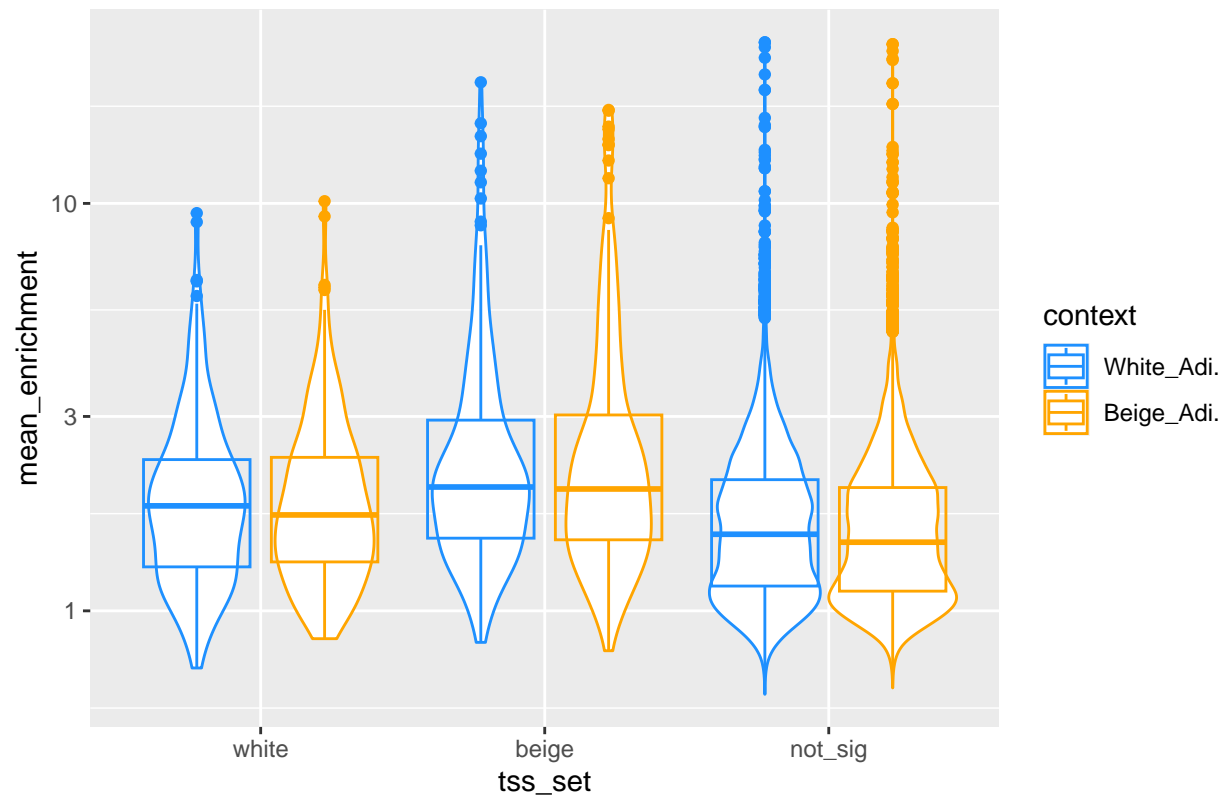
```
##           chr    start       end White_Adi. Beige_Adi. tss_set
## beige.1 chr1  6614356  6615357   1.338763   1.164247   beige
## beige.2 chr1 14923629 14924630   1.110319   1.043586   beige
## beige.3 chr1 23799901 23800902   2.502486   2.227372   beige
## beige.4 chr1 45339454 45340455   1.707652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

