Project

Oksana Ivanova

```
stringsAsFactors = F

library(rapportools)
library(data.table)
library(DESeq2)
library(limma)
library(fgsea)
library(BiocParallel)
library(WGCNA)
library(ggplot2)
library(ggrepel)
library(reshape)
library(Biobase)
library("AnnotationDbi")
library("org.Hs.eg.db")
library(devtools)
library(rUtils)
```

es raw

```
#Download sample_tables
samples1 <- fread("./featureCounts_counts.txt")
samples2 <- fread("./3RNASeq_21.10.2018.cnt")
head(samples1)
```

```
##               Geneid
## 1: ENSG00000223972.5
## 2: ENSG00000227232.5
## 3: ENSG00000278267.1
## 4: ENSG00000243485.5
## 5: ENSG00000284332.1
## 6: ENSG00000237613.2
##                                                              Chr
## 1:           chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1
## 2: chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1;chr1
## 3:                                                        chr1
## 4:                                chr1;chr1;chr1;chr1;chr1
## 5:                                                        chr1
## 6:                                chr1;chr1;chr1;chr1;chr1
##                                                                  Start
## 1:             11869;12010;12179;12613;12613;12975;13221;13221;13453
## 2: 14404;15005;15796;16607;16858;17233;17606;17915;18268;24738;29534
## 3:                                                             17369
## 4:                                     29554;30267;30564;30976;30976
## 5:                                                             30366
## 6:                                     34554;35245;35277;35721;35721
##                                                                    End
```

```
## 1:             12227;12057;12227;12721;12697;13052;13374;14409;13670
## 2: 14501;15038;15947;16765;17055;17368;17742;18061;18366;24891;29570
## 3:                                                                17436
## 4:                                     30039;30667;30667;31109;31097
## 5:                                                                30503
## 6:                                     35174;35481;35481;36073;36081
##                   Strand Length 3RNA-2-1_S11_L001_R1_001.fastq.gz.bam
## 1:     +;+;+;+;+;+;+;+;+   1735                                       0
## 2: -;-;-;-;-;-;-;-;-;-;-   1351                                       0
## 3:                       -     68                                       0
## 4:              +;+;+;+;+   1021                                       0
## 5:                      +    138                                       0
## 6:              -;-;-;-;-   1219                                       0
##    3RNA-7_S1_L001_R1_001.fastq.gz.bam
## 1:                                  0
## 2:                                  0
## 3:                                  0
## 4:                                  0
## 5:                                  0
## 6:                                  0
##    3RNA-9-1_S12_L001_R1_001.fastq.gz.bam
## 1:                                    0
## 2:                                    0
## 3:                                    0
## 4:                                    0
## 5:                                    0
## 6:                                    0
##    3RNA-11_S2_L001_R1_001.fastq.gz.bam 3RNA-13_S9_L001_R1_001.fastq.gz.bam
## 1:                                   0                                   0
## 2:                                   0                                   0
## 3:                                   0                                   0
## 4:                                   0                                   0
## 5:                                   0                                   0
## 6:                                   0                                   0
##    3RNA-14_S10_L001_R1_001.fastq.gz.bam
## 1:                                    0
## 2:                                    0
## 3:                                    0
## 4:                                    0
## 5:                                    0
## 6:                                    0
##    3RNA-19_S5_L001_R1_001.fastq.gz.bam 3RNA-22_S7_L001_R1_001.fastq.gz.bam
## 1:                                   0                                   0
## 2:                                   0                                   0
## 3:                                   0                                   0
## 4:                                   0                                   0
## 5:                                   0                                   0
## 6:                                   0                                   0
##    3RNA-HF1-1_S3_L001_R1_001.fastq.gz.bam
## 1:                                      0
## 2:                                      0
## 3:                                      0
## 4:                                      0
## 5:                                      0
```

```
## 6:                                      0
##    3RNA-HF1-2_S4_L001_R1_001.fastq.gz.bam
## 1:                                      0
## 2:                                      0
## 3:                                      0
## 4:                                      0
## 5:                                      0
## 6:                                      0
##    3RNA-HF12_S6_L001_R1_001.fastq.gz.bam
## 1:                                     0
## 2:                                     0
## 3:                                     0
## 4:                                     0
## 5:                                     0
## 6:                                     0
##    3RNA-Mal_S8_L001_R1_001.fastq.gz.bam
## 1:                                    0
## 2:                                    0
## 3:                                    0
## 4:                                    0
## 5:                                    0
## 6:                                    0
```

