# Exploratory Data Analysis

```r
# Load packages and read Files
library(tidyverse)
library(data.table)
library(scales)
library(metR)

df.ZIKV <- fread("GSE207347_A1B1_vs_A2B2_ZIKV_ribodiff_name.txt.gz")
df.DENV <- fread("GSE207347_A1B1_vs_A3B3_DENV_ribodiff_name.txt.gz")
df.ZIKV_DE <- fread("GSE207347_ZIKV_DESeq2_result_name.txt.gz")
df.DENV_DE <- fread("GSE207347_DENV_DESeq2_result_name.txt.gz")
```

```r
# Figure 1 B
fig1b <- ggplot(subset(df.ZIKV_DE, !is.na(padj)),
                aes(log2FoldChange, -log10(padj)))

fig1b +
  geom_point(aes(colour = -log10(padj) > -log10(0.05)), show.legend = FALSE) +
  labs(title = "ZIKV") + ylab("p adj(-log10)") +
  xlab("RNA log2Fold change \n (Infected vs uninfected)")  +
  geom_vline(xintercept = 0) + geom_hline(yintercept = -log10(0.05)) +
  scale_color_manual(values = c('black', 'red')) +
  annotate("text", x = -1.5, y = 20,
           label = "Down regulated \n q<0.05 \n (n=335)",
           col = "black") +
  annotate("text", x = 1.5, y = 20,
           label = "Up regulated \n q<0.05 \n (n=445)",
