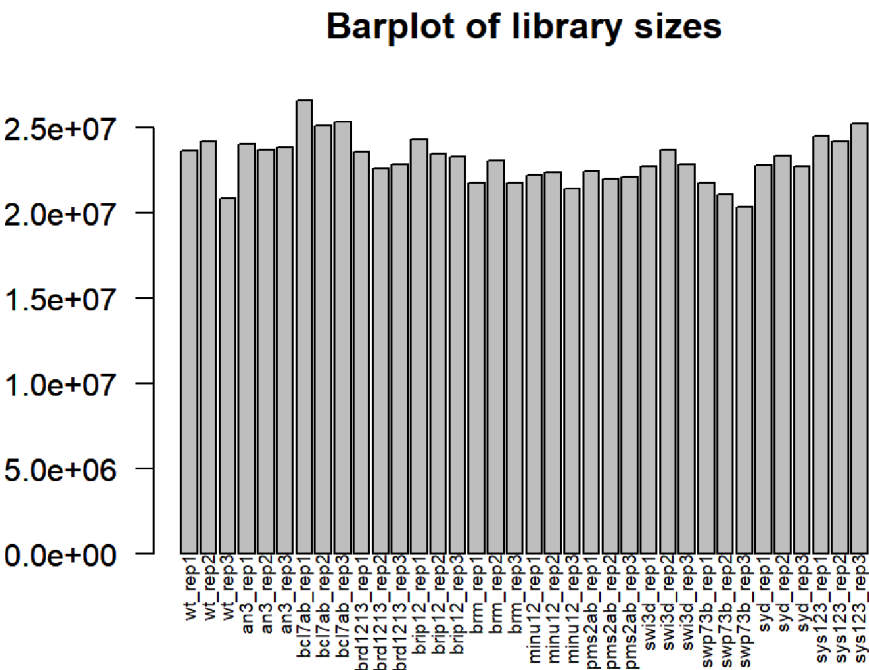


3.1 MDS plot

library size



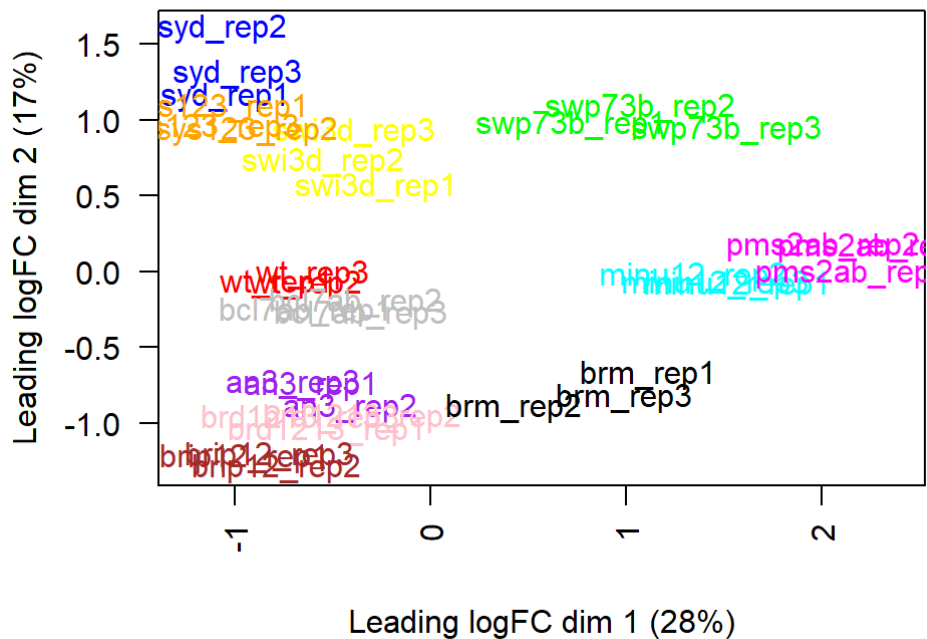
screen: I only preserve genes which's cpm > 0.5. ([As a general rule, a good threshold can be chosen for a CPM value that corresponds to a count of 10](#), and 0.5 equals to 10 in these data, which is showed by my plots.)

before screen: 32548 36

after screen: 20923 36

After screen, the MDS plot:

MDS plot



The MDS plot tells us

- the samples within a group are near each other and the samples of different groups are far from each other, this tells us that the data is consistent with our prior knowledge that some samples are from the same cluster.