```
head(samples2)
```

```
##                   V1 3RNA-11_S2_L001_R1_001.fastq.gz
## 1: ENSG00000000003.14                               1
## 2:  ENSG00000000005.5                               1
## 3: ENSG00000000419.12                               0
## 4: ENSG00000000457.13                               3
## 5: ENSG00000000460.16                               0
## 6: ENSG00000000938.12                               2
##    3RNA-13_S9_L001_R1_001.fastq.gz 3RNA-14_S10_L001_R1_001.fastq.gz
## 1:                               1                                15
## 2:                               0                                 0
## 3:                               0                                 2
## 4:                               2                                 6
## 5:                               0                                 0
## 6:                               2                                 9
##    3RNA-19_S5_L001_R1_001.fastq.gz 3RNA-2-1_S11_L001_R1_001.fastq.gz
## 1:                               2                                  8
## 2:                               3                                  1
## 3:                               2                                  8
## 4:                               2                                  3
## 5:                               0                                  1
## 6:                               5                                  1
##    3RNA-22_S7_L001_R1_001.fastq.gz 3RNA-7_S1_L001_R1_001.fastq.gz
## 1:                               7                              2
## 2:                               0                              3
## 3:                               6                              1
## 4:                               2                              0
## 5:                               1                              0
## 6:                               3                              1
##    3RNA-9-1_S12_L001_R1_001.fastq.gz 3RNA-HF1-1_S3_L001_R1_001.fastq.gz
## 1:                                 2                                  1
```

```
## 2:                                0                                2
## 3:                                4                                2
## 4:                                0                                1
## 5:                                0                                0
## 6:                                1                                1
##    3RNA-HF1-2_S4_L001_R1_001.fastq.gz 3RNA-HF12_S6_L001_R1_001.fastq.gz
## 1:                                  2                                2
## 2:                                  1                                3
## 3:                                  3                                3
## 4:                                  1                                0
## 5:                                  0                                0
## 6:                                  1                                0
##    3RNA-Mal_S8_L001_R1_001.fastq.gz
## 1:                                6
## 2:                                8
## 3:                                0
## 4:                                2
## 5:                                0
## 6:                                1
```

```
samples1$Geneid <- sub("\\..*", "", samples1$Geneid)
samples1 <- unique(samples1, by = "Geneid")
samples1[, entrez := mapIds(org.Hs.eg.db, keys=Geneid, keytype="ENSEMBL", column="ENTREZID")]
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
samples1[, symbol := mapIds(org.Hs.eg.db, keys=Geneid, keytype="ENSEMBL", column="SYMBOL")]
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
samples1$Chr <- NULL
samples1$Start <- NULL
samples1$End <- NULL
samples1$Length <- NULL
samples1$Strand <- NULL
samples1 <- samples1[!is.na(samples1$symbol)]
head(samples1)
```

```
##             Geneid 3RNA-2-1_S11_L001_R1_001.fastq.gz.bam
## 1: ENSG00000223972                                    0
## 2: ENSG00000227232                                    0
## 3: ENSG00000278267                                    0
## 4: ENSG00000284332                                    0
## 5: ENSG00000237613                                    0
## 6: ENSG00000186092                                    0
##    3RNA-7_S1_L001_R1_001.fastq.gz.bam
## 1:                                  0
## 2:                                  0
## 3:                                  0
## 4:                                  0
## 5:                                  0
## 6:                                  0
##    3RNA-9-1_S12_L001_R1_001.fastq.gz.bam
## 1:                                     0
## 2:                                     0
## 3:                                     0
```

```
## 4:                                     0
## 5:                                     0
## 6:                                     0
##    3RNA-11_S2_L001_R1_001.fastq.gz.bam 3RNA-13_S9_L001_R1_001.fastq.gz.bam
## 1:                                   0                                    0
## 2:                                   0                                    0
## 3:                                   0                                    0
## 4:                                   0                                    0
## 5:                                   0                                    0
## 6:                                   0                                    0
##    3RNA-14_S10_L001_R1_001.fastq.gz.bam
## 1:                                    0
## 2:                                    0
## 3:                                    0
## 4:                                    0
## 5:                                    0
## 6:                                    0
##    