

# MetaStudy

Moritz Madern

2023-01-14

This script compiles differential expression data from files of different experiments (each provided in a standardized format & containing gene-wise FC vs p-values) and performs meta-analysis on specified genes.

## Specify parameters

```
## specify file path to datasets to read in (all datasets should be located in the same folder)
filepath_datasets = "./Datasets"

## specify file path to table that defines the conditions within those datasets of interest
filepath_tableConditions = "Condition_datasets_11012023.csv"

## specify genes to be analyzed (note that they will be converted to capital letters to make pattern s
vector_geneNames = c("Cdc45", "eIF2A", "Rpi7", "Gapdh", "Ncor1", "Stat1", "Gata3", "Stat4")

## specify subset of those genes (from vector_geneNames) to be illustrated in a Heatmap.
vector_geneNames_heatmap = c("Cdc45", "Ncor1", "Stat1", "Stat4", "Gata3")

## specify heatmap color scheme. Supported colorschemes are "viridis" and "redblue"
heatmap_colorscheme = "redblue"
```

## Read in datasets

Read in conditions table:

```
##  dataset condition.number
## 1      1      condition1
## 2      2      condition3
## 3      3      condition1
## 4      4      condition2
## 5      5      condition1
## 6      7      condition2
## 7      8      condition2
## 8      9      condition3
## 9     11      condition1
```

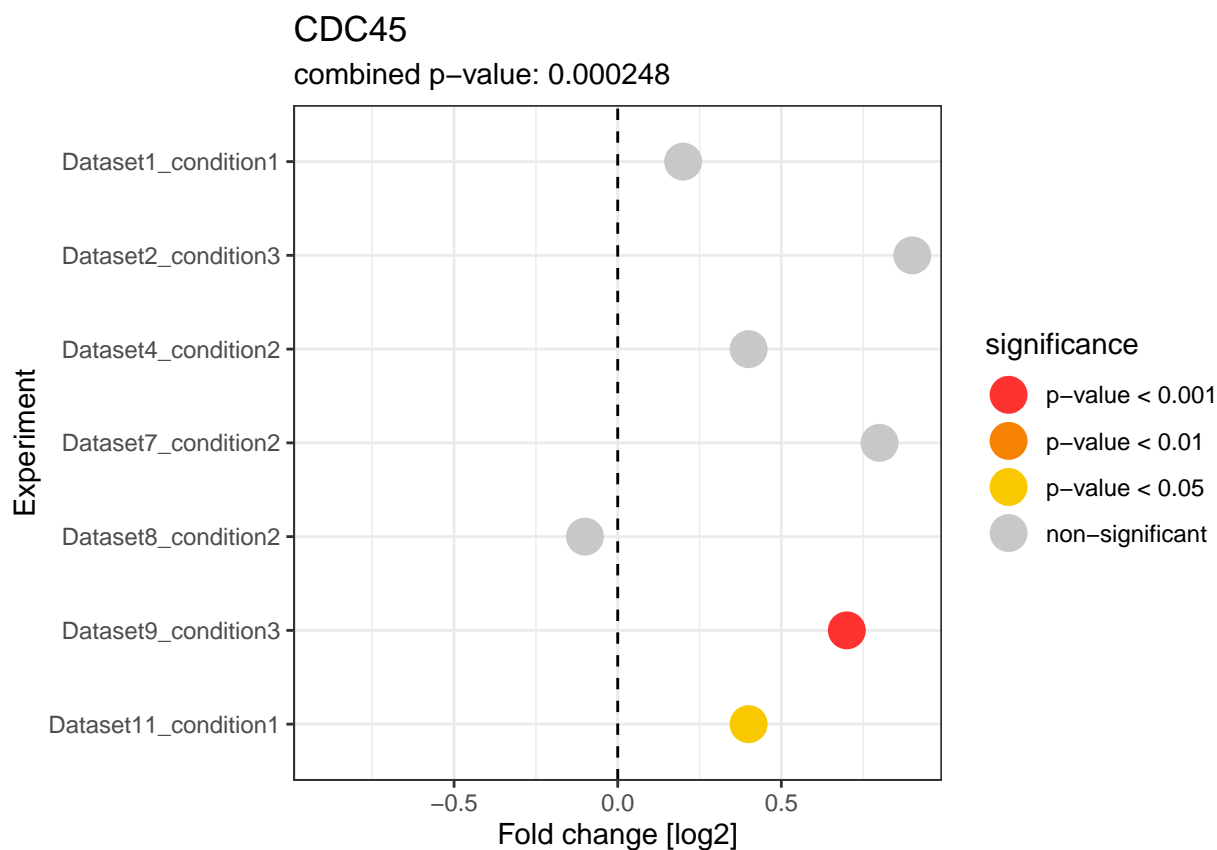
Read in all datasets as specified by filepath\_datasets:

```
## [1] "./Datasets/Dataset1 E-GEOD-57945-query-results-.tsv"
## [1] "./Datasets/Dataset11 E-GEOD-42768-A-AFFY-130-query-results.tsv"
## [1] "./Datasets/Dataset2 E-MTAB-7860-query-results-2.tsv"
## [1] "./Datasets/Dataset3 E-GEOD-2461-A-AFFY-33-query-results.tsv"
## [1] "./Datasets/Dataset4 E-GEOD-65114-A-GEOD-16686-query-results.tsv"
## [1] "./Datasets/Dataset5 E-GEOD-6731-A-AFFY-1-query-results.tsv"
## [1] "./Datasets/Dataset7 E-MTAB-9850-query-results-2.tsv"
## [1] "./Datasets/Dataset8 E-GEOD-20621-A-AFFY-45-query-results-2.tsv"
## [1] "./Datasets/Dataset9 E-GEOD-27302-A-AFFY-45-query-results.tsv"
```

