

2023 Analysis Graphics

Michelle Bang

2023-02-28

```
library(tidyverse)

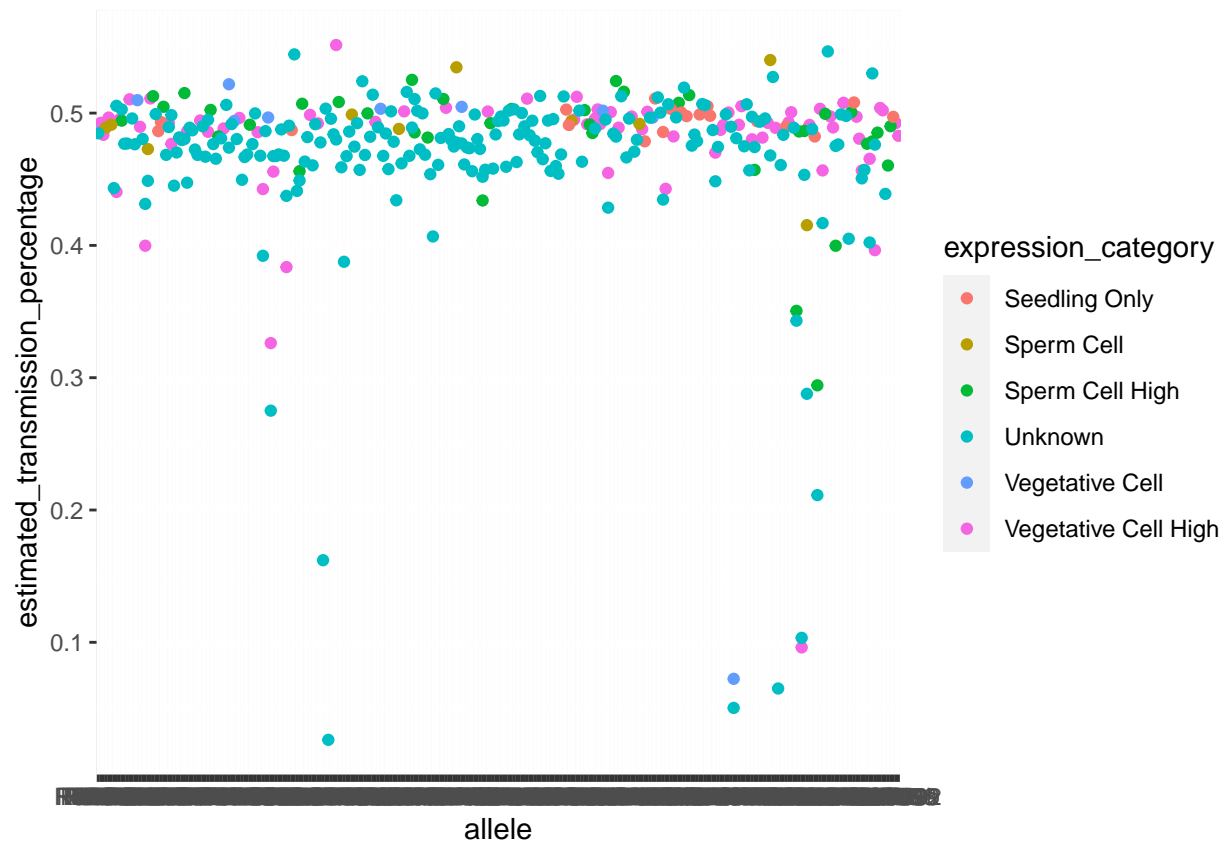
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.6    v dplyr  1.0.7
## v tidyr   1.1.4    v stringr 1.4.0
## v readr   2.1.1    v forcats 0.5.1

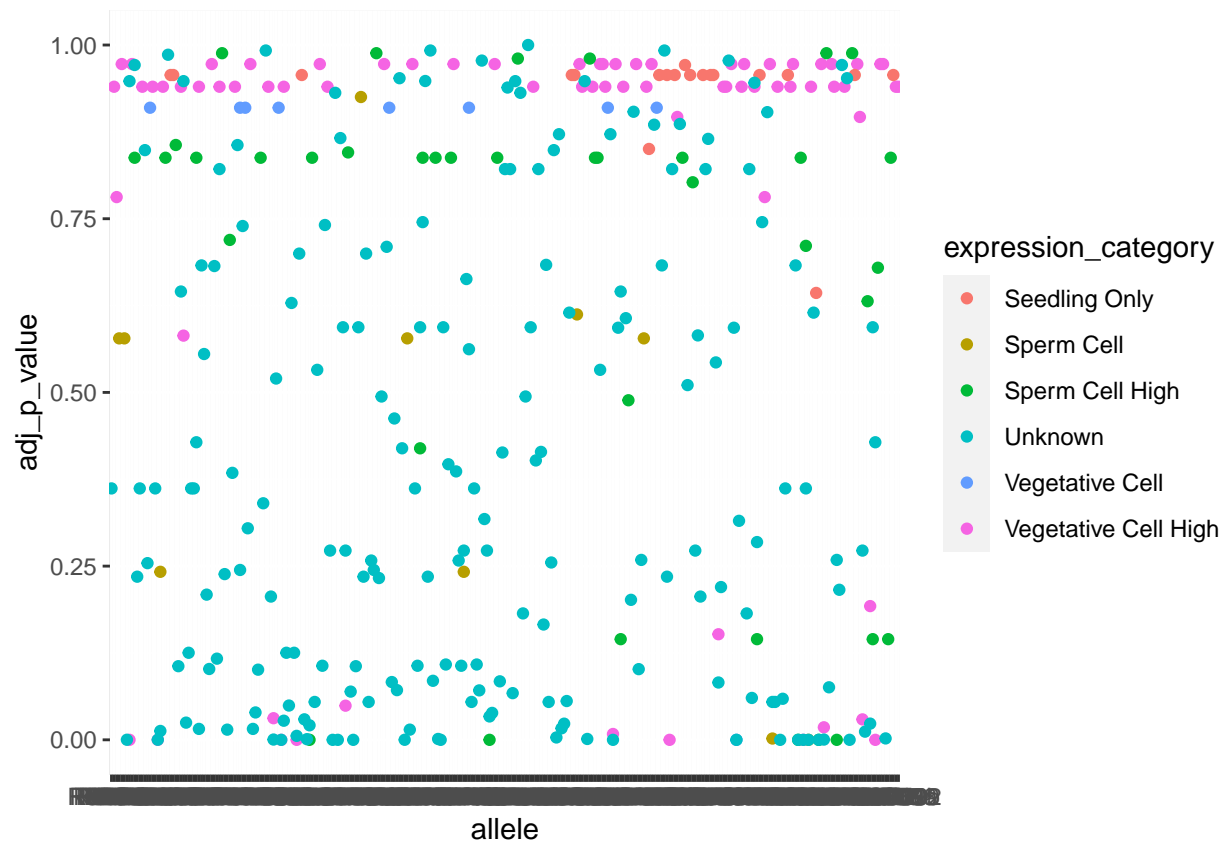
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

asn <- read.table(file = "2023_analysis_sum.tsv",
                  header = TRUE,
                  sep = '\t')

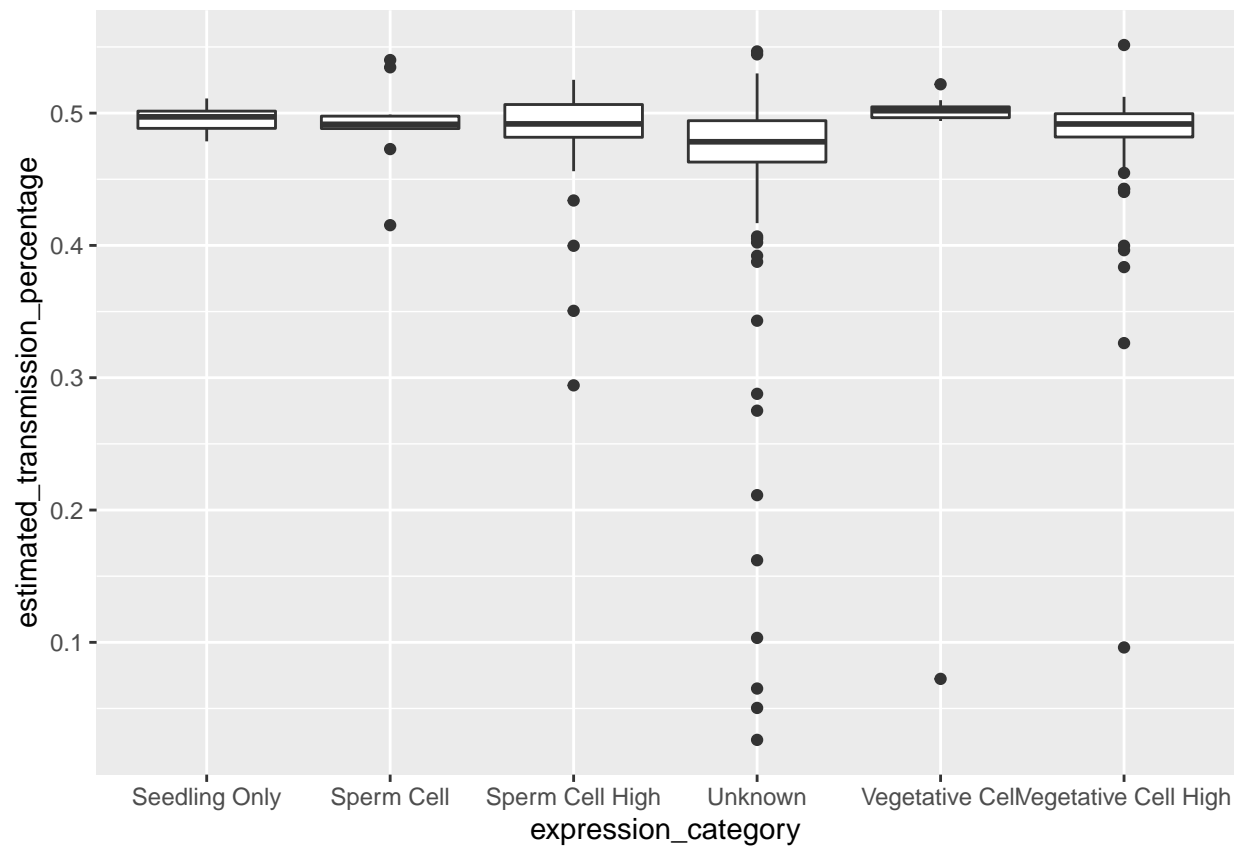
ggplot(asn, mapping = aes(x = allele,
                          y = estimated_transmission_percentage,
                          col = expression_category)) + geom_point()
```