3RNA-19_S5_L001_R1_001.fastq.gz.bam 3RNA-22_S7_L001_R1_001.fastq.gz.bam
## 1:                                   0                                    0
## 2:                                   0                                    0
## 3:                                   0                                    0
## 4:                                   0                                    0
## 5:                                   0                                    0
## 6:                                   0                                    0
##    3RNA-HF1-1_S3_L001_R1_001.fastq.gz.bam
## 1:                                      0
## 2:                                      0
## 3:                                      0
## 4:                                      0
## 5:                                      0
## 6:                                      0
##    3RNA-HF1-2_S4_L001_R1_001.fastq.gz.bam
## 1:                                      0
## 2:                                      0
## 3:                                      0
## 4:                                      0
## 5:                                      0
## 6:                                      0
##    3RNA-HF12_S6_L001_R1_001.fastq.gz.bam
## 1:                                     0
## 2:                                     0
## 3:                                     0
## 4:                                     0
## 5:                                     0
## 6:                                     0
##    3RNA-Mal_S8_L001_R1_001.fastq.gz.bam    entrez    symbol
## 1:                                   0 100287102   DDX11L1
## 2:                                   0    653635    WASH7P
## 3:                                   0 102466751 MIR6859-1
## 4:                                   0 100302278 MIR1302-2
## 5:                                   0    645520   FAM138A
## 6:                                   0     79501     OR4F5
```

```
samples2$Geneid <- sub("\\..*", "", samples2$V1)
samples2$V1 <- NULL
samples2 <- unique(samples2, by = "Geneid")
samples2[, entrez := mapIds(org.Hs.eg.db, keys=Geneid, keytype="ENSEMBL", column="ENTREZID")]
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
samples2[, symbol := mapIds(org.Hs.eg.db, keys=Geneid, keytype="ENSEMBL", column="SYMBOL")]
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
samples2 <- samples2[!is.na(samples2$symbol)]
head(samples2)
```

```
##    3RNA-11_S2_L001_R1_001.fastq.gz 3RNA-13_S9_L001_R1_001.fastq.gz
## 1:                               1                               1
## 2:                               1                               0
## 3:                               0                               0
## 4:                               3                               2
## 5:                               0                               0
## 6:                               2                               2
##    3RNA-14_S10_L001_R1_001.fastq.gz 3RNA-19_S5_L001_R1_001.fastq.gz
## 1:                               15                               2
## 2:                                0                               3
## 3:                                2                               2
## 4:                                6                               2
## 5:                                0                               0
## 6:                                9                               5
##    3RNA-2-1_S11_L001_R1_001.fastq.gz 3RNA-22_S7_L001_R1_001.fastq.gz
## 1:                                 8                                7
## 2:                                 1                                0
## 3:                                 8                                6
## 4:                                 3                                2
## 5:                                 1                                1
## 6:                                 1                                3
##    3RNA-7_S1_L001_R1_001.fastq.gz 3RNA-9-1_S12_L001_R1_001.fastq.gz
## 1:                              2                                2
## 2:                              3                                0
## 3:                              1                                4
## 4:                              0                                0
## 5:                              0                                0
## 6:                              1                                1
##    3RNA-HF1-1_S3_L001_R1_001.fastq.gz 3RNA-HF1-2_S4_L001_R1_001.fastq.gz
## 1:                                  1                                  2
## 2:                                  2                                  1
## 3:                                  2                                  3
## 4:                                  1                                  1
## 5:                                  0                                  0
## 6:                                  1                                  1
##    3RNA-HF12_S6_L001_R1_001.fastq.gz 3RNA-Mal_S8_L001_R1_001.fastq.gz
## 1:                                 2                                6
## 2:                                 3                                8
## 3:                                 3                                0
## 4:                                 0                                2
## 5:                                 0                                0
```

```
## 6:                                0                                1
##             Geneid entrez   symbol
## 1: ENSG00000000003   7105   TSPAN6
## 2: ENSG00000000005  64102     TNMD
## 3: ENSG00000000419   8813     DPM1
## 4: ENSG00000000457  57147    SCYL3
## 5: ENSG00000000460  55732 C1orf112
## 6: ENSG00000000938   2268      FGR
```

```
nrow(samples1)
```

```
## [1] 26990
```

```
nrow(samples2)
```

```
## [1] 19359
```

