lab1.r

bauricenafack

2022-10-27

```
setwd("~/Documents/computational-genomics/kap1-experiment")
library(edgeR)
```

```
## Loading required package: limma
```

```
counts<-read.csv("data/GSE105128_genecount.csv")
head(counts)
```

```
##             geneName  WT1  WT2  WT3  KO1  KO2  KO3
## 1 ENSMUSG00000000001 2022 2141 2313 2311 2415 2748
## 2 ENSMUSG00000000003    0    0    0    0    0    0
## 3 ENSMUSG00000000028  157  128   97   98  112  145
## 4 ENSMUSG00000000031    0    0    0    0    0    0
## 5 ENSMUSG00000000037    5    5    8    7   10    8
## 6 ENSMUSG00000000049    0    0    0    0    0    2
```

```
dim(counts)
```

```
## [1] 37991     7
```

```
"WT: normal cell, KO: knockout cell. The counts has 7 columns, the first is the gene name.
It contains the id that identifies the different genes, every raw is related to a single gene in the mou
The column WT1 represent the wise type relative to the expression level of every normal gene in the mic
replication 1, the second WT2 contains the expression level of the normal T regulatory cell  and same fo
KO1 contains expression level of the T-reg cell where KAP1 is shut down and cannot produce the gene.
Same interpretation for replicate 2 and 3. "
```

```
## [1] "WT: normal cell, KO: knockout cell. The counts has 7 columns, the first is the gene name. \nIt c
```

```
"This experiment/data aim to identify the genes that behave differently. We check whether they have dif
average levels in Ko t-reg cells compared to With type cells. In this experiment, we will be able to de
which portions of a gene in the mouse genome in the T-reg cell are probably regulated by kap1. It is imp
because we need to know all the different relationships between cells, especially for a cell associated
disease, to identify the type of gene responsible for a specific disease. Kap1 off implies the autoimmun
```

```
## [1] "This experiment/data aim to identify the genes that behave differently. We check whether they ha
```

```
# Let's explore the data
"Quality check using edgeR"
```

```
## [1] "Quality check using edgeR"
```

```
"1) Load data in edgeR
In this data, rows are gene and columns are relative to samples"
```

```
## [1] "1) Load data in edgeR\nIn this data, rows are gene and columns are relative to samples"
```

```
data<-DGEList(counts=counts[,2:7], genes = counts[,1]) # counts receive the matrix of sample

class(data)
```

```
## [1] "DGEList"
## attr(,"package")
## [1] "edgeR"
```

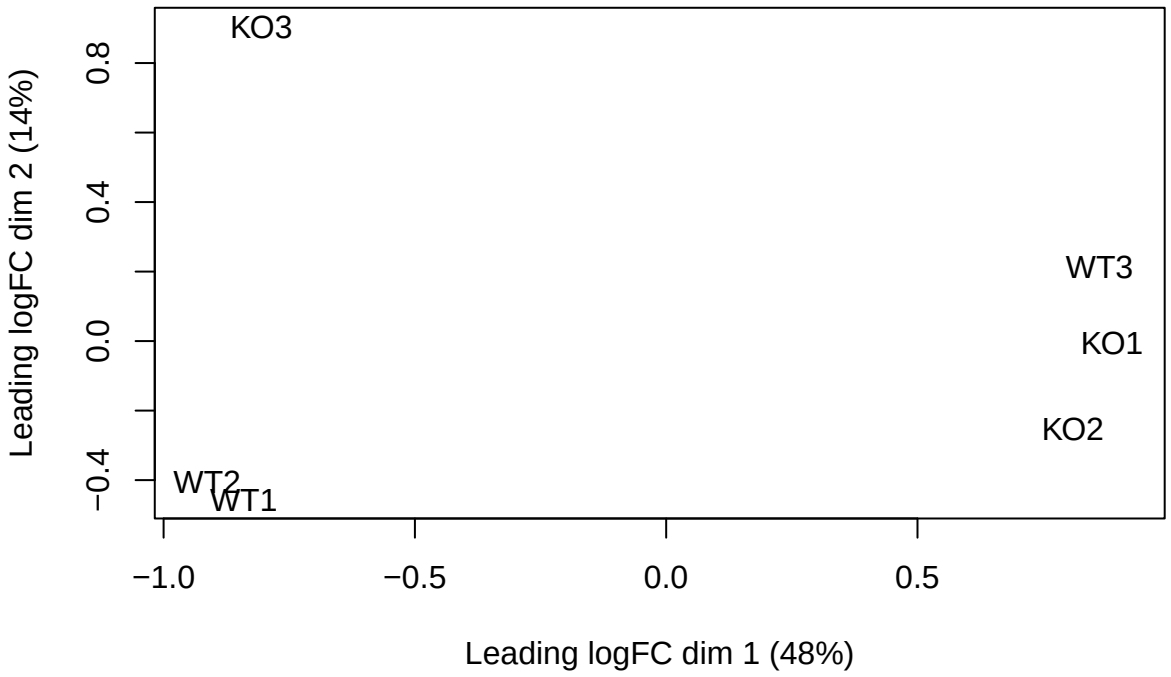
```
data
```

```
## An object of class "DGEList"
## $counts
##    WT1  WT2  WT3  KO1  KO2  KO3
## 1 2022 2141 2313 2311 2415 2748
## 2    0    0    0    0    0    0
## 3  157  128   97   98  112  145
## 4    0    0    0    0    0    0
## 5    5    5    8    7   10    8
## 37986 more rows ...
##
## $samples
##     group lib.size norm.factors
## WT1     1 10103117            1
## WT2     1 10180281            1
## WT3     1 10300626            1
## KO1     1 10080881            1
## KO2     1  9954344            1
## KO3     1 10994745            1
##
## $genes
##                genes
## 1 ENSMUSG00000000001
## 2 ENSMUSG00000000003
## 3 ENSMUSG00000000028
## 4 ENSMUSG00000000031
## 5 ENSMUSG00000000037
## 37986 more rows ...
```

```
"a) Let's apply the PCA"
```

```
## [1] "a) Let's apply the PCA"
```

```
plotMDS(data)
```




```
"Interpretation:
We see in the x axis that there are two groups, but this is not we expected. This is telling us that th
We just need to say there are something wrong as assumption. "
```

```
## [1] "Interpretation:\nWe see in the x axis that there are two groups, but this is not we expected. T
```