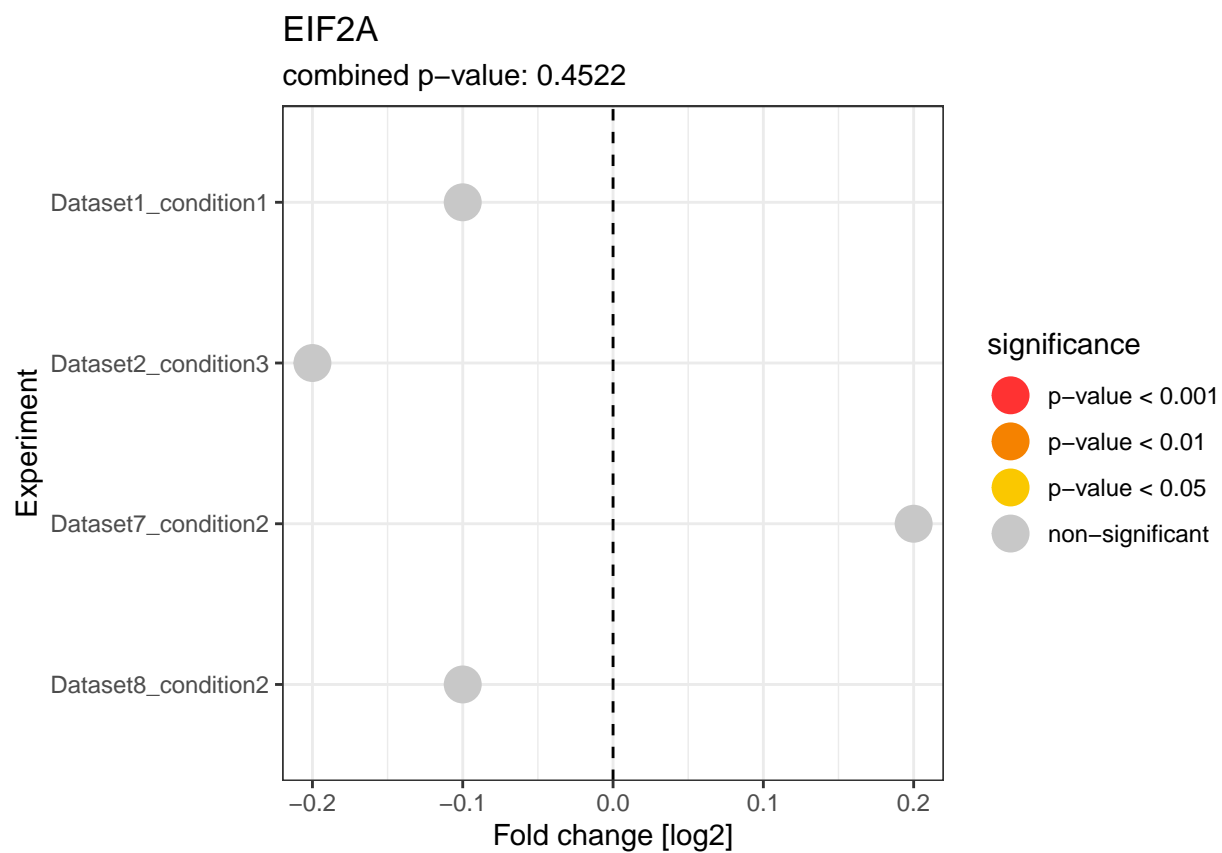
## Plot data for each specified gene

Plots show foldChange vs experiment, points are colored by statistical significance. Note that experiments in which the respective gene was not quantified are not shown. Furthermore, for each gene, the individual p-values were combined to an overall p-value via Fisher's method ([https://en.wikipedia.org/wiki/Fisher%27s\\_method](https://en.wikipedia.org/wiki/Fisher%27s_method)), rounded to 6 digits. This method assumes that the individual p-values to be pooled are independent but test for the same Null hypothesis. If these assumptions are not met, ignore the combined p-value.