Download conditions

```
conditions_full <- read.table("./conditions_full.txt", header=T,row.names=1)
row.names(conditions_full) <- sub("\\.bam", "", row.names(conditions_full))
head(conditions_full[,-1])
```

```
##                                  number Condition
## 3RNA-2-1_S11_L001_R1_001.fastq.gz     N1    before
## 3RNA-7_S1_L001_R1_001.fastq.gz        N2    before
## 3RNA-9-1_S12_L001_R1_001.fastq.gz     N1     after
## 3RNA-11_S2_L001_R1_001.fastq.gz       N2     after
## 3RNA-13_S9_L001_R1_001.fastq.gz       N3    before
## 3RNA-14_S10_L001_R1_001.fastq.gz      N3     after
```

```
conditions1 <- conditions_full$Condition
patients1 <- conditions_full$number
conditions2 <- conditions_full[(samples2[,c(1:12)][ , order(row.names(conditions_full))]),3]
patients2 <- conditions_full[(samples2[,c(1:12)][ , order(row.names(conditions_full))]),2]
```

Almost create es

```
res1 <- samples1

res1$Geneid <- NULL
res1$entrez <- NULL
res1$symbol <- NULL
rownames(res1) <- samples1$Geneid

res2 <- samples2
res2$Geneid <- NULL
res2$entrez <- NULL
res2$symbol <- NULL
rownames(res2) <- samples2$Geneid

sum(res1)
```

```
## [1] 8397223
```

```
sum(res2)
```

```
## [1] 8967523
```

```
inters_genes <- dplyr::intersect(row.names(res1),row.names(res2))
length(inters_genes)
```

```
## [1] 19359
```

es1

```
es1 <- ExpressionSet(as.matrix(res1))
fData(es1)$entrez <- samples1$entrez
fData(es1)$geneid <- samples1$Geneid
fData(es1)$symbol <- samples1$symbol
rownames(es1) <- fData(es1)$geneid
pData(es1) <- cbind(pData(es1), conditions1)
pData(es1) <- cbind(pData(es1), patients1)

es1
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 26990 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: 3RNA-2-1_S11_L001_R1_001.fastq.gz.bam
##     3RNA-7_S1_L001_R1_001.fastq.gz.bam ...
##     3RNA-Mal_S8_L001_R1_001.fastq.gz.bam (12 total)
##   varLabels: conditions1 patients1
##   varMetadata: labelDescription
## featureData
##   featureNames: ENSG00000223972 ENSG00000227232 ...
##     ENSG00000198727 (26990 total)
##   fvarLabels: entrez geneid symbol
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

es2

```
es2 <- ExpressionSet(as.matrix(res2))
fData(es2)$entrez <- samples2$entrez
fData(es2)$geneid <- samples2$Geneid
fData(es2)$symbol <- samples2$symbol
rownames(es2) <- fData(es2)$geneid
pData(es2) <- cbind(pData(es2), conditions2)
pData(es2) <- cbind(pData(es2), patients2)

