

Let's practice – extra – 1/2

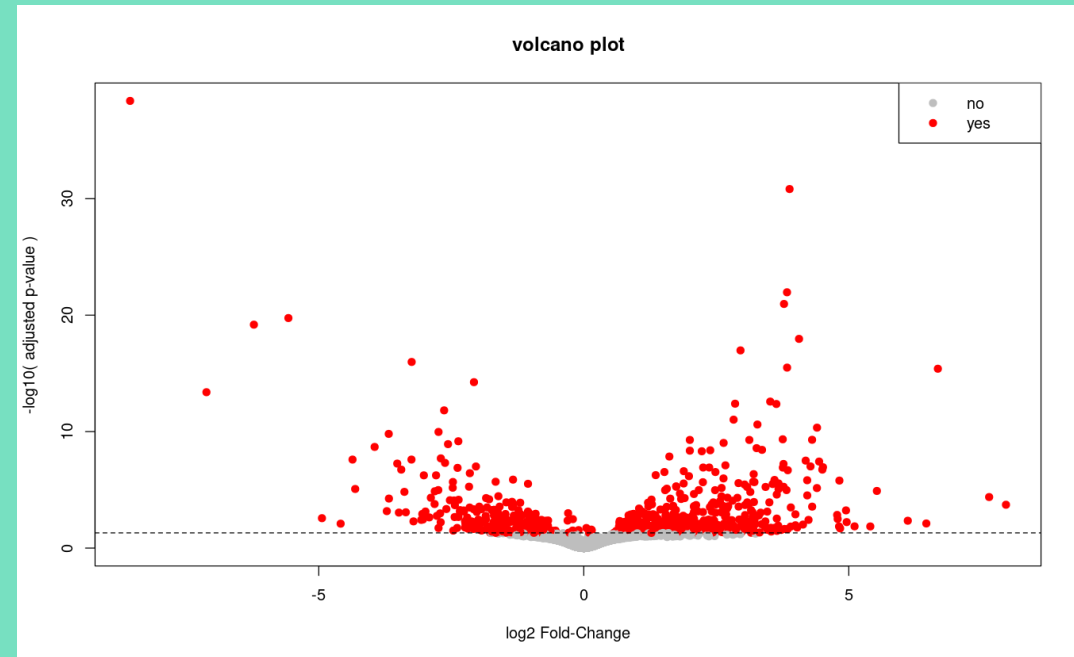
In this extra practice session we are going to read in a csv file that contain the result of a Differential Gene Expression analysis from RNAseq data.

Briefly, this dataset consists in a table where each line is a gene and the different columns contains information about the change in expression level between two experimental conditions.

The most important columns here are:

- **padj** : contains the adjusted p-value (a low p-value denotes a gene whose expression is significantly different between the conditions)
- **log2FoldChange** : logarithm2 of the fold-change between the two conditions (eg, a value of 1 mean that the expressed twice a much ; -1 means that the gene is expressed half as much)

Our goal is to quickly analyse this data and ultimately generate a [volcano plot](#), which should look somewhat like this:



The Instructions are in the next slide.

Reminder: You can find a solution script in the solutions/ folder

Let's practice – extra – 2/2

1. read the csv file `course_datasets/Ruhland2016.DESeq2.results.csv` as a `data.frame`, using the first row as row names.
2. Create a new column named "significant" which contains the text "yes" if the adjusted p-value (column `padj`) is below the significance threshold 0.05, and "no" otherwise.
3. how many genes have a significant adjusted p-value?
4. What is the average of the absolute value of log-fold change? (nb: you may have to google the function to compute the absolute value)
 1. For all genes?
 2. For significant genes only?
5. creating a volcano plot
 1. Create a new column containing $-\log_{10}(\text{adjusted p-value})$ (it will be used for the y-axis)
 2. basic plot: plot the log-fold change (x-axis) with $-\log_{10}(\text{adjusted p-value})$ (y-axis)
 3. adding colors: adapt the previous plot to change the `pch` to 19, and color in grey non-significant genes and in red significant ones (you can use the "significant" column we created earlier)
 4. adding threshold line: add an horizontal dashed black line to visualize the 0.05 threshold on adjusted p-value
 5. beautify: add a title, axis labels, and legend