

Differential expression analysis

Tuesday - part 5

Source of data:

<https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP051848>

Data download: Case pre

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/001/SRR1747301/SRR1747301_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/001/SRR1747301/SRR1747301_2.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/009/SRR1747299/SRR1747299_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/009/SRR1747299/SRR1747299_2.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/003/SRR1747303/SRR1747303_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/003/SRR1747303/SRR1747303_2.fastq.gz
```

Data download: Control pre

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/009/SRR1747399/SRR1747399_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/009/SRR1747399/SRR1747399_2.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/007/SRR1747397/SRR1747397_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/007/SRR1747397/SRR1747397_2.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/005/SRR1747395/SRR1747395_1.fastq.gz
```

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR174/005/SRR1747395/SRR1747395_2.fastq.gz
```

Reads quality control: before filtering

```
mkdir FASTQC_out
```

```
mkdir FASTQC_out/SRR1747301_raw
```

```
mkdir FASTQC_out/SRR1747299_raw
```

```
mkdir FASTQC_out/SRR1747303_raw
```

```
mkdir FASTQC_out/SRR1747399_raw
```

```
mkdir FASTQC_out/SRR1747397_raw
```

```
mkdir FASTQC_out/SRR1747395_raw
```

```
FastQC/fastqc SRR1747301_1.fastq.gz SRR1747301_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747301_raw
```

```
FastQC/fastqc SRR1747299_1.fastq.gz SRR1747299_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747299_raw
```

```
FastQC/fastqc SRR1747303_1.fastq.gz SRR1747303_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747303_raw
```

```
FastQC/fastqc SRR1747399_1.fastq.gz SRR1747399_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747399_raw
```

```
FastQC/fastqc SRR1747397_1.fastq.gz SRR1747397_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747397_raw
```

```
FastQC/fastqc SRR1747395_1.fastq.gz SRR1747395_2.fastq.gz --quiet --noextract --nogroup --outdir  
FASTQC_out/SRR1747395_raw
```

```
gunzip *.gz
```

```
mkdir TRIMMED
```

```
mkdir STATUS
```

Reads filtering: bbduk2

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747301_1.fastq in2=SRR1747301_2.fastq  
out=TRIMMED/SRR1747301_bbduk2_R1.fastq out2=TRIMMED/SRR1747301_bbduk2_R2.fastq  
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe  
minlength=50 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747301.bbduk2_stats.txt 2>  
STATUS/SRR1747301.bbduk2_trimming.txt
```

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747299_1.fastq in2=SRR1747299_2.fastq  
out=TRIMMED/SRR1747299_bbduk2_R1.fastq out2=TRIMMED/SRR1747299_bbduk2_R2.fastq  
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe  
minlength=30 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747299.bbduk2_stats.txt 2>  
STATUS/SRR1747299.bbduk2_trimming.txt
```

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747303_1.fastq in2=SRR1747303_2.fastq  
out=TRIMMED/SRR1747303_bbduk2_R1.fastq out2=TRIMMED/SRR1747303_bbduk2_R2.fastq  
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe  
minlength=30 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747303.bbduk2_stats.txt 2>  
STATUS/SRR1747303.bbduk2_trimming.txt
```

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747399_1.fastq in2=SRR1747399_2.fastq
out=TRIMMED/SRR1747399_bbduk2_R1.fastq out2=TRIMMED/SRR1747399_bbduk2_R2.fastq
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe
minlength=30 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747399.bbduk2_stats.txt 2>
STATUS/SRR1747399.bbduk2_trimming.txt
```

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747397_1.fastq in2=SRR1747397_2.fastq
out=TRIMMED/SRR1747397_bbduk2_R1.fastq out2=TRIMMED/SRR1747397_bbduk2_R2.fastq
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe
minlength=30 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747397.bbduk2_stats.txt 2>
STATUS/SRR1747397.bbduk2_trimming.txt
```

```
bbmap/bbduk2.sh -Xmx2g threads=2 in=SRR1747395_1.fastq in2=SRR1747395_2.fastq
out=TRIMMED/SRR1747395_bbduk2_R1.fastq out2=TRIMMED/SRR1747395_bbduk2_R2.fastq
qtrim=w trimq=20 maq=10 rref=bbmap/resources/adapters.fa k=23 mink=11 hdist=1 tbo tpe
minlength=30 removeifeitherbad=t overwrite=t stats=STATUS/SRR1747395.bbduk2_stats.txt 2>
STATUS/SRR1747395.bbduk2_trimming.txt
```