es2
```

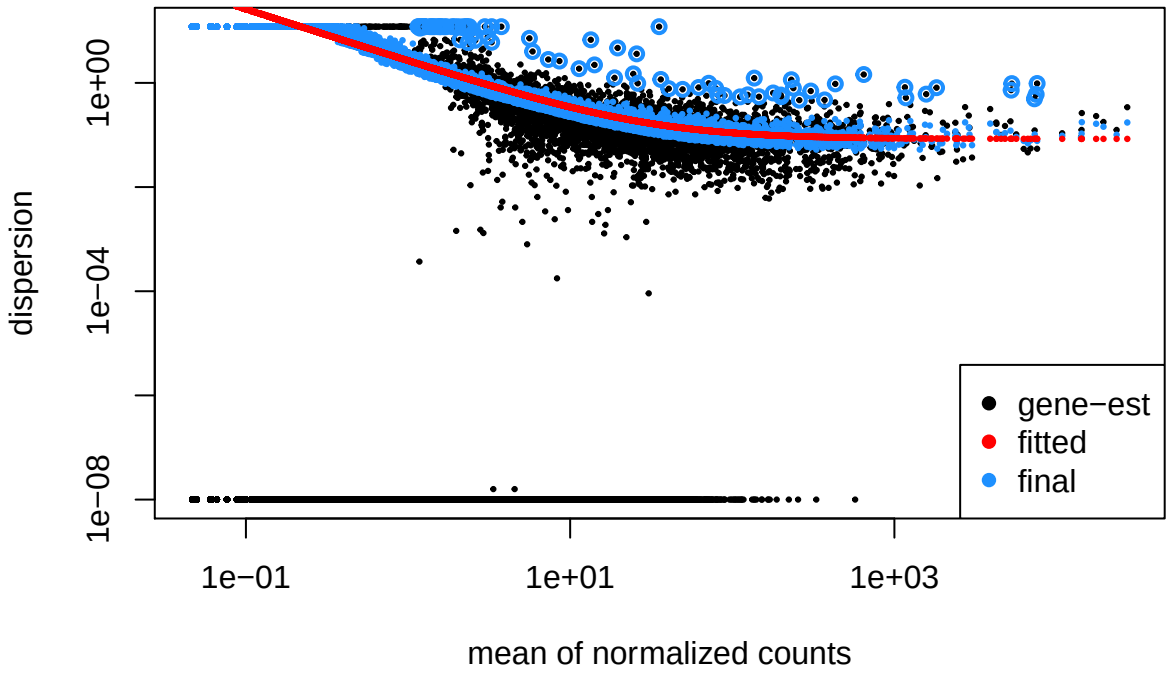
```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 19359 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: 3RNA-11_S2_L001_R1_001.fastq.gz
##     3RNA-13_S9_L001_R1_001.fastq.gz ...
##     3RNA-Mal_S8_L001_R1_001.fastq.gz (12 total)
##   varLabels: conditions2 patients2
##   varMetadata: labelDescription
## featureData
##   featureNames: ENSG00000000003 ENSG00000000005 ...
##     ENSG00000285472 (19359 total)
##   fvarLabels: entrez geneid symbol
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

DESeq 1

```
cond1 <- "before"
cond2 <- "after"

dds1        <- DESeqDataSetFromMatrix(countData=exprs(es1),
                                      colData=pData(es1),
                                      design = ~ conditions1 + patients1)
dds1      <- DESeq(dds1)
```

```
## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing
```

```
plotDispEsts(dds1)
```




```
res1        <- results(dds1, contrast=c("conditions1",cond2,cond1))
res1 <- cbind(res1, fData(es1))
resord1     <- as.data.frame(res1[order(res1$pvalue),])
head(resord1)
```

```
##                   baseMean log2FoldChange     lfcSE      stat       pvalue
## ENSG00000142871 119.150224      1.5090348 0.2986459  5.052923 4.350986e-07
## ENSG00000130222  35.804427     -1.3221862 0.3609967 -3.662599 2.496694e-04
## ENSG00000189060  24.999156     -1.7652880 0.4857233 -3.634349 2.786833e-04
## ENSG00000100292  35.746328      1.7210152 0.4745797  3.626399 2.874012e-04
## ENSG00000170345   7.819877      2.4234908 0.6728656  3.601746 3.160873e-04
## ENSG00000167863 106.383475     -0.9642941 0.2802939 -3.440296 5.810784e-04
##                        padj entrez          geneid  symbol
## ENSG00000142871 0.006988989   3491 ENSG00000142871   CYR61
## ENSG00000130222 0.999747434  10912 ENSG00000130222 GADD45G
## ENSG00000189060 0.999747434   3005 ENSG00000189060    H1F0
## ENSG00000100292 0.999747434   3162 ENSG00000100292   HMOX1
## ENSG00000170345 0.999747434   2353 ENSG00000170345     FOS
## ENSG00000167863 0.999747434  10476 ENSG00000167863  ATP5PD
```