3.2 different expressed genes

use FDR to find different expressed genes, use first 2 comparisons as examples.

```
> topTags(et_12_1, n=10)
Comparison of groups: an3-wt
      logFC      logCPM      PValue      FDR
AT1G64380 -4.407562  3.2211569  4.230813e-75  8.852130e-71
AT4G30350 -1.648675  5.1434481  4.031713e-70  4.217777e-66
AT3G05955  6.181067  0.9144233  1.533801e-65  1.069724e-61
AT1G64795 -7.874744  0.9953956  2.883384e-53  1.508226e-49
AT1G78070 -3.853441  6.4030389  2.995819e-46  1.253630e-42
AT2G47060 -1.509678  5.8883224  2.139378e-37  7.460369e-34
AT1G01720 -2.284484  5.5775999  4.675619e-37  1.397543e-33
AT3G11410 -3.024735  6.1405742  1.717610e-34  4.492195e-31
AT3G11420 -1.385513  5.3979381  2.197468e-33  5.108626e-30
AT5G59220 -5.599080  4.5423172  4.485019e-33  9.384005e-30
```

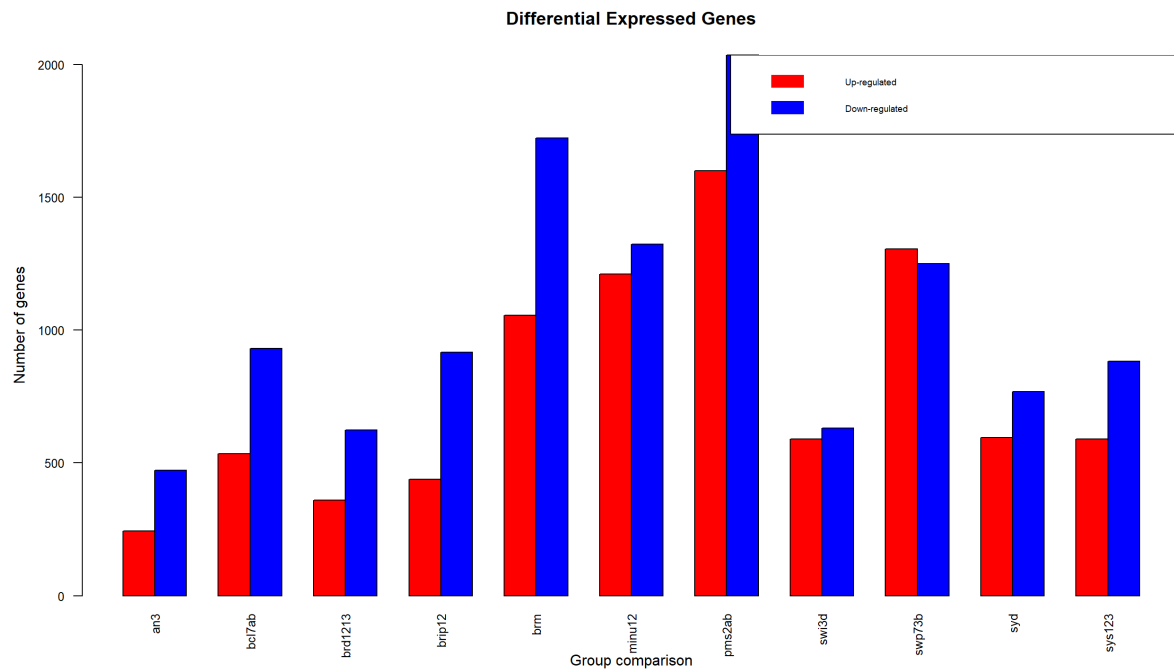
```
> topTags(et_12_2, n=10)
Comparison of groups: bcl7ab-wt
      logFC      logCPM      PValue      FDR
AT5G55210 -6.716477  3.6362928  2.925769e-180  6.121587e-176
AT1G52100  2.977161  4.7005055  2.571290e-100  2.689955e-96
AT5G19310  1.776345  4.7941702  9.535594e-84  6.650441e-80
AT4G22320 -4.896863  2.5592672  4.895474e-64  2.560700e-60
AT2G47060 -1.925201  5.8883224  4.634788e-58  1.939473e-54
AT1G64380 -3.593844  3.2211569  6.846575e-57  2.387515e-53
AT4G30350 -1.430295  5.1434481  3.256452e-54  9.733536e-51
AT3G05955  5.727379  0.9144233  1.185986e-51  3.101798e-48
AT1G64795 -5.402176  0.9953956  6.272573e-45  1.458234e-41
AT4G33490 -1.628355  4.9115621  2.784600e-44  5.826218e-41
```