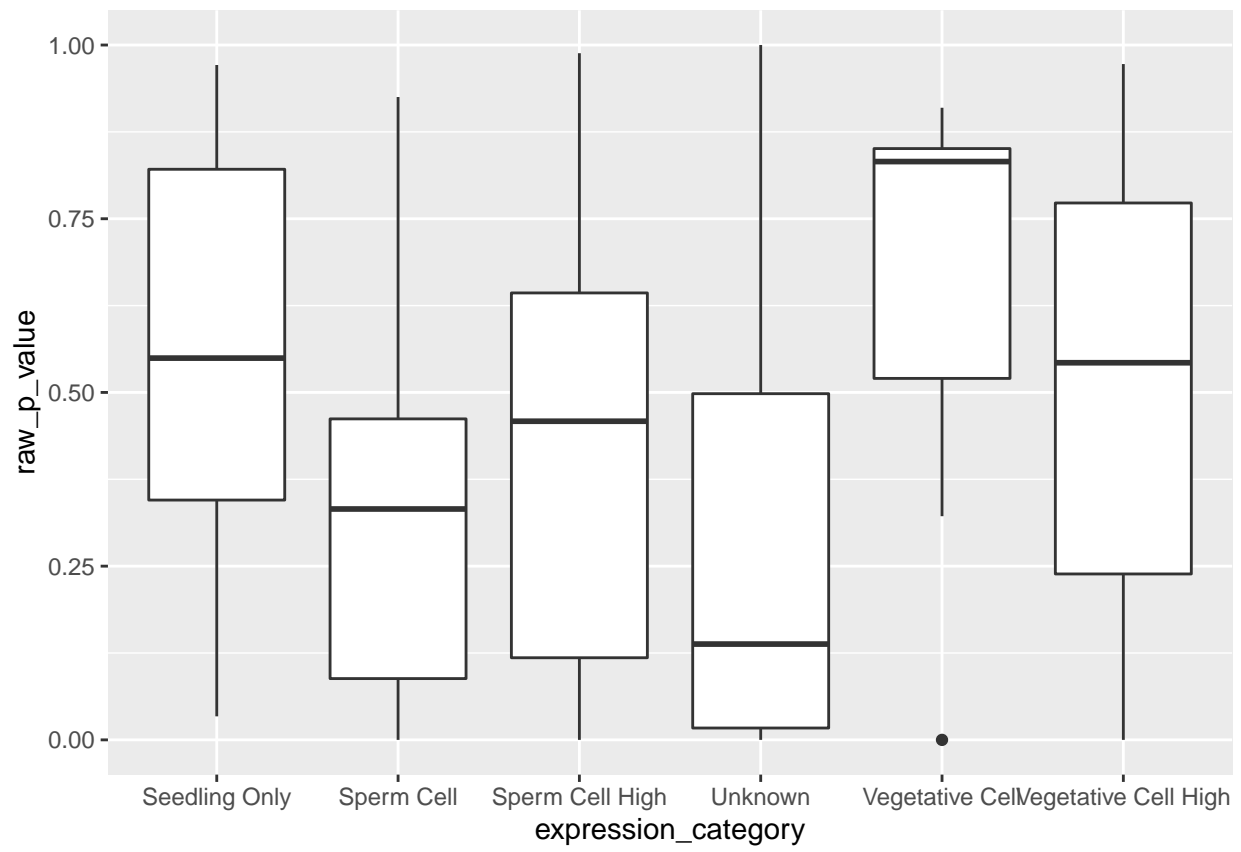
```
ggplot(asn, mapping = aes(x = allele,
                           y = adj_p_value,
                           col = expression_category)) + geom_point()
```



```
ggplot(asn, mapping = aes(x = expression_category,
                          y = estimated_transmission_percentage)) + geom_boxplot()
```



```
ggplot(asn, mapping = aes(x = expression_category,
                           y = raw_p_value)) + geom_boxplot()
```