DESeq 2

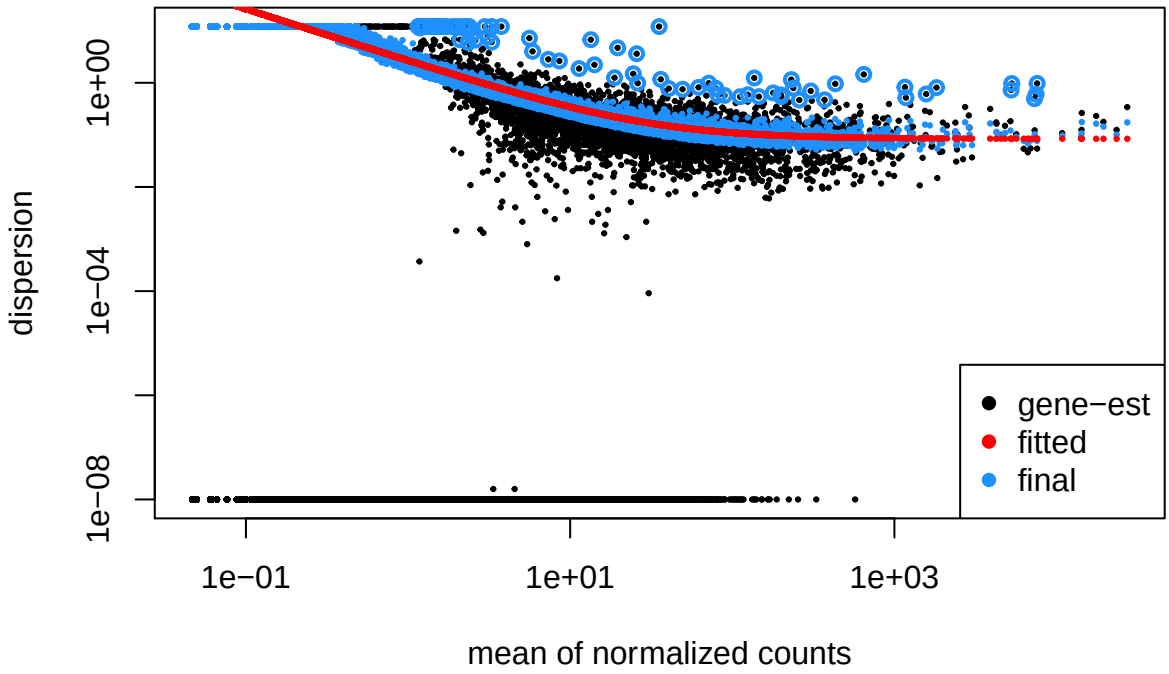
```
dds2        <- DESeqDataSetFromMatrix(countData=exprs(es2),
                                      colData=pData(es2),
                                      design = ~ conditions2 + patients2)
dds2      <- DESeq(dds2)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
plotDispEsts(dds1)
```




```
res2        <- results(dds2, contrast=c("conditions2",cond2,cond1))
res2 <- cbind(res2, fData(es2))
resord2     <- as.data.frame(res2[order(res2$pvalue),])
head(resord2)
```