According to the paper, the criterion of up

- $\log_2(FC) > 1$
- $FDR < 0.05$

the criterion of down

- $\log_2(FC) < -1$
- $FDR < 0.05$



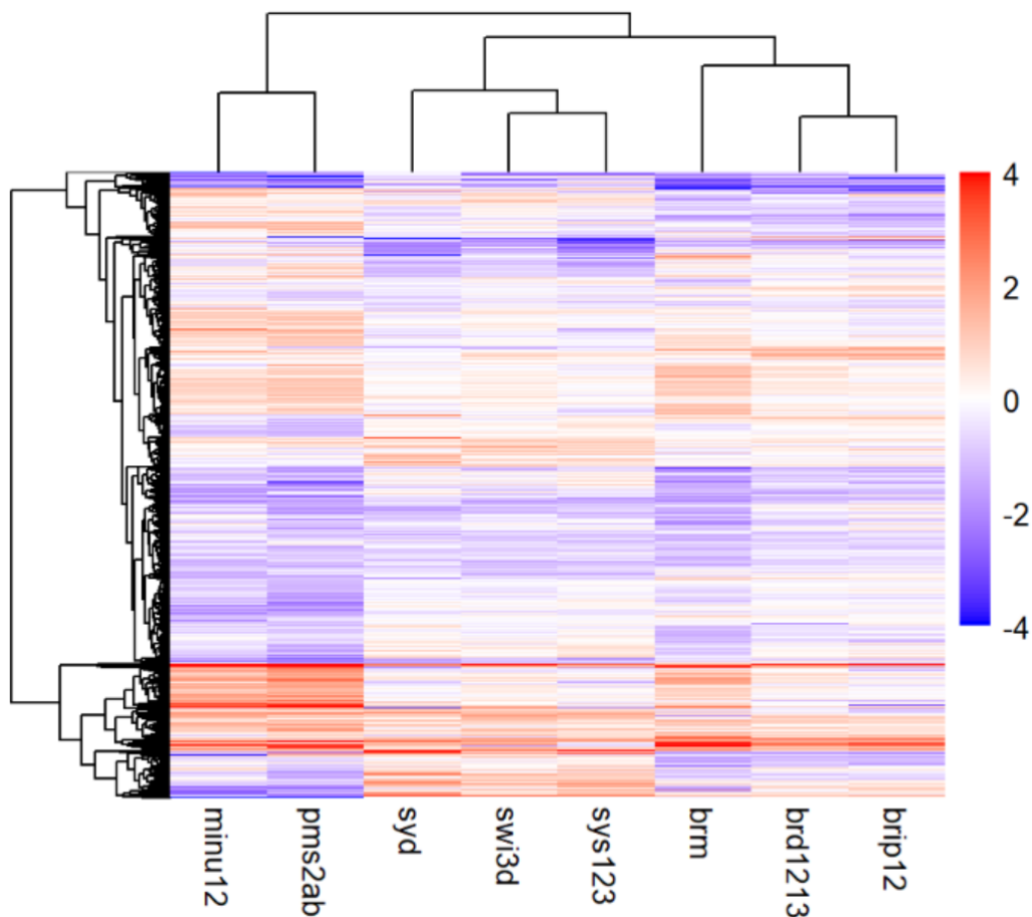
The graph of mine is very similar to figure 3.2.

The bar plot tells us:

- down-regulated genes are more than up-regulated genes.