           col = "black") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black",
                                    fill = NA, linewidth = 1.5))
```

ZIKV

Down regulated
q<0.05
(n=335)

Up regulated
q<0.05
(n=445)
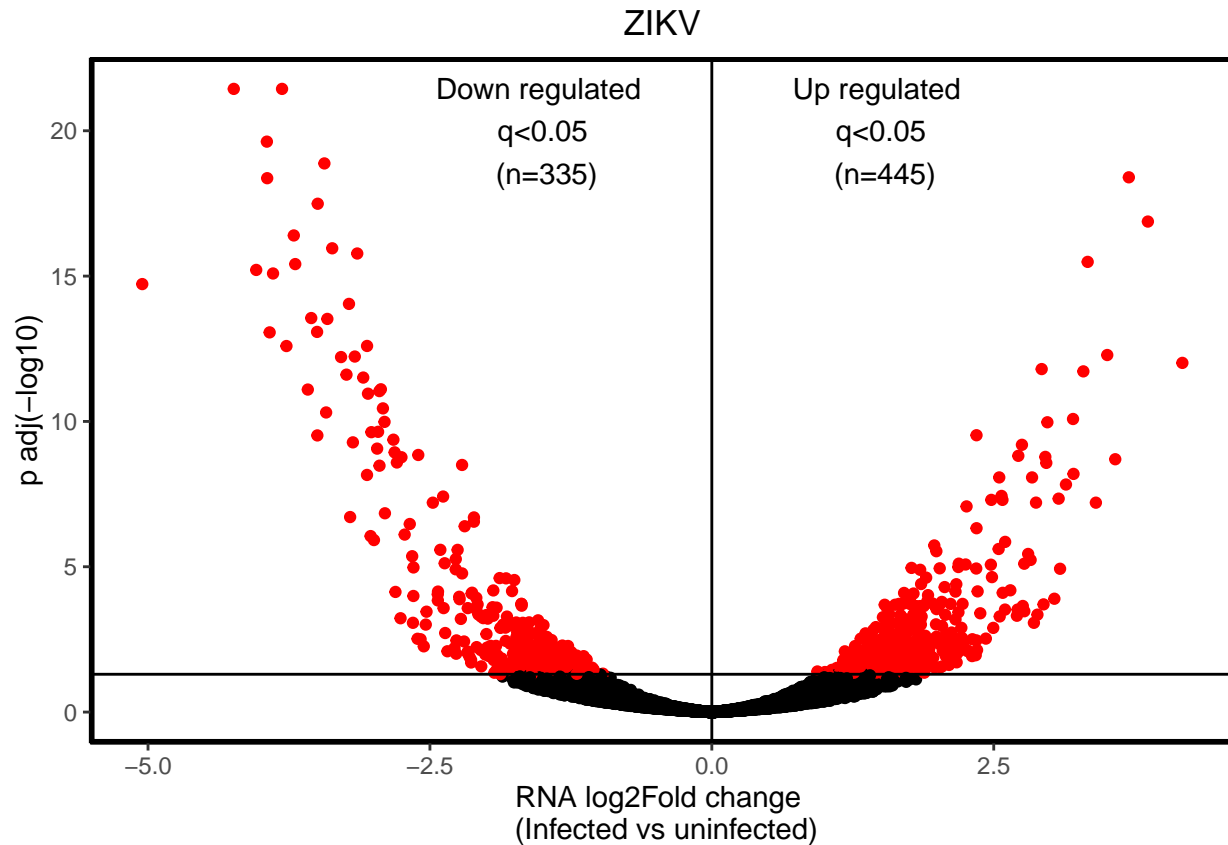
```r
df.ZIKV_DE.sub <- subset(df.ZIKV_DE, !is.na(padj))
# Count of Up regulated with q < 0.05
(df.ZIKV_DE.sub$log2FoldChange > 0 & df.ZIKV_DE.sub$padj < 0.05) %>% sum()
```

```
## [1] 445
```

```r
# Count of Down regulated with q < 0.05
(df.ZIKV_DE.sub$log2FoldChange < 0 & df.ZIKV_DE.sub$padj < 0.05) %>% sum()
```
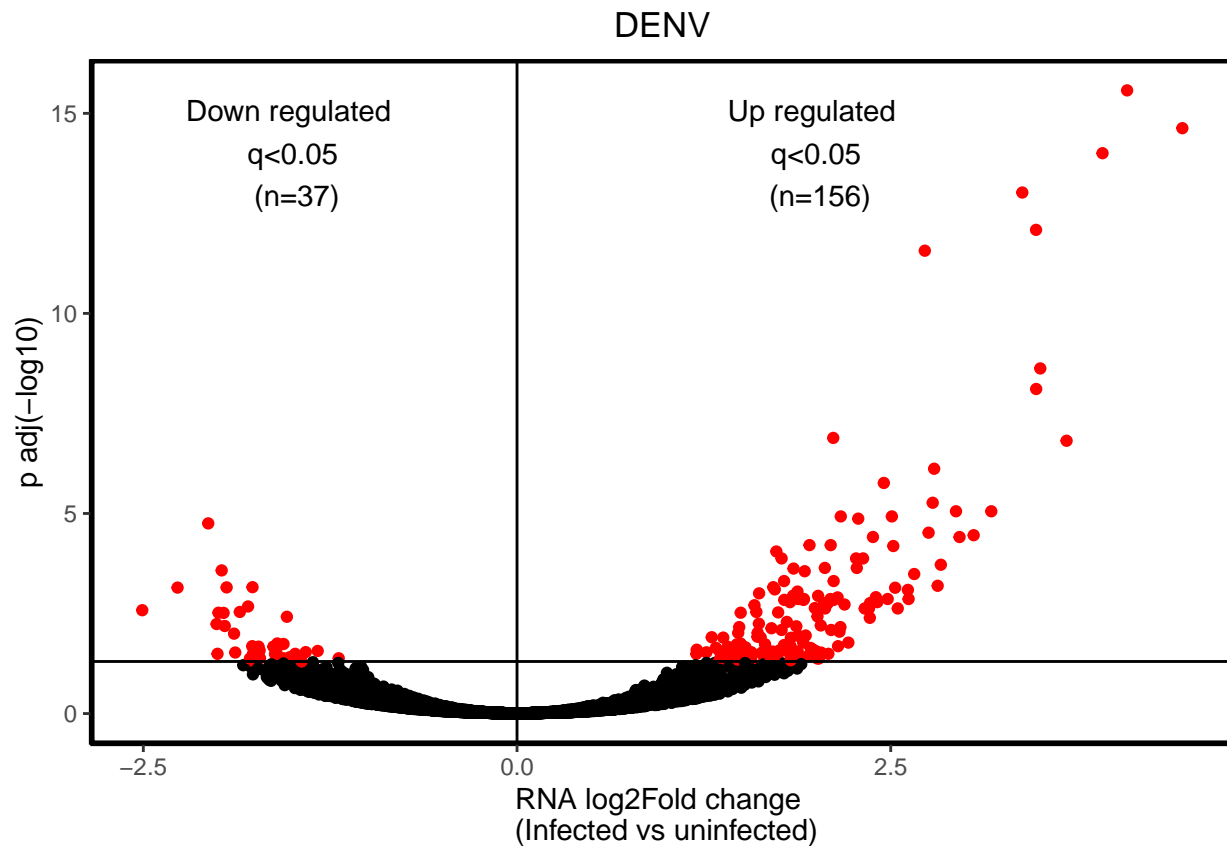
```
## [1] 335
```

```r
# Figure 1 C
fig1c <- ggplot(subset(df.DENV_DE, !is.na(padj)),
                aes(log2FoldChange, -log10(padj)))

fig1c +
  geom_point(aes(colour = -log10(padj) > -log10(0.05)), show.legend = FALSE) +
  labs(title = "DENV") + ylab("p adj(-log10)") +
  xlab("RNA log2Fold change \n (Infected vs uninfected)") +
  geom_vline(xintercept = 0) + geom_hline(yintercept = -log10(0.05)) +
  scale_color_manual(values = c('black', 'red')) +
  annotate("text", x = -1.5, y = 14,
           label = "Down regulated \n q<0.05 \n (n=37)",
           col = "black") +
```

```r
  annotate("text", x = 2, y = 14,
           label = "Up regulated \n q<0.05 \n (n=156)",
           col = "black") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black",
                                    fill = NA, linewidth = 1.5))
```



```r
df.DENV_DE.sub <- subset(df.DENV_DE, !is.na(padj))
# Count of Up regulated with q < 0.05
(df.DENV_DE.sub$log2FoldChange > 0 & df.DENV_DE.sub$padj < 0.05) %>% sum()
```

```
## [1] 156
```

```r
# Count of Down regulated with q < 0.05
(df.DENV_DE.sub$log2FoldChange < 0 & df.DENV_DE.sub$padj < 0.05) %>% sum()
```

```
## [1] 37
```

```
# Figure 2 A

ZIKV.up <- df.ZIKV_DE.sub$log2FoldChange > 0 & df.ZIKV_DE.sub$padj < 0.05
df.ZIKV_DE.up <- df.ZIKV_DE.sub[ZIKV.up,]

fig2a <- ggplot(subset(df.ZIKV_DE.up, !is.na(padj)),
                aes(x = -log10(padj), y = log2FoldChange))

fig2a +
  geom_point(aes(colour = -log10(padj) > -log10(0.001)), show.legend = FALSE) +
  xlim(0, 20) + ylim(0, 5) + labs(title = "ZIKV Upregulated RNA") +
  xlab("p adj(-log10)") +
  ylab("Up regulated RNA \n log2Fold change \n (Infected vs uninfected)") +
  scale_color_manual(values = c('black', 'red')) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.line = element_line(colour = "black"))
```
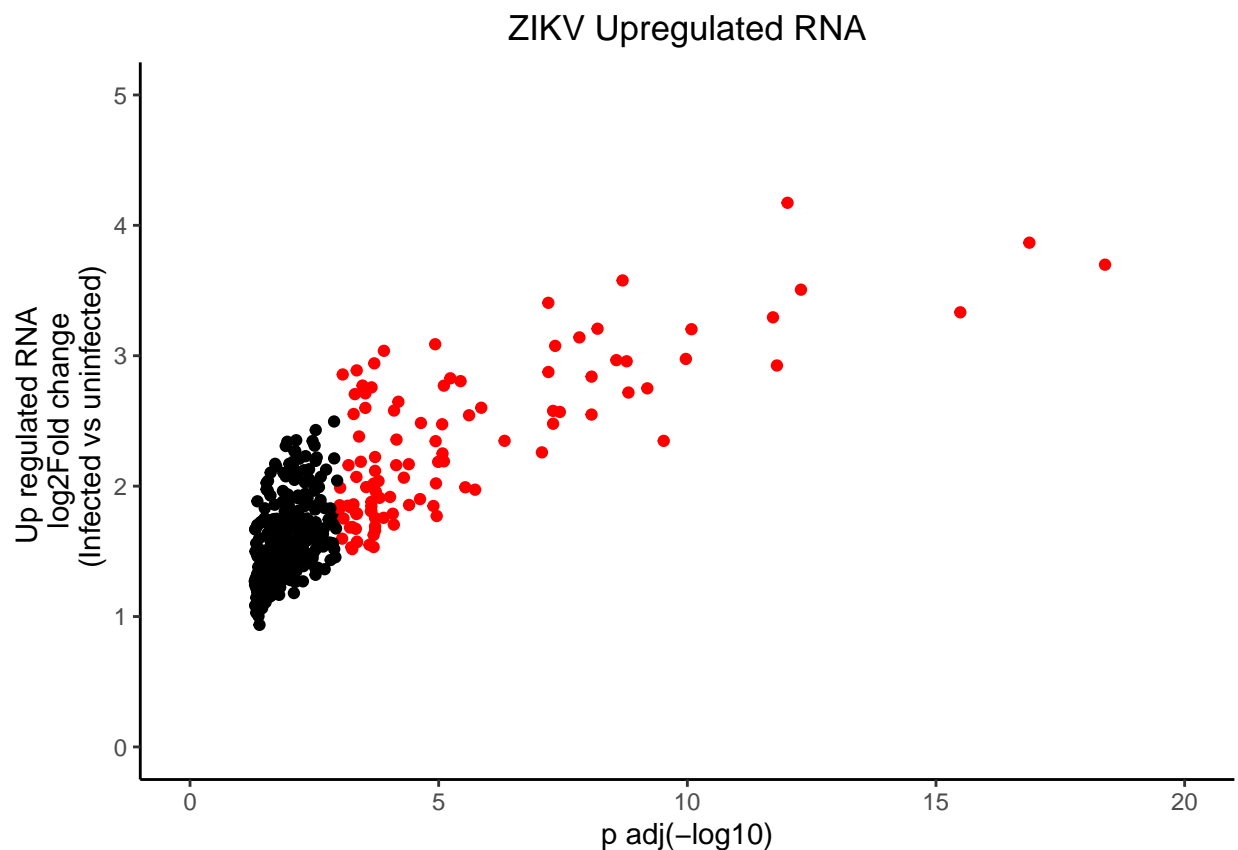


```
# Count of q < 0.001
sum(df.ZIKV_DE.up$padj < 0.001)
```

```
## [1] 104
```

```
# Figure 2 B

DENV.up <- df.DENV_DE.sub$log2FoldChange > 0 & df.DENV_DE.sub$padj < 0.05
df.DENV_DE.up <- df.DENV_DE.sub[DENV.up,]

fig2b <- ggplot(subset(df.DENV_DE.up, !is.na(padj)),
                aes(x = -log10(padj), y = log2FoldChange))

fig2b +
  geom_point(aes(colour = -log10(padj) > -log10(0.001)), show.legend = FALSE) +
  xlim(0, 20) + ylim(0, 5) + labs(title = "DENV Upregulated RNA") +
  xlab("p adj(-log10)") +
  ylab("Up regulated RNA \n log2Fold change \n (Infected vs uninfected)") +
  scale_color_manual(values = c('black', 'red')) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.line = element_line(colour = "black"))
```
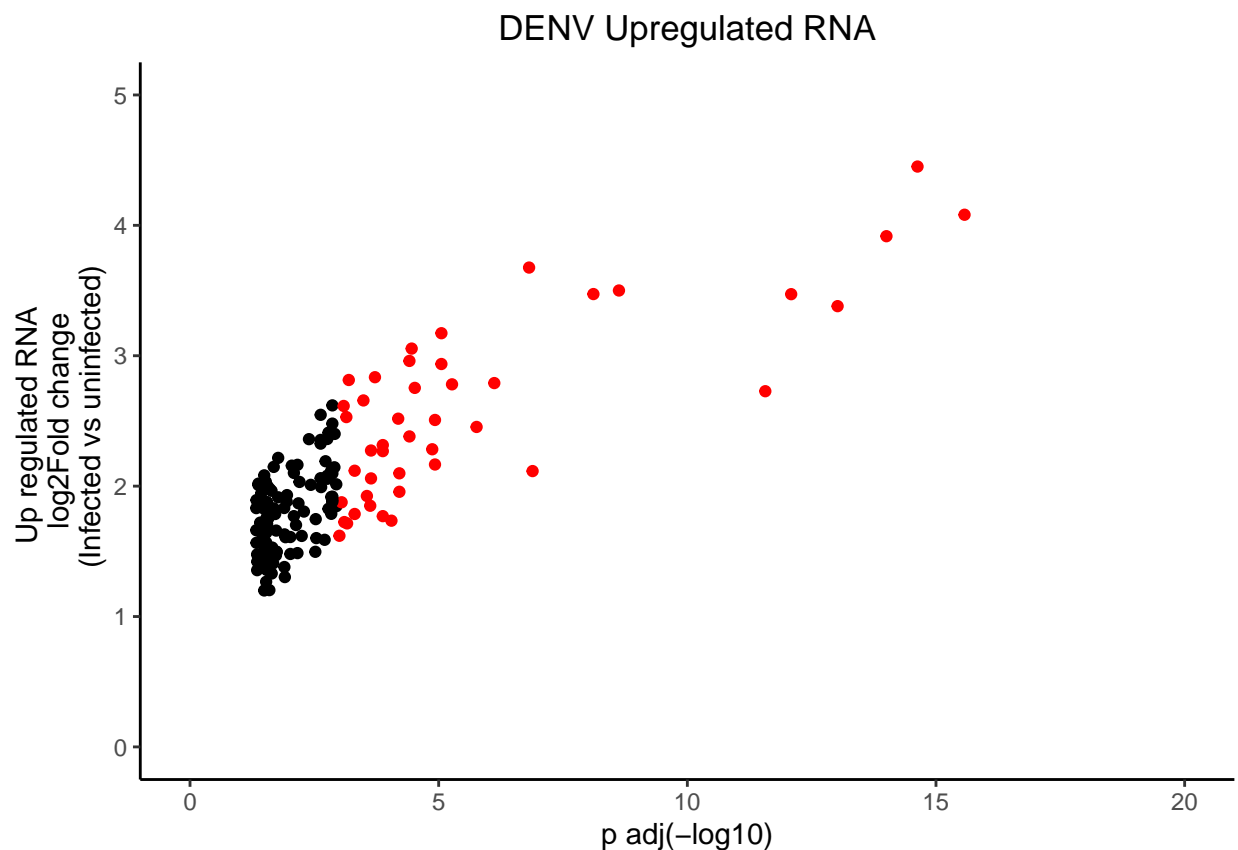


```
# Count of q < 0.001
sum(df.DENV_DE.up$padj < 0.001)
```

```
## [1] 44
```

```r
# Figure 3 A

colnames(df.ZIKV)[7] <- "log2FC_TE_DrugTreated_vs_Control"

df.ZIKV.sub <- subset(df.ZIKV, !is.na(padj))
df.ZIKV.sub <- df.ZIKV.sub[df.ZIKV.sub$padj < 0.5,]

fig3a <- ggplot(df.ZIKV.sub,
                aes(x = log2FC_TE_DrugTreated_vs_Control, y = padj))

fig3a +
  geom_point(aes(color = padj > 0.05), show.legend = FALSE) +
  scale_y_continuous(trans = reverselog_trans(base = 10),
                     breaks = c(1, 0.1, 0.01, 0.001)) +
  scale_x_continuous(limits = c(-8, 8),
                     breaks = seq(-8, 8, 2)) +
  scale_color_manual(values = c('red', 'black')) +
  xlab("log2FC_TE(ZIKV infected vs control)") +
  ylab("q adj(-log10)") +
  geom_hline(yintercept=0.05, linetype=2, col="black") +
  geom_vline(xintercept=0, linetype=2, col="black") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black",
                                    fill = NA, linewidth = 1))
```
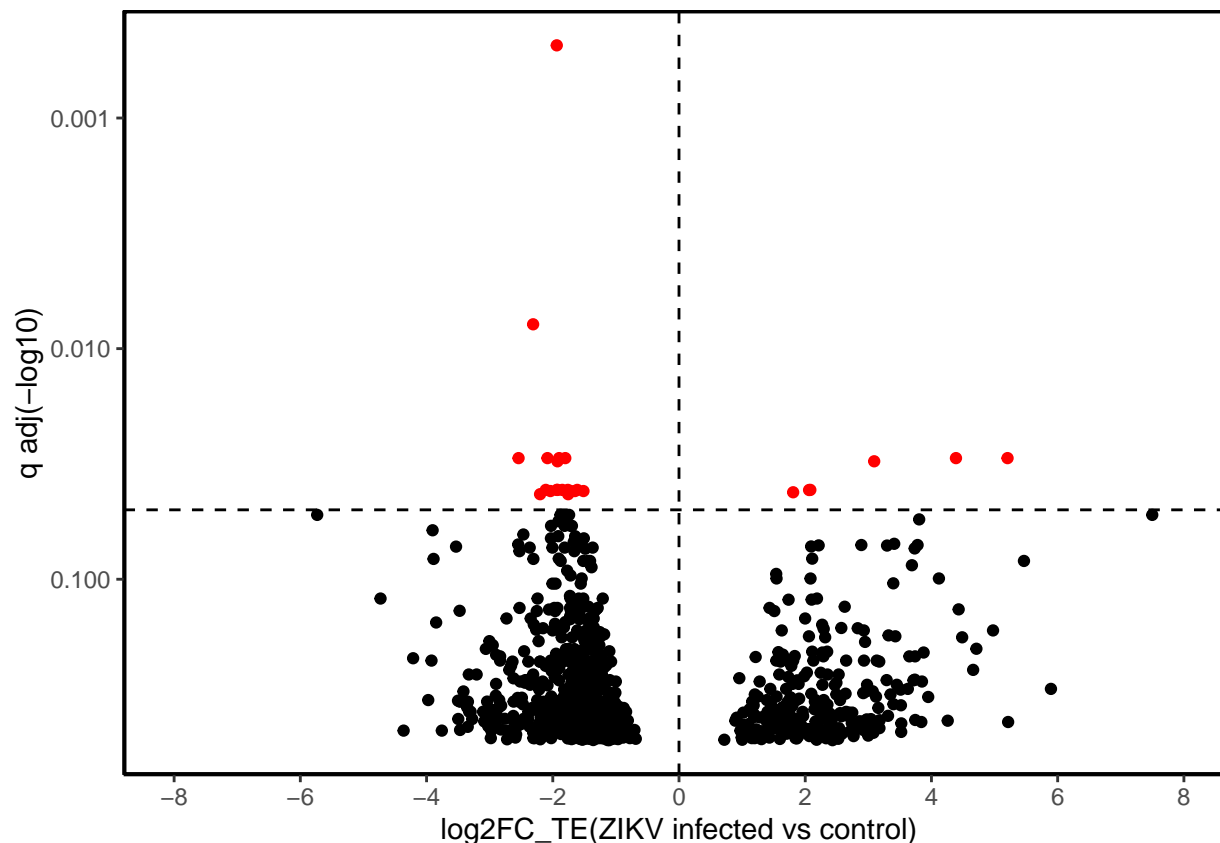
```
# Figure 3 B

colnames(df.DENV)[7] <- "log2FC_TE_DrugTreated_vs_Control"

df.DENV.sub <- subset(df.DENV, !is.na(padj))
df.DENV.sub <- df.DENV.sub[df.DENV.sub$padj < 0.5,]

fig3b <- ggplot(df.DENV.sub,
                aes(x = log2FC_TE_DrugTreated_vs_Control, y = padj))

fig3b +
  geom_point(aes(color = padj > 0.05), show.legend = FALSE) +
  scale_y_continuous(trans = reverselog_trans(base = 10),
                     labels = trans_format("log10", math_format(10^.x)),
                     breaks = trans_breaks("log10", function(x) 10^x)) +
  scale_x_continuous(limits = c(-8, 8),
                     breaks = seq(-8, 8, 2)) +
  scale_color_manual(values = c('red', 'black')) +
  xlab("log2FC_TE(DENV infected vs control)") +
  ylab("q adj(-log10)") +
  geom_hline(yintercept=0.05, linetype=2, col="black") +
  geom_vline(xintercept=0, linetype=2, col="black") +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
```

```
        axis.line = element_line(colour = "black"),
        panel.border = element_rect(colour = "black",
                                    fill = NA, linewidth = 1))
```