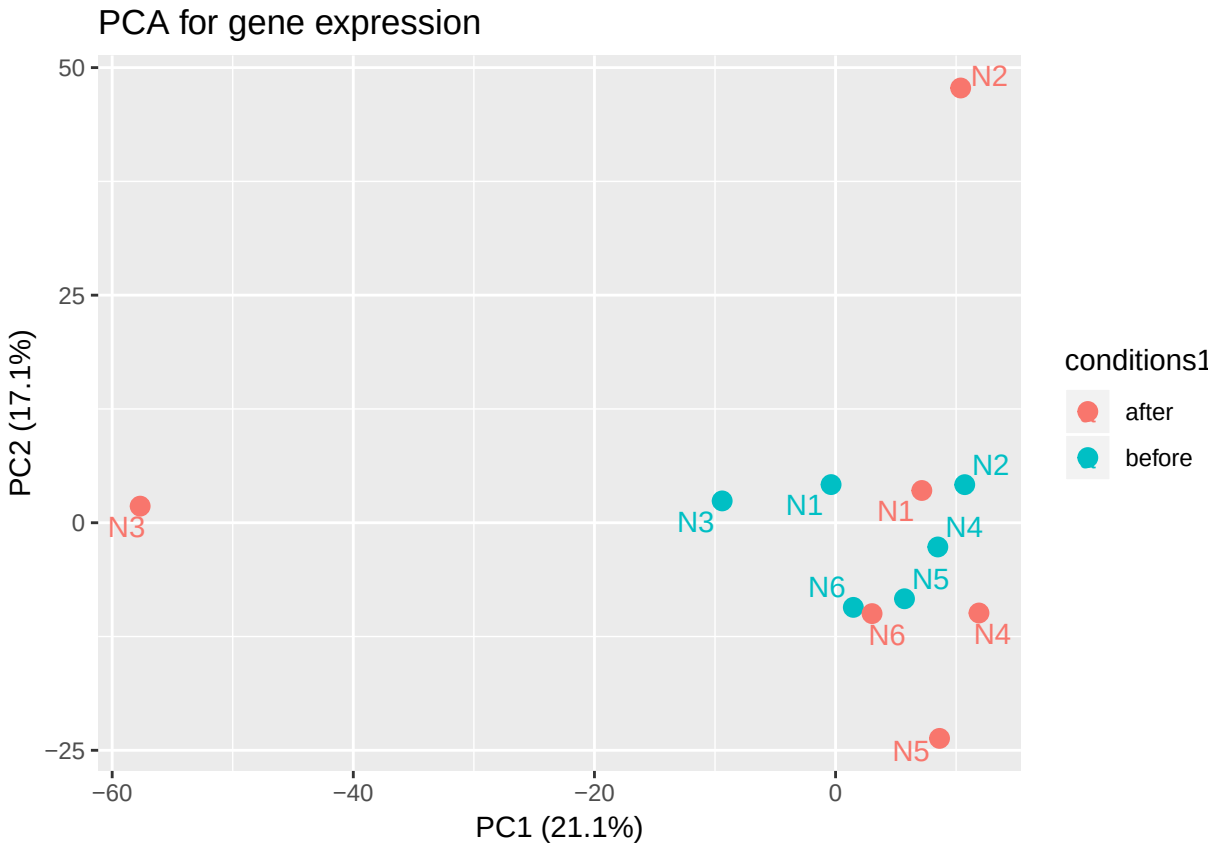
```
##                  baseMean log2FoldChange     lfcSE      stat       pvalue
## ENSG00000142871 118.813398       1.502266 0.2957836  5.078938 3.795513e-07
## ENSG00000130222  35.822536      -1.329800 0.3578862 -3.715706 2.026367e-04
## ENSG00000171223   9.371887       2.179586 0.5895916  3.696772 2.183584e-04
## ENSG00000189060  25.051735      -1.775738 0.4817697 -3.685865 2.279274e-04
## ENSG00000100292  35.672485       1.715670 0.4708210  3.643997 2.684364e-04
## ENSG00000167863 111.889550      -0.990619 0.2758036 -3.591755 3.284585e-04
##                        padj entrez          geneid  symbol
## ENSG00000142871 0.005471231   3491 ENSG00000142871   CYR61
## ENSG00000130222 0.754284011  10912 ENSG00000130222 GADD45G
## ENSG00000171223 0.754284011   3726 ENSG00000171223    JUNB
## ENSG00000189060 0.754284011   3005 ENSG00000189060    H1F0
## ENSG00000100292 0.754284011   3162 ENSG00000100292   HMOX1
## ENSG00000167863 0.754284011  10476 ENSG00000167863  ATP5PD
```

For PCA

```
es.norm1 <- es1
exprs(es.norm1) <- getVarianceStabilizedData(dds1)
```

```
es.norm1 <- es.norm1[head(order(apply(exprs(es.norm1), 1, mean), decreasing = T), 12000), ]

es.norm2 <- es2
exprs(es.norm2) <- getVarianceStabilizedData(dds2)
es.norm2 <- es.norm2[head(order(apply(exprs(es.norm2), 1, mean), decreasing = T), 12000), ]
```

```
pcaStyle <-  list(
  geom_text_repel(aes(label=patients1)),
  aes(color=conditions1),
  ggtitle("PCA for gene expression"))

pcaPlot(es.norm1, 1, 2) + pcaStyle
```




```
pcaPlot(es.norm1,2,3) + pcaStyle
```


```
pcaStyle <-  list(
  geom_text_repel(aes(label=patients2)),
  aes(color=conditions2),
  ggtitle("PCA for gene expression"))

pcaPlot(es.norm2, 1, 2) + pcaStyle
```


```
pcaPlot(es.norm2, 2, 3) + pcaStyle
```


PCA for gene expression
40
20
0
−20
PC3 (13.1%)
−25
0
25
50
PC2 (16.9%)
N3
N5
N5
N4
N4
N2
N6
N2
N6
N3
N1
N1
conditions2
after
before