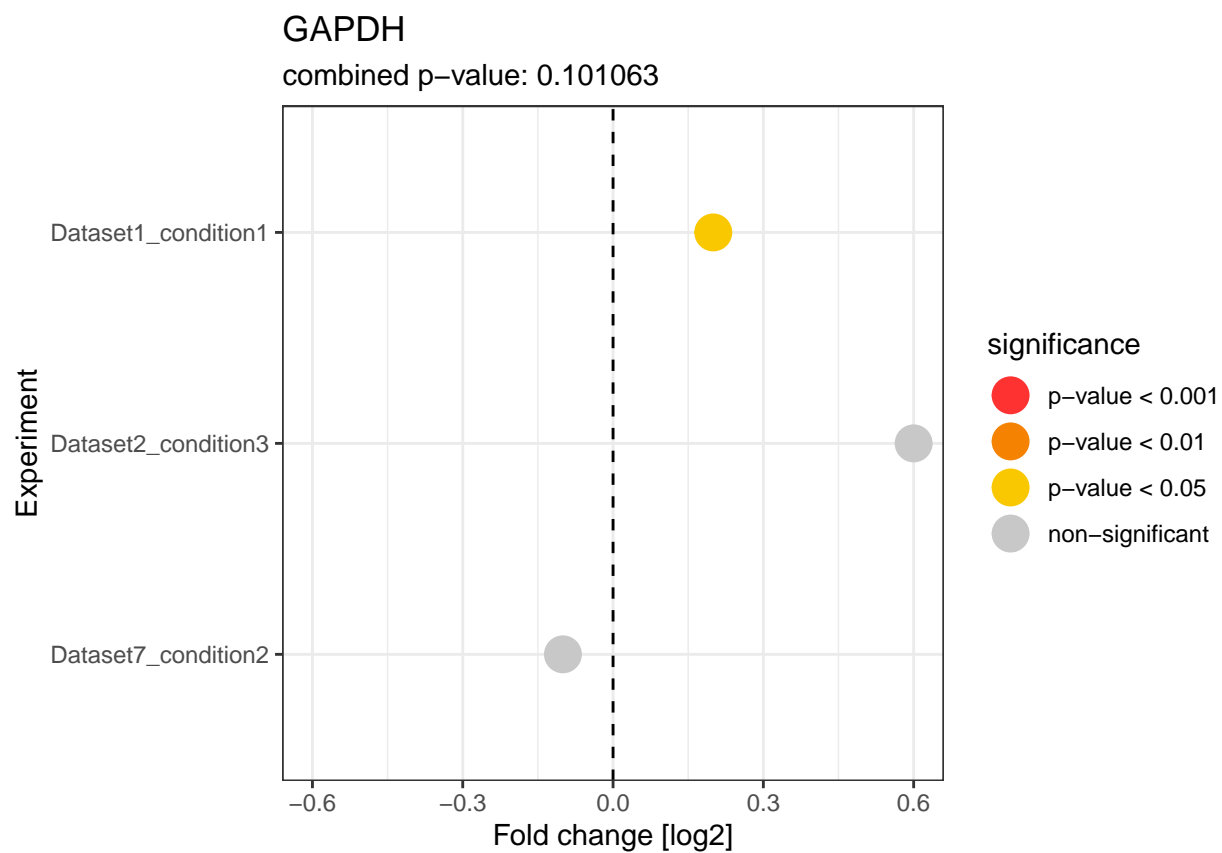
```
## [[1]]
```



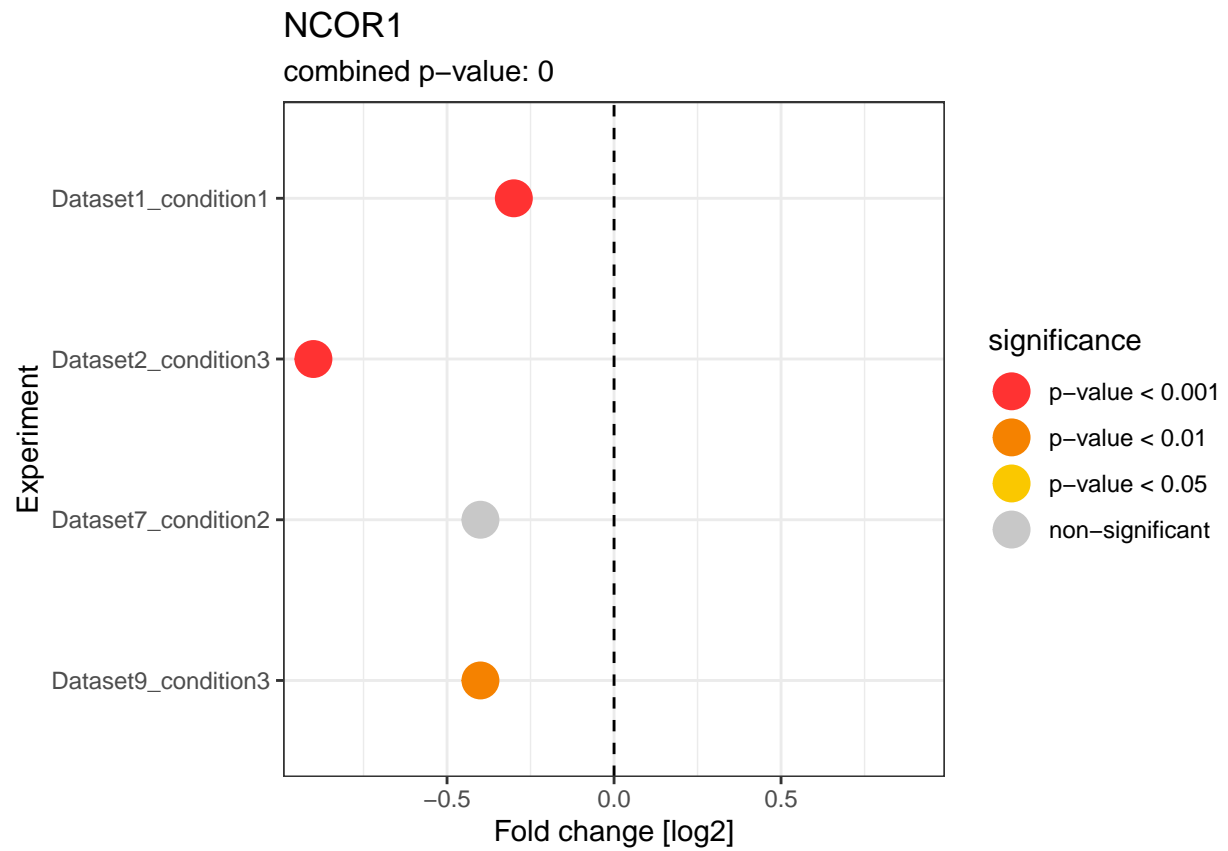
```
##
## [[2]]
```