3.3 heat map of $\log_2 FC$

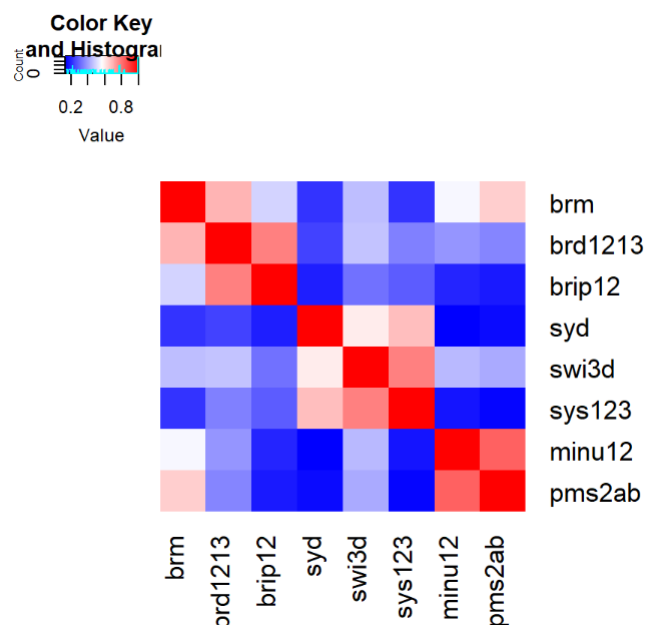
Do Or operation to 11 groups of 3.2 graph, the number of gene is 7290, which is similar to 7753 in the original paper.



The heat map tells us:

- down-regulated genes are more than up-regulated genes.
- the cluster of columns tell use that the data is consistent with our prior knowledge that some samples are from the same cluster.

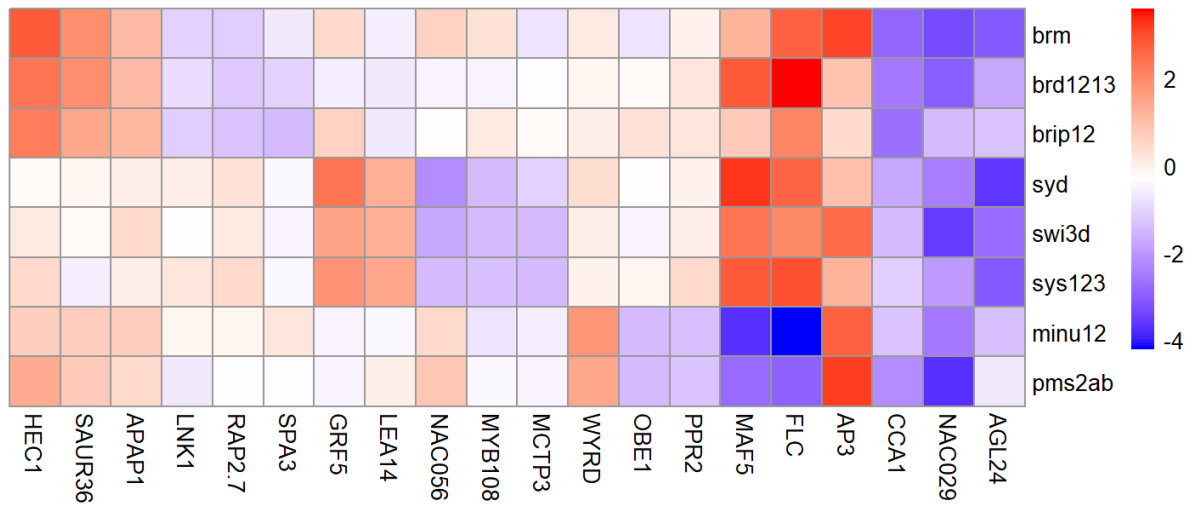
3.4 heat map of correlation matrix



The heat map of correlation matrix tells us:

- the cluster of columns (3,3,2) tell use that the data is consistent with our prior knowledge that some samples are from the same cluster.

3.5 heat map of some genes' log2 FC



Original paper use 26 genes, but I only have 20 of them after screen.

Reference

[tutorial 1: I use it to draw the MDS plot](#)

[tutorial 2: I use it to calculate FDR to find DEGs](#)

[a tutorial using NBA stats to help me learn how to use pheatmap](#)

[a YouTube video for pheatmap](#)

[Official user guide of edgeR](#)