#ADDITIONAL REQUESTS FROM JOHN (02/28)

```
ordered_df <- asn %>%  
  arrange(desc(estimated_transmission_percentage))
```

```
write.table(ordered_df, file = "2023_analysis_sum_ordered.tsv", sep = "\t", row.names=FALSE)
```

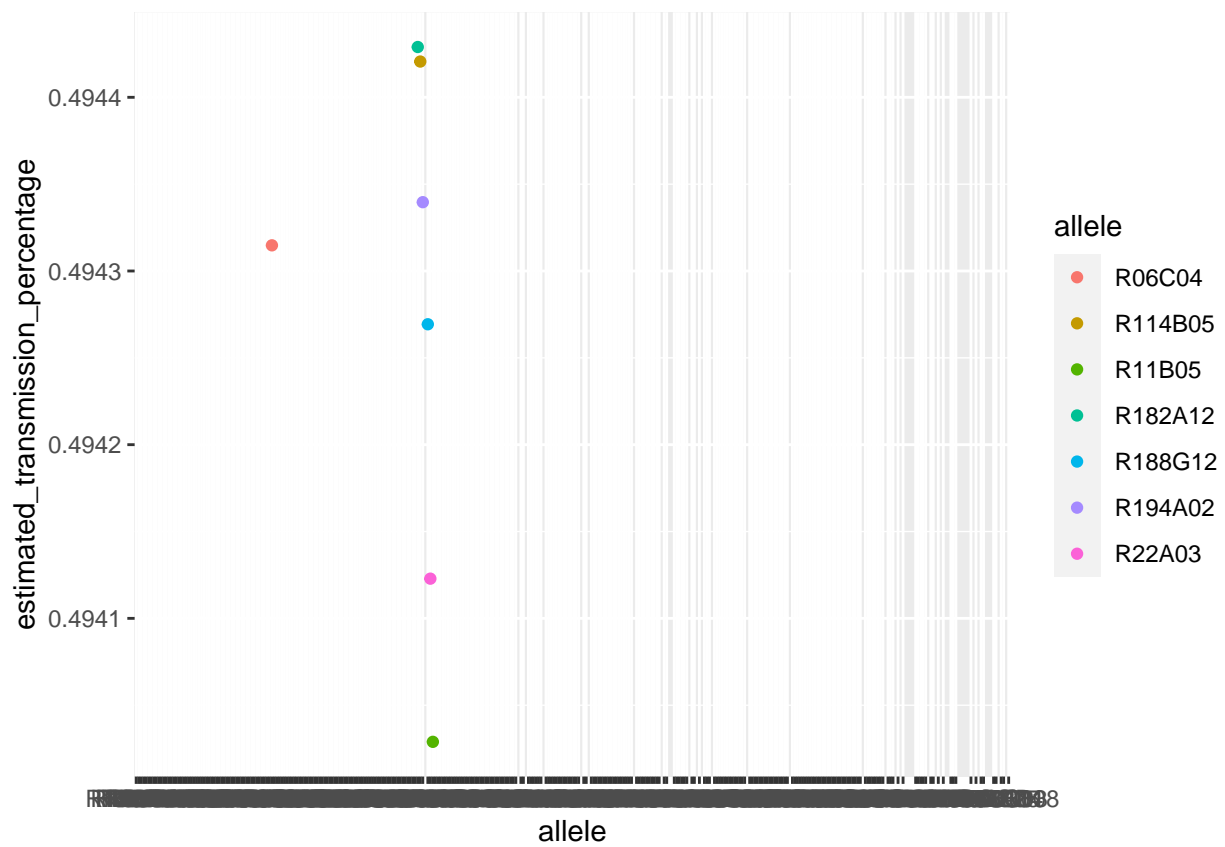
ignore this part... this is me looking into the graph...

```
ordered_df %>%  
  filter(allele == "R06C04")
```

```
##   allele logit_of_percentage estimated_transmission_percentage  
## 1 R06C04      0.01111123              0.5027778  
## 2 R06C04     -0.02274185              0.4943148  
##   lower_confidence_interval upper_confidence_interval raw_p_value adj_p_value  
## 1              0.4367791              0.5686798    0.9345915    0.9712513  
## 2              0.4689328              0.5197261    0.6613608    0.8377236  
##   expression_category  
## 1              Unknown  
## 2      Sperm Cell High
```

```
small_df <- ordered_df %>%  
  filter(estimated_transmission_percentage > .4940 & estimated_transmission_percentage < .4945)
```

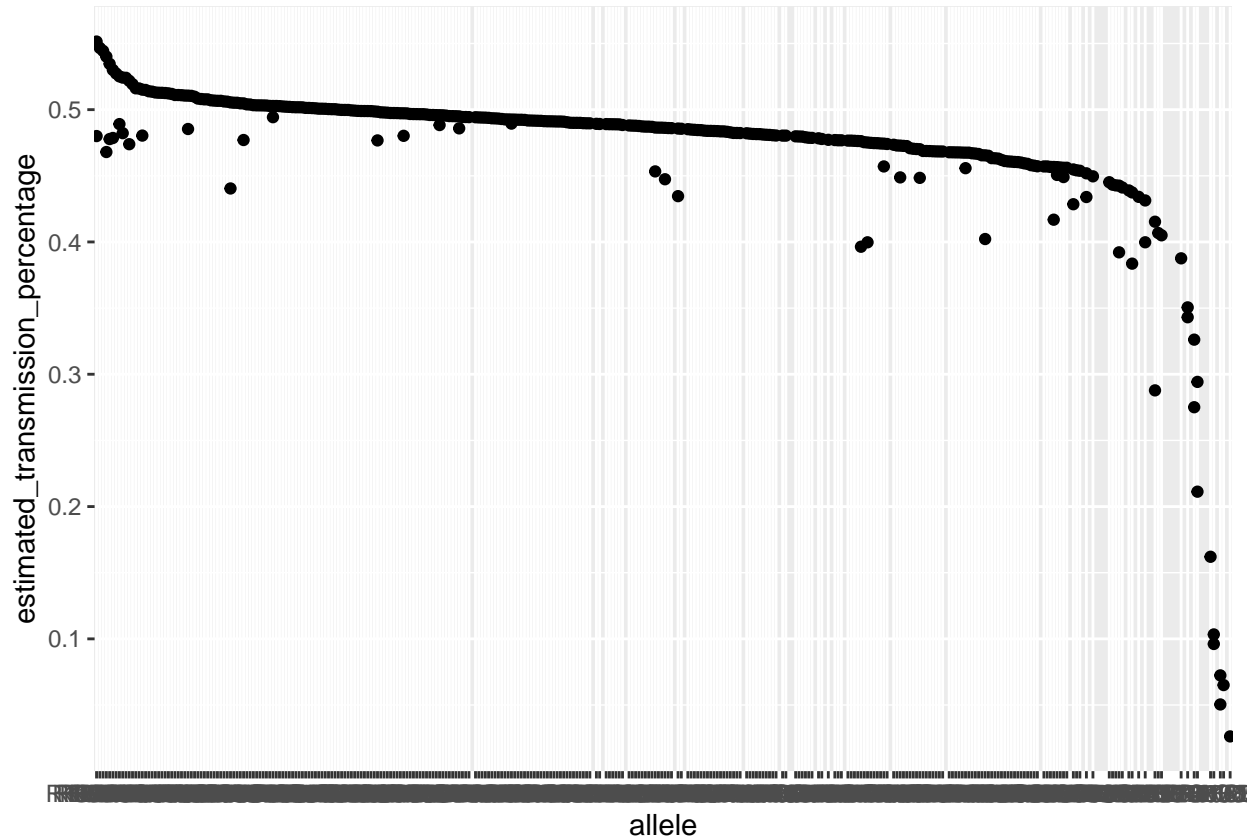
```
sm_plot <- ggplot(small_df, mapping = aes(x = allele,  
                                           y = estimated_transmission_percentage,  
                                           col = allele)) + geom_point()  
sm_plot + scale_x_discrete(limits = ordered_df[["allele"]])
```



—ignore up to here—

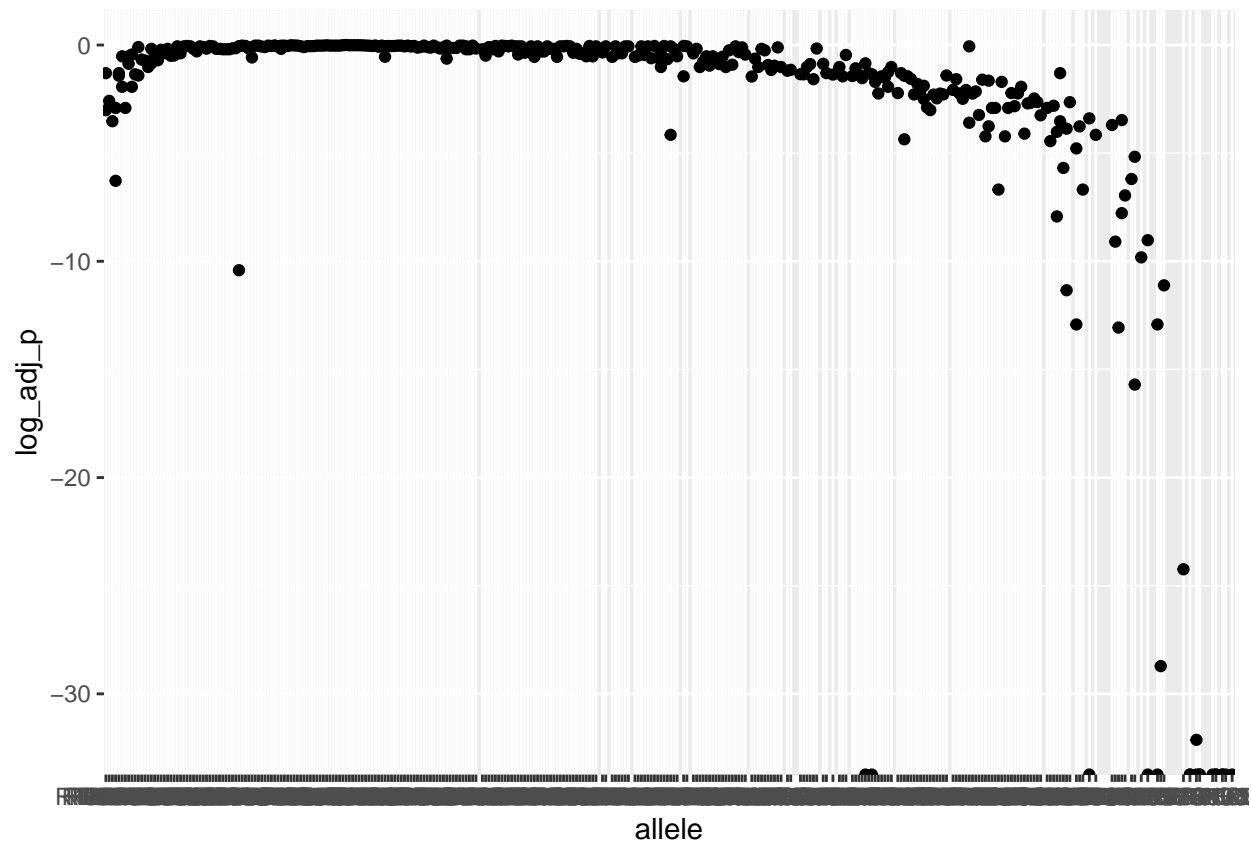
GRAPHS

```
ordered_plot <- ggplot(ordered_df, mapping = aes(x = allele,  
                                                  y = estimated_transmission_percentage )) + geom_point()  
  
ordered_plot + scale_x_discrete(limits = ordered_df[["allele"]])
```