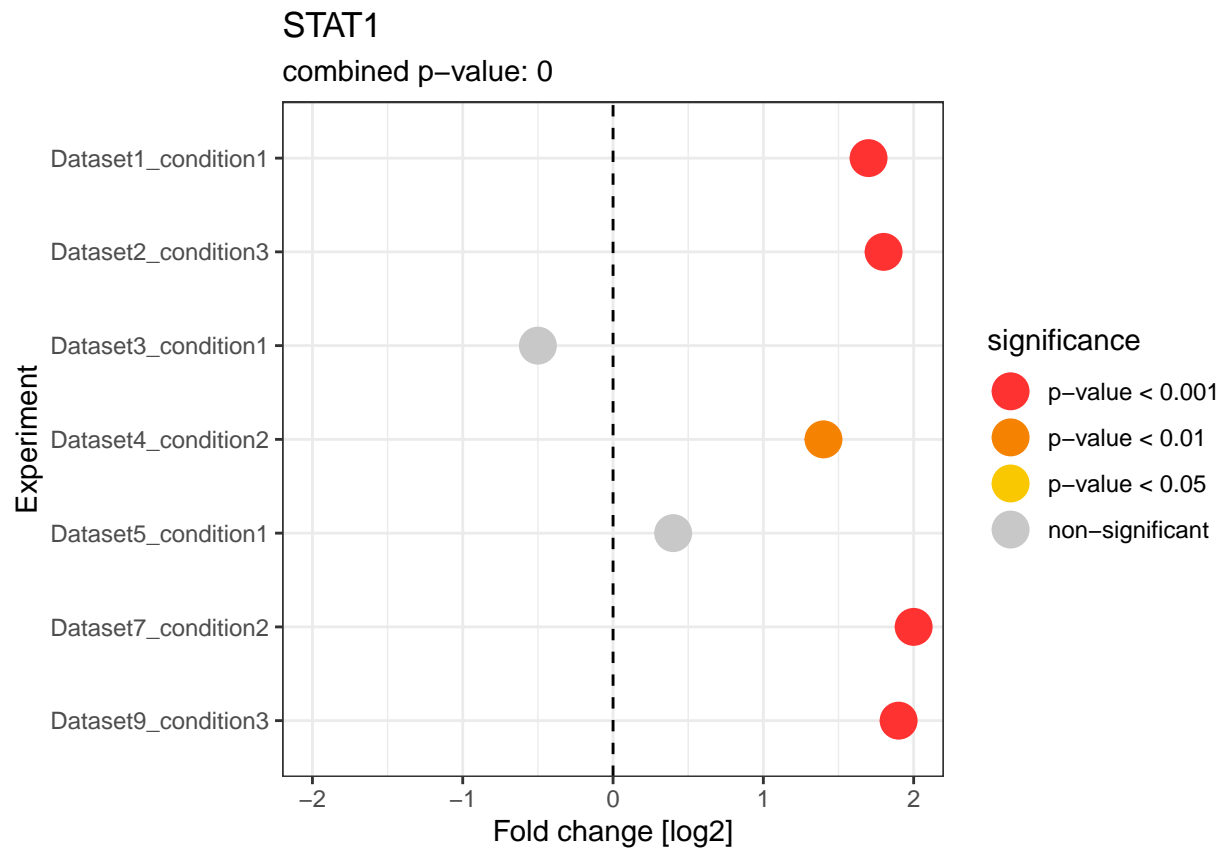
```
##  
## [[3]]
```



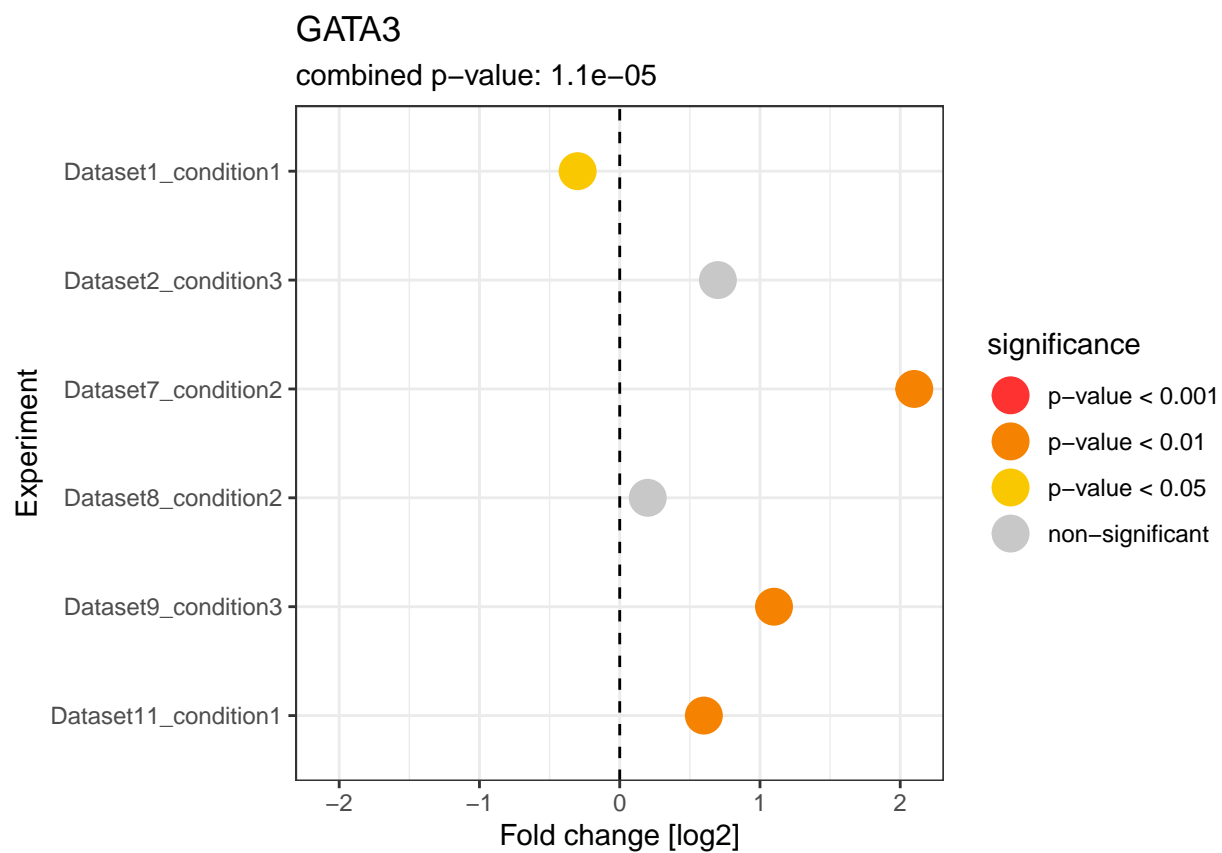
```
##  
## [[4]]
```



