

# Introduction to Bulk RNAseq data analysis

## Visualisation of Differential Expression Results - Solutions

Last modified: 01 Oct 2024

## Contents

Exercise 1 - Volcano plot for 33 days . . . . .	1
Exercise 2 - MA plot for day 33 with ggplot2 . . . . .	2
Exercise 3 - Strip Chart . . . . .	2

## Exercise 1 - Volcano plot for 33 days

Now it's your turn! We just made the volcano plot for the 11 days contrast, you will make the one for the 33 days contrast.

If you haven't already make sure you load in our data and annotation. You can copy and paste the code below.

```
# First load data and annotations
results.d33 <- readRDS("RObjects/DESeqResults.interaction_d33.rds")
annot <- readRDS("RObjects/Ensembl_annotations.rds")
```

(a) Shrink the results for the 33 days contrast.

```
#Shrink our values
shrink.33 <- lfcShrink(ddsObj,
                      res = results.d33,
                      type = "ashr")

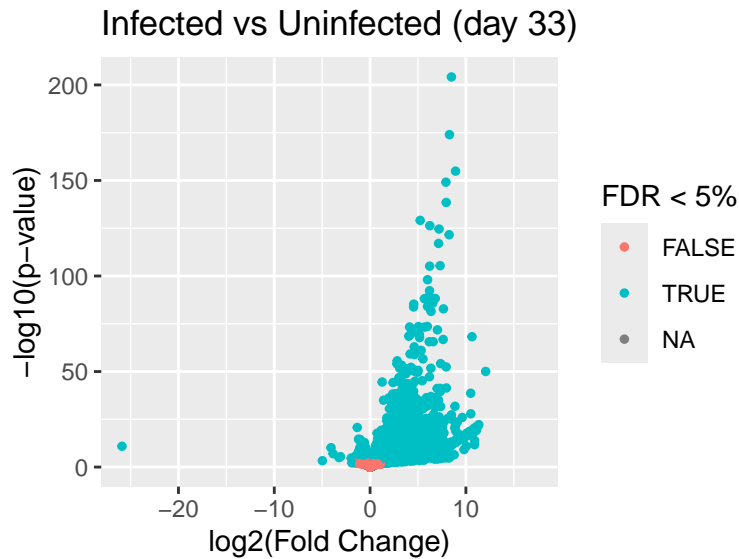
## using 'ashr' for LFC shrinkage. If used in published research, please cite:
## Stephens, M. (2016) False discovery rates: a new deal. Biostatistics, 18:2.
## https://doi.org/10.1093/biostatistics/kxw041

shrinkTab.33 <- as.data.frame(shrink.33) %>%
  rownames_to_column("GeneID") %>%
  left_join(annot, "GeneID")
```

(b) Create a plot with points coloured by P-value < 0.05 similar to how we did in the first volcano plot

```
ggplot(shrinkTab.33, aes(x = log2FoldChange, y = -log10(pvalue))) +
  geom_point(aes(colour = padj < 0.05), size = 1) +
  labs(x = "log2(Fold Change)", y = "-log10(p-value)", colour = "FDR < 5%",
       title = "Infected vs Uninfected (day 33)")
```

```
## Warning: Removed 47 rows containing missing values or values outside the scale
## range (`geom_point()`).
```

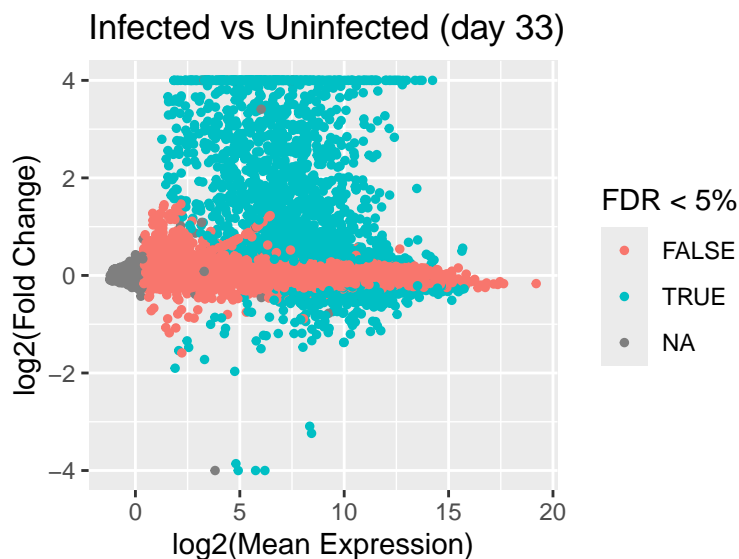


## Exercise 2 - MA plot for day 33 with ggplot2

For this exercise create an MA plot for day 33 like the ones we plotted with `plotMA` from **DESeq2** but this time using `ggplot2`.

The x-axis (M) should be the  $\log_2$  of the mean gene expression across all samples, and the y-axis should be the  $\log_2$  of the fold change between Infected and Uninfected.

```
ggplot(shrinkTab.33, aes(x = log2(baseMean), y = log2FoldChange)) +
  geom_point(aes(colour = padj < 0.05), size = 1) +
  scale_y_continuous(limit = c(-4, 4), oob = scales::squish) +
  labs(x = "log2(Mean Expression)",
       y = "log2(Fold Change)",
       colour = "FDR < 5%",
       title = "Infected vs Uninfected (day 33)")
```



## Exercise 3 - Strip Chart

For this exercise create another strip chart for the gene *Jchain*.

```

geneID_33 <- shrinkTab.33 %>%
  filter(Symbol == "Jchain") %>%
  pull(GeneID)

plotCounts(ddsObj,
  gene = geneID_33,
  intgroup = c("TimePoint", "Status", "Replicate"),
  returnData = TRUE) %>%
  ggplot(aes(x = Status, y = log2(count))) +
  geom_point(aes(fill = Replicate), shape = 21, size = 2) +
  facet_wrap(~ TimePoint) +
  expand_limits(y = 0) +
  labs(title = "Normalised counts - Immunoglobulin Joining Chain")

```