Removing rRNA-mapping reads

```
bowtie2-build human_rRNA.fasta index/human_rRNA
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747301_bbduk2_R1.fastq -2
TRIMMED/SRR1747301_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc
SRR1747301.fastq > /dev/null
```

```
mv SRR1747301.1.fastq TRIMMED/SRR1747301_clean_R1.fastq
mv SRR1747301.2.fastq TRIMMED/SRR1747301_clean_R2.fastq
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747299_bbduk2_R1.fastq -2
TRIMMED/SRR1747299_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc
SRR1747299.fastq > /dev/null
```

```
mv SRR1747299.1.fastq TRIMMED/SRR1747299_clean_R1.fastq
mv SRR1747299.2.fastq TRIMMED/SRR1747299_clean_R2.fastq
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747303_bbduk2_R1.fastq -2
TRIMMED/SRR1747303_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc
SRR1747303.fastq > /dev/null
```

```
mv SRR1747303.1.fastq TRIMMED/SRR1747303_clean_R1.fastq
mv SRR1747303.2.fastq TRIMMED/SRR1747303_clean_R2.fastq
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747399_bbduk2_R1.fastq -2
TRIMMED/SRR1747399_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc
SRR1747399.fastq > /dev/null
```

```
mv SRR1747399.1.fastq TRIMMED/SRR1747399_clean_R1.fastq
mv SRR1747399.2.fastq TRIMMED/SRR1747399_clean_R2.fastq
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747397_bbduk2_R1.fastq -2  
TRIMMED/SRR1747397_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc  
SRR1747397.fastq > /dev/null
```

```
mv SRR1747397.1.fastq TRIMMED/SRR1747397_clean_R1.fastq  
mv SRR1747397.2.fastq TRIMMED/SRR1747397_clean_R2.fastq
```

```
bowtie2 -t -p 4 -X 1000 -1 TRIMMED/SRR1747395_bbduk2_R1.fastq -2  
TRIMMED/SRR1747395_bbduk2_R2.fastq -x index/human_rRNA --fast --un-conc  
SRR1747395.fastq > /dev/null
```

```
mv SRR1747395.1.fastq TRIMMED/SRR1747395_clean_R1.fastq  
mv SRR1747395.2.fastq TRIMMED/SRR1747395_clean_R2.fastq
```

Reads quality control: after filtering

```
mkdir FASTQC_out/SRR1747301_filtered  
mkdir FASTQC_out/SRR1747299_filtered  
mkdir FASTQC_out/SRR1747303_filtered  
mkdir FASTQC_out/SRR1747399_filtered  
mkdir FASTQC_out/SRR1747397_filtered  
mkdir FASTQC_out/SRR1747395_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747301_clean_R1.fastq TRIMMED/SRR1747301_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747301_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747299_clean_R1.fastq TRIMMED/SRR1747299_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747299_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747303_clean_R1.fastq TRIMMED/SRR1747303_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747303_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747399_clean_R1.fastq TRIMMED/SRR1747399_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747399_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747397_clean_R1.fastq TRIMMED/SRR1747397_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747397_filtered
```

```
FastQC/fastqc TRIMMED/SRR1747395_clean_R1.fastq TRIMMED/SRR1747395_clean_R2.fastq --  
quiet --noextract --nogroup --outdir FASTQC_out/SRR1747395_filtered
```

Estimating gene expressions with SALMON

Preparing an index of human transcriptome

Downloading data from ENSEMBL 88

```
wget ftp://ftp.ensembl.org/pub/release-  
88/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh38.cdna.all.fa.gz
```

```
wget ftp://ftp.ensembl.org/pub/release-88/fasta/homo_sapiens/ncrna/Homo_sapiens.GRCh38.ncrna.fa.gz
```