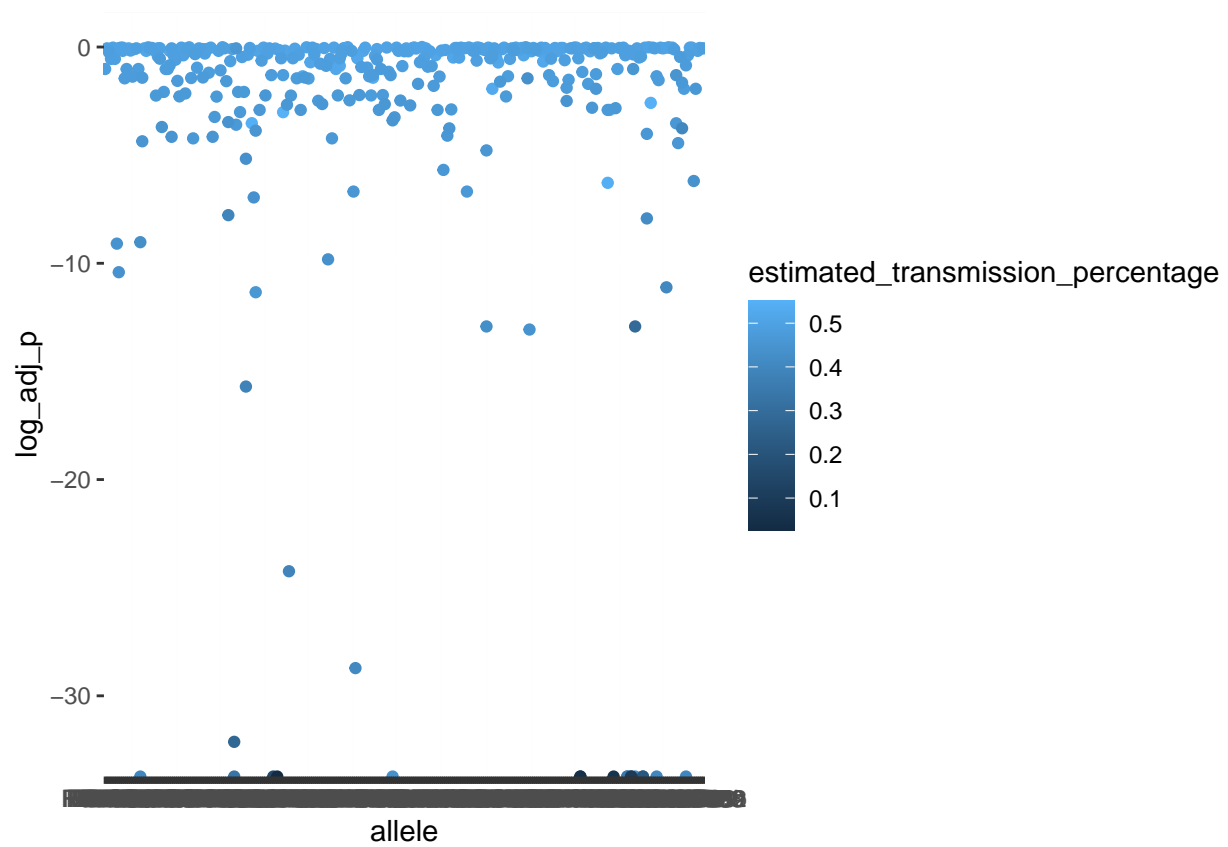


```
with_log_df <- ordered_df %>%  
  mutate(log_adj_p = log(adj_p_value)) %>%  
  #only R187B08 had log(adj_p_value) of 0  
  mutate(log_adj_p = if_else(log_adj_p == 0, 1.0*10(-15), log_adj_p))
```

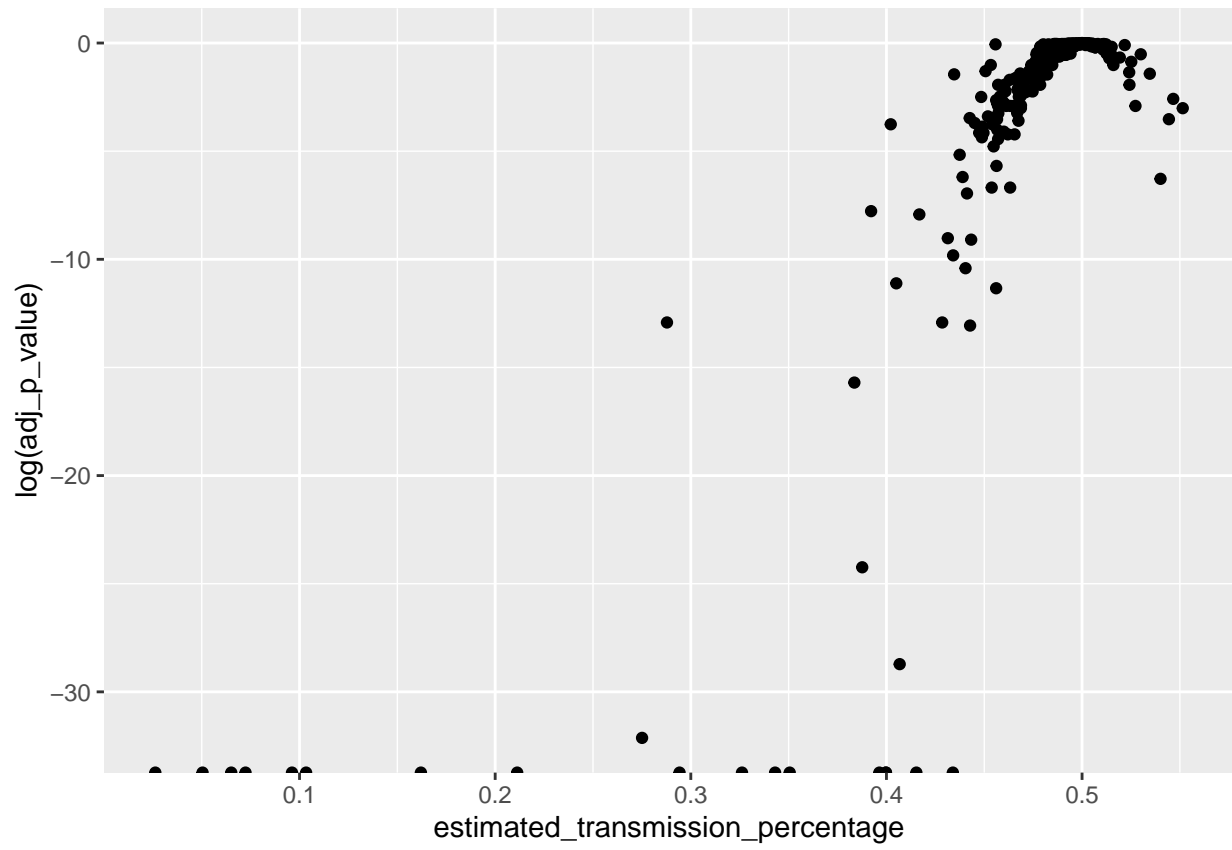
```
ordered_log <- ggplot(with_log_df, mapping = aes(x = allele,  
                                                  y = log_adj_p)) + geom_point()  
  
ordered_log + scale_x_discrete(limits = ordered_df[["allele"]])
```

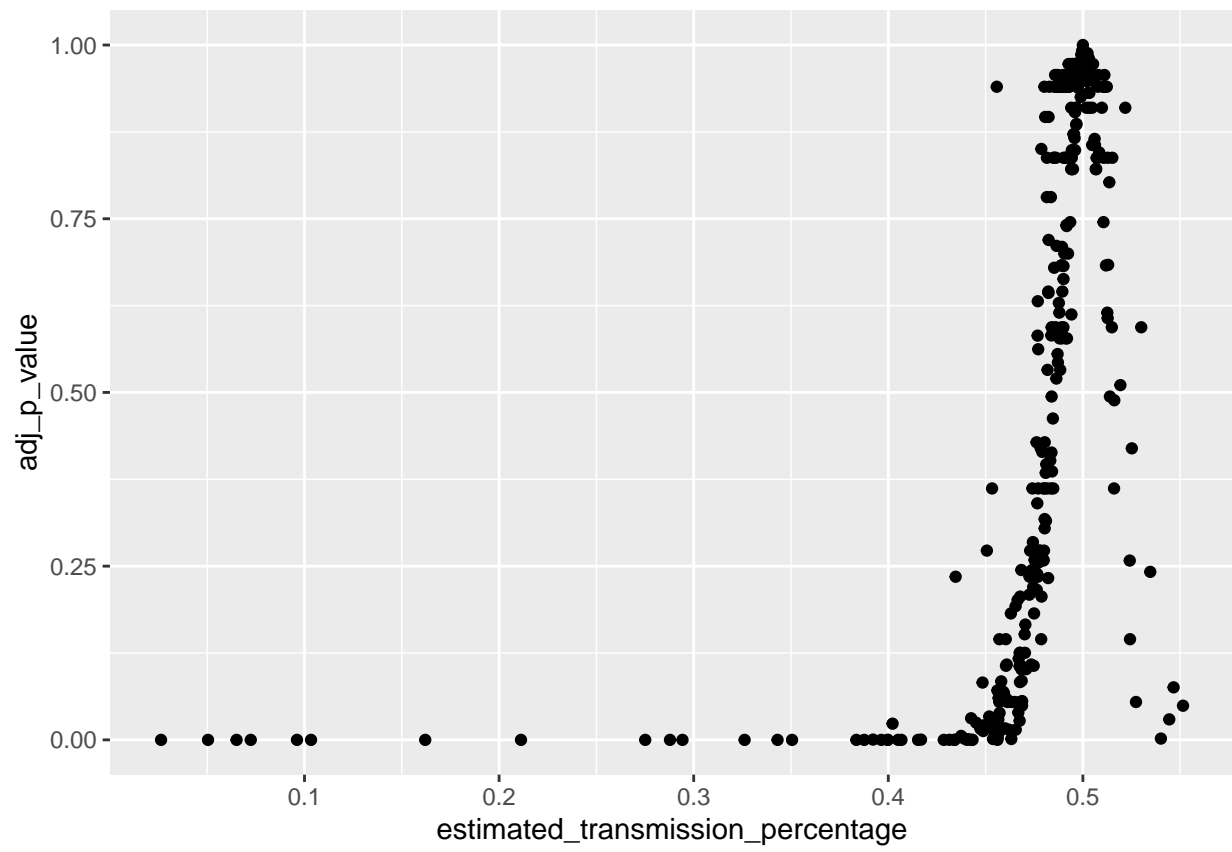
```
ggplot(with_log_df, mapping = aes(x = allele,  
  y = log_adj_p,  
  col = estimated_transmission_percentage)) + geom_point()
```



```
ggplot(with_log_df, mapping = aes(x = estimated_transmission_percentage,  
                                  y = log(adj_p_value))) + geom_point()
```



```
ggplot(asn, mapping = aes(x = estimated_transmission_percentage,  
                           y = adj_p_value)) + geom_point()
```