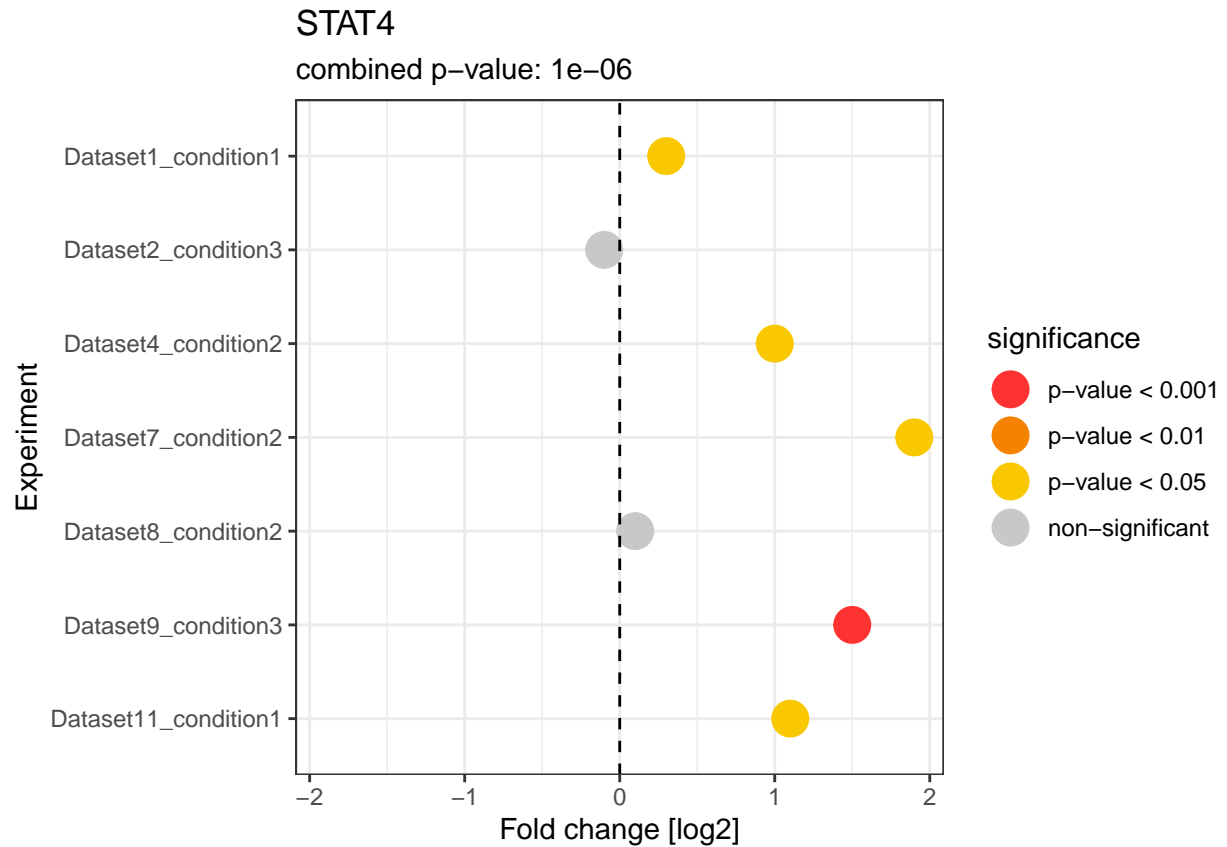
```
##  
## [[5]]
```



```
##  
## [[6]]
```



```
##
## [[7]]
```



Note: Upon compiling the script to pdf, each plot is saved as separate pdf and png file in the newly generated folder called “figures” in your working directory.

## Plot Heatmap

For selected genes, a heatmap is plotted that depicts fold changes. Missing values (i.e. NAs) are colored as grey.