Merging sequences into a single transcriptome

```
cat Homo_sapiens.GRCh38.cdna.all.fa Homo_sapiens.GRCh38.ncrna.fa > human_transcriptome.fasta
```

```
mkdir SALMON_OUT # tutaj przechowywane są wyniki analizy ekspresji
```

```
Salmon/bin/salmon index -p 2 -t human_transcriptome.fasta -i index/human_transcriptome --type quasi -k 31
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747301_clean_R1.fastq -2 TRIMMED/SRR1747301_clean_R2.fastq -o SALMON_OUT/SRR1747301
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747299_clean_R1.fastq -2 TRIMMED/SRR1747299_clean_R2.fastq -o SALMON_OUT/SRR1747299
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747303_clean_R1.fastq -2 TRIMMED/SRR1747303_clean_R2.fastq -o SALMON_OUT/SRR1747303
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747399_clean_R1.fastq -2 TRIMMED/SRR1747399_clean_R2.fastq -o SALMON_OUT/SRR1747399
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747397_clean_R1.fastq -2 TRIMMED/SRR1747397_clean_R2.fastq -o SALMON_OUT/SRR1747397
```

```
Salmon/bin/salmon quant -p 4 --useVBOpt -i index/human_transcriptome -l IU -1 TRIMMED/SRR1747395_clean_R1.fastq -2 TRIMMED/SRR1747395_clean_R2.fastq -o SALMON_OUT/SRR1747395
```

Preparing a file with raw expression values (not normalized)

Look into a file **expression_samples.txt**

```
python prepare_for_deseq_salmon.py expression_samples.txt control case SALMON_OUT ensembl_data_human.txt case_vs_control.txt
```

The program returns this message:

Control: control

Treatment: case

Control samples: SRR1747399, SRR1747397, SRR1747395

Treatment samples: SRR1747301, SRR1747299, SRR1747303

Control: samples described as „control” # our control samples

Case: samples described as „case” # our test samples

SALMON_OUT: a directory with Salmon results

case_vs_control.txt: a result file

ensembl_data_human.txt: a file matching transcript IDs to gene IDs. It contains the following fields in this order:

1. Gene stable ID
2. Transcript stable ID
3. Protein stable ID
4. Chromosome/scaffold name
5. Strand Gene name
6. Gene type
7. Transcript type
8. Gene description

Identification of differentially expressed genes using three different programs

Let's start R by typing *R*.