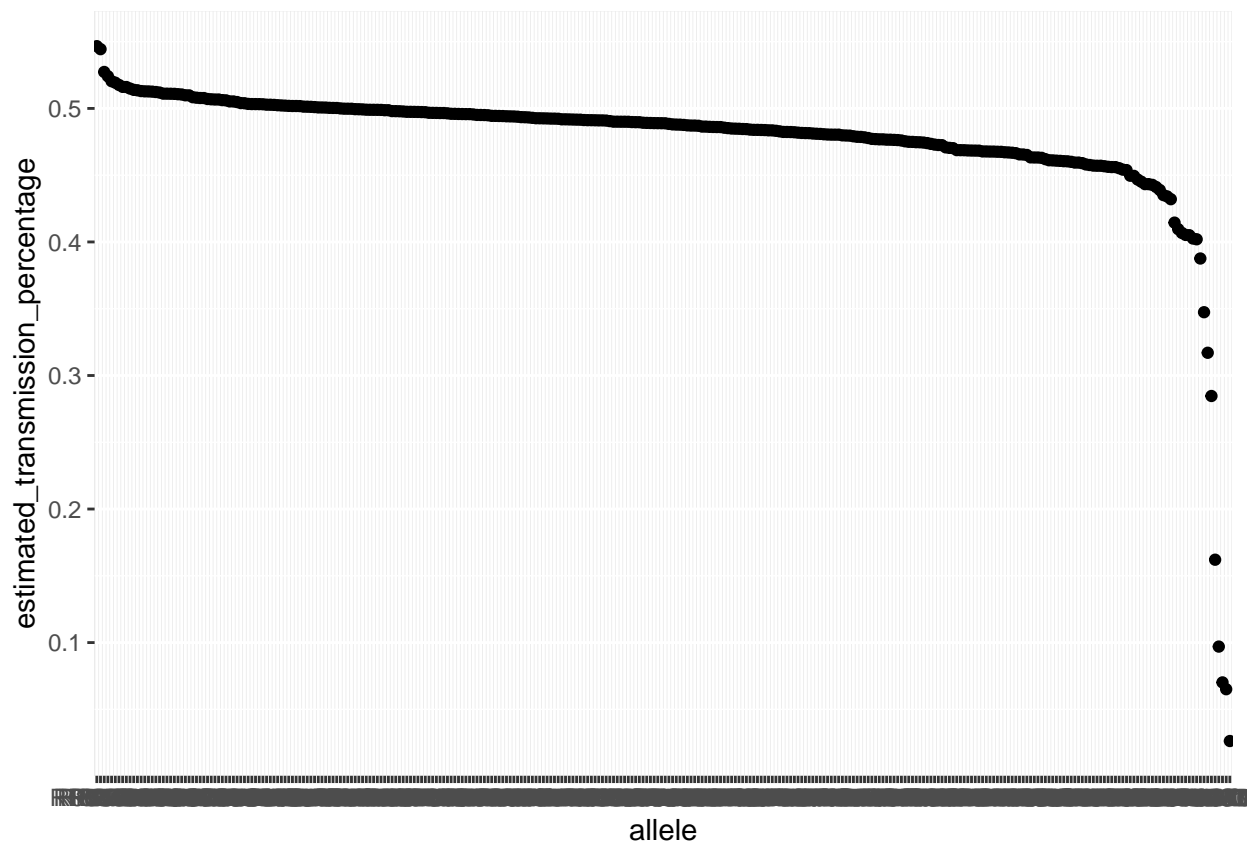


#NEW GRAPHICS...

```
asn_new <- read.table(file = "2023_noEC_analysis_sum.tsv",  
                      header = TRUE,  
                      sep = '\t')
```