```
## [1] 24000
```

```
long = pivot_longer(annot, grep("Adi.",colnames(annot)), names_to = "context", values_to = "mean_enrichm
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","orar
  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)
```
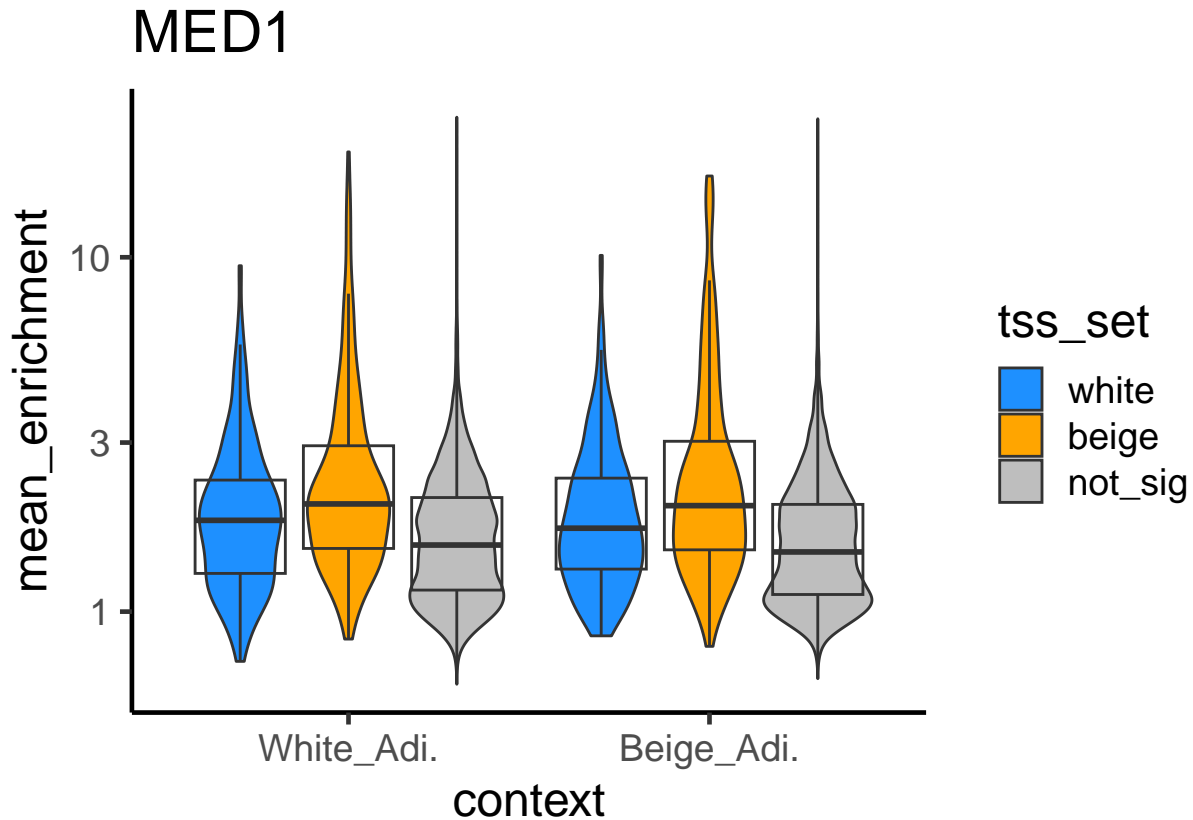
MED1


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

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

```r
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   14.5   7.273 237.931 <2e-16 ***
## context           1    2.4   2.419  79.134 <2e-16 ***
## tss_set:context   2    0.1   0.045   1.459  0.232
## Residuals     47994 1467.0   0.031
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
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. logenrichm~ white  beige  1.25e- 4 1.3e- 4 0.00013  ***      Wilco~
## 2 White_Adi. logenrichm~ white  not_s~ 5.48e- 6 1.6e- 5 5.5e-06  ****     Wilco~
## 3 White_Adi. logenrichm~ beige  not_s~ 3.98e-22 2  e-21 < 2e-16  ****     Wilco~
## 4 Beige_Adi. logenrichm~ white  beige  6.71e- 5 1.3e- 4 6.7e-05  ****     Wilco~
```

```
## 5 Beige_Adi. logenrichm~ white  not_s~ 2.59e- 8 1  e- 7 2.6e-08  ****    Wilco~
## 6 Beige_Adi. logenrichm~ beige  not_s~ 1.01e-25 6.1e-25 < 2e-16  ****    Wilco~
```

```r
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~ 7.71e- 1 1  e+ 0 0.77     ns       Wilco~
## 3 not_sig logenrichment White_~ Beige~ 2.55e-21 7.6e-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.