DESeq 2

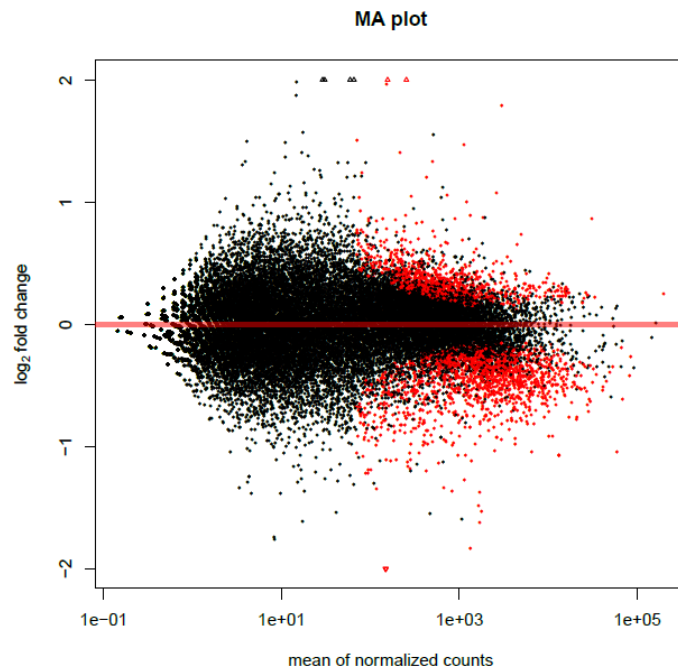
```
library(DESeq2)
library(gplots)

counts <- read.delim('case_vs_control.txt', sep="\t", header=T, row.names=1)
counts <- as.matrix(counts)
design <- data.frame( condition=factor( c("control", "control", "control", "test", "test", "test") ) )
rownames(design) <- colnames(counts)
dataset <- DESeqDataSetFromMatrix(countData = counts, colData = design, design = ~condition)
dataset <- DESeq(dataset)
de_results <- results(dataset)
new_columns <- data.frame(GeneID=rownames(de_results))
de_results <- cbind( new_columns, de_results)

de_results <- de_results[ de_results$padj < 0.05 & complete.cases(de_results$padj), ]
de_results <- de_results[order(de_results$padj),]

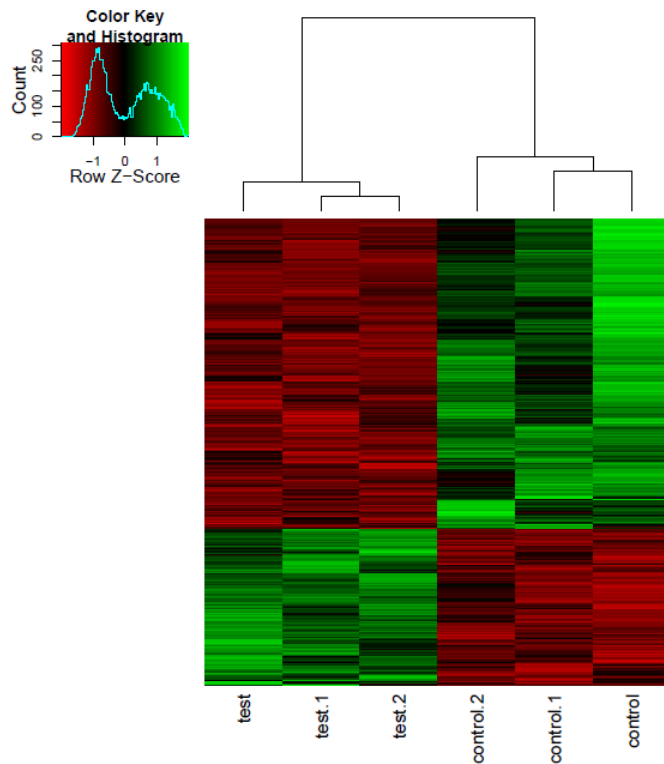
write.table(de_results, file='deseq2.tsv', sep="\t", quote=F, row.names=F)

pdf("case_vs_control_MA_plot.pdf")
plotMA(dataset, main="MA plot", ylim=c(-2, 2))
dev.off()
```

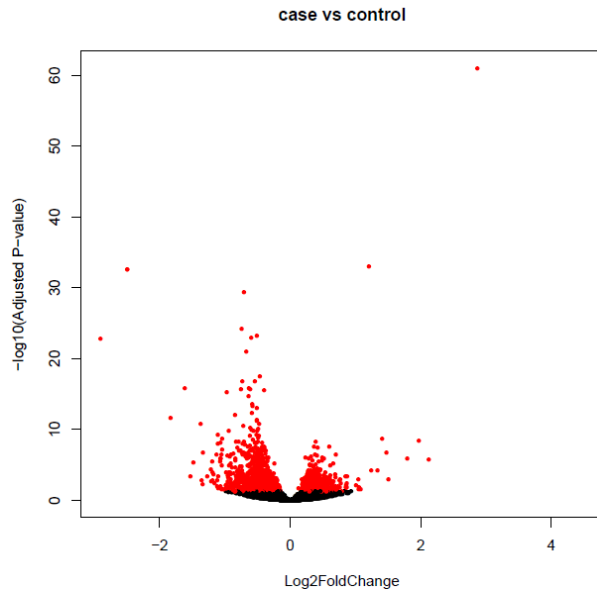


```
normalized_expression <- counts(dataset, normalized=T)
de_genes <- results(dataset)
```

```
genes <- normalized_expression[de_genes$padj < 0.05 & complete.cases(de_genes$padj),]
pdf("case_vs_control_heatmap.pdf")
heatmap.2( as.matrix(genes), labRow=F, col=redgreen(100), scale="row", dendrogram="column",
  trace="none", cexCol=1.2, hclust=function(x) hclust(x,method="centroid"), distfun=function(x)
  as.dist(1-cor(t(x))) )
dev.off()
```



```
pdf("case_vs_control_VolcanoPlot.pdf")
plot(main = "case vs control", de_genes$log2FoldChange, -
log10(de_genes$padj), pch=19, cex=0.5, xlab="Log2FoldChange", ylab="-log10(Adjusted P-
value)", col=ifelse(de_genes$padj<0.05, "red", "black"))
dev.off()
```