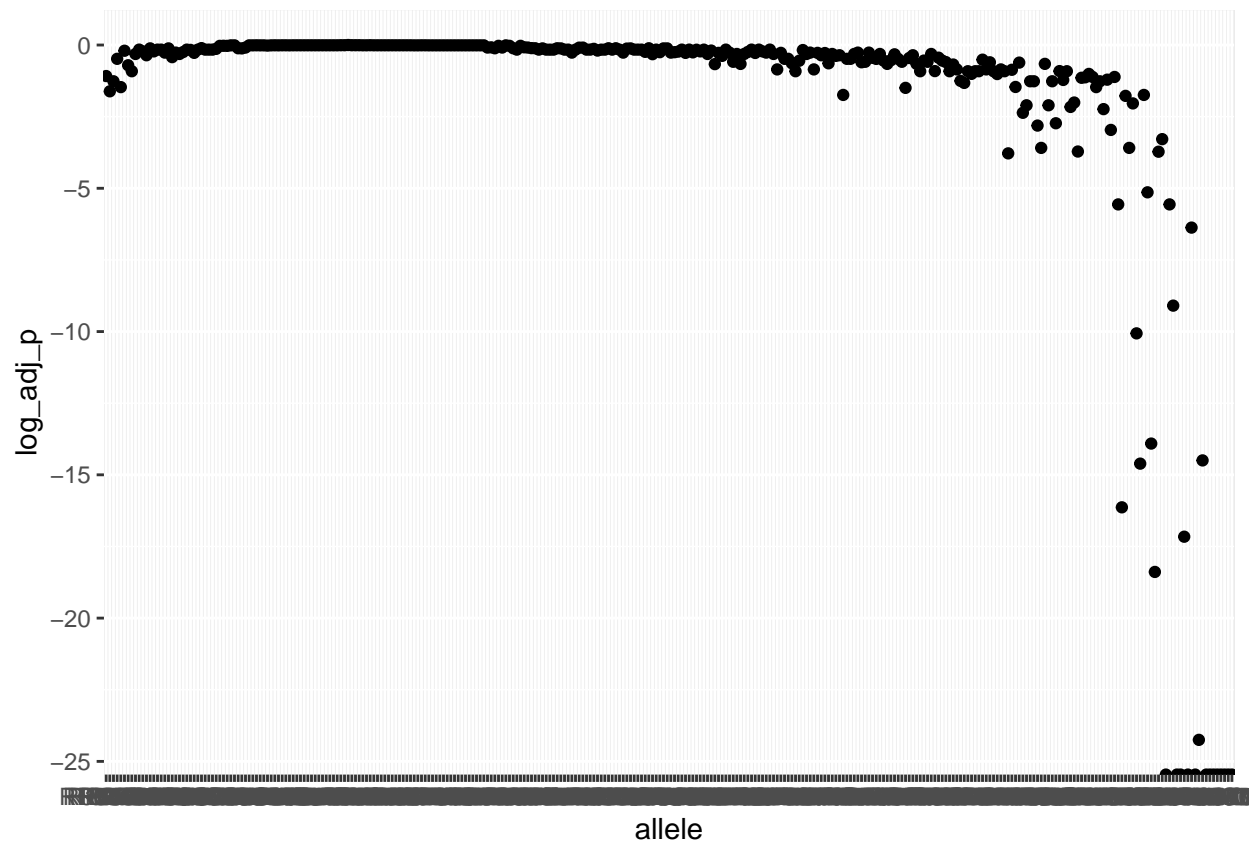
```
ordered_new <- asn_new %>%  
  arrange(desc(estimated_transmission_percentage))
```

```
ordered_plot_new <- ggplot(ordered_new, mapping = aes(x = allele,  
                                                       y = estimated_transmission_percentage )) + geom_point()  
  
ordered_plot_new + scale_x_discrete(limits = ordered_new[["allele"]])
```

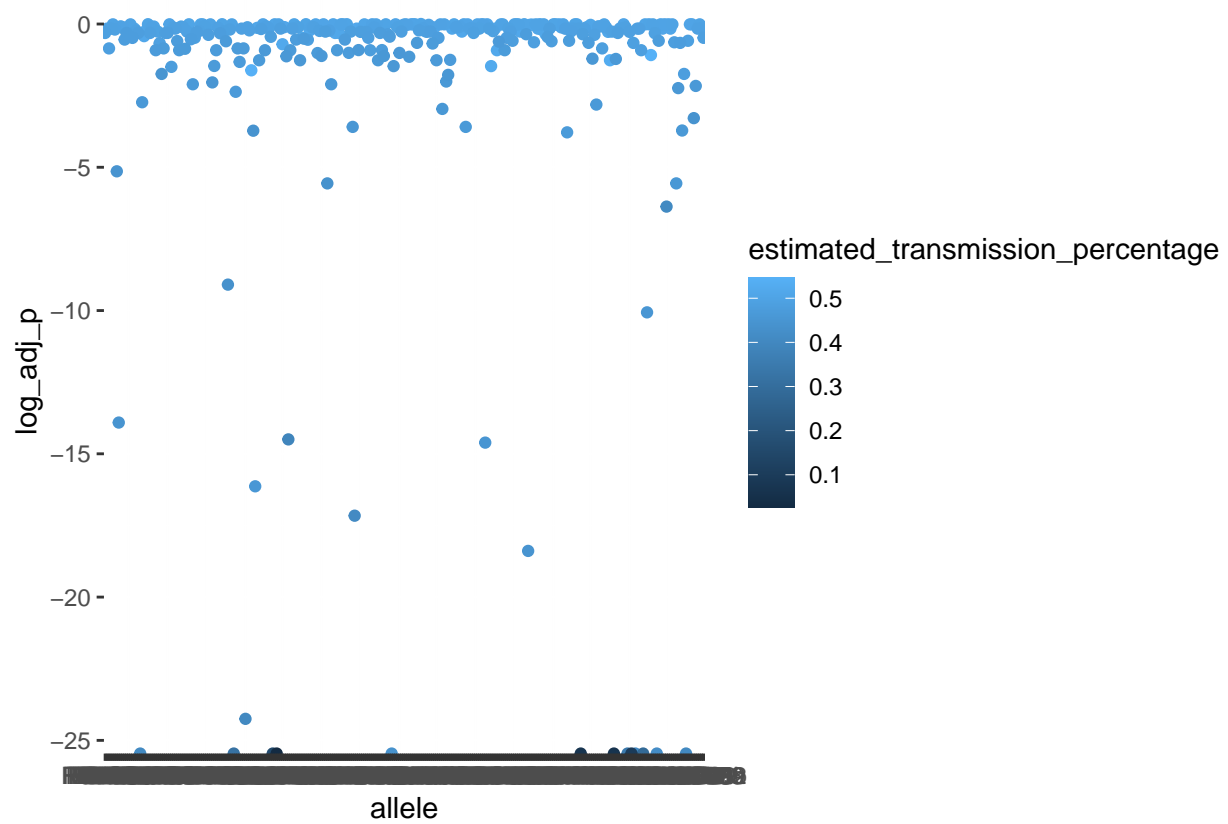


```
with_log_df_new <- ordered_new %>%  
  mutate(log_adj_p = log(adj_p_value)) %>%  
  #only R187B08 had log(adj_p_value) of 0  
  mutate(log_adj_p = if_else(log_adj_p == 0, 1.0*10(-15), log_adj_p))
```

```
ordered_log_new <- ggplot(with_log_df_new, mapping = aes(x = allele,  
                                                         y = log_adj_p)) + geom_point()  
  
ordered_log_new + scale_x_discrete(limits = ordered_new[["allele"]])
```



```
ggplot(with_log_df_new, mapping = aes(x = allele,  
  y = log_adj_p,  
  col = estimated_transmission_percentage)) + geom_point()
```



```
logcolor <- ggplot(with_log_df_new, mapping = aes(x = allele,
  y = log_adj_p,
  col = estimated_transmission_percentage)) + geom_point()

logcolor + scale_x_discrete(limits = ordered_new[["allele"]])
```