Annotating the results

```
python deseq2_annotate.py deseq2.tsv ensembl_data_human.txt
```

deseq2.tsv: a result file

ensembl_data_human.txt: annotations (same file as above)

Look into the output file: **deseq2_annotated.tsv**

DESeq

```
library(DESeq)

counts <- read.delim("case_vs_control.txt", row.names="gene")

design <- factor( c("control", "control", "control", "stress", "stress", "stress") )

dataset <- newCountDataSet( counts, design )

dataset <- estimateSizeFactors( dataset )

dataset <- estimateDispersions( dataset )

de_results <- nbinomTest( dataset, "control", "stress" )
```



```
colnames(de_results)[1] <- "gene"

de_results <- de_results[ de_results$padj < 0.05 & complete.cases(de_results$padj), ]

de_results <- de_results[ order(de_results$padj, decreasing=F),]

write.table(de_results, file="deseq.tsv", sep="\t", quote=F, row.names=F)
```

edgeR

```
library(edgeR)

counts <- read.delim("case_vs_control.txt", row.names="gene")

design <- factor( c("control", "control", "control", "stress", "stress", "stress") )

y <- DGEList(counts=counts, group=design)

y <- calcNormFactors(y)

y <- estimateCommonDisp(y)

y <- estimateTagwiseDisp(y)

de_results <- exactTest(y, pair=c("control", "stress") )

de_results <- topTags(de_results, n=Inf)

de_results <- de_results$table

de_results <- de_results[ de_results$FDR < 0.05 & complete.cases(de_results$FDR), ]

de_results <- de_results[ order(de_results$FDR, decreasing=F),]

de_results <- cbind( data.frame(Geneid=rownames(de_results), Fold_change=2**de_results$logFC),
de_results)

write.table(de_results, file="edger.tsv", sep="\t", quote=F, row.names=F)
```

Venn diagrams

Let's create a Venn diagram to compare results obtained with DESeq, DESeq 2 and edgeR.

Loading the library

```
library(VennDiagram)
```

Loading tables with differentially expressed genes

```
deseq2 <- read.delim( "deseq2.tsv", sep="\t", header=T )
```

```
deseq <- read.delim( "deseq.tsv", sep="\t", header=T )
```

```
edger <- read.delim( "edger.tsv", sep="\t", header=T )
```

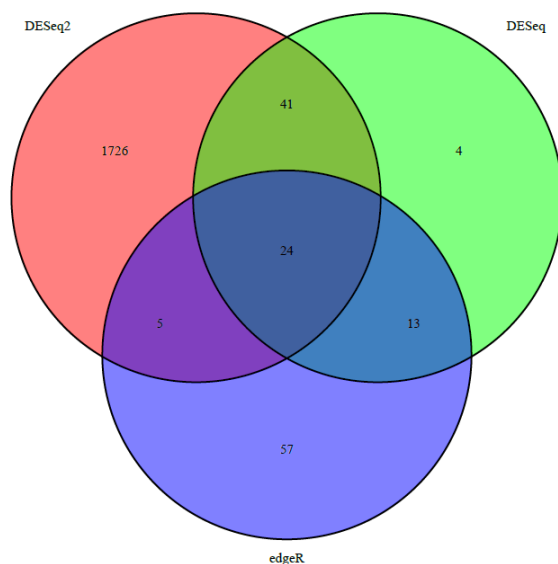
Drawing the Venn diagram

```
pdf("VennDiagram.pdf")
```

```
venn.plot <- venn.diagram( list( DESeq2 = deseq2$GeneID, DESeq = deseq$gene, edgeR =  
edger$Geneid), fill = c("red", "green", "blue"), alpha = c(0.5, 0.5, 0.5), height = 2000, width = 2000,  
filename = NULL, euler.d=F, scaled=F);
```

```
grid.draw(venn.plot);
```

```
dev.off()
```



Downstream analysis

ConsensusPathDB: over-representation analysis

Go to ConsensusPathDB (<http://consensuspathdb.org/>) and perform *over-representation analysis* (pathways-based sets + gene ontology categories), using DEG from DESeq2.

Enriched pathway-based sets (download) (show word cloud)

794 genes (71.9%) from the input list are present in at least one pathway.
The total number of genes present in at least one pathway and identifiable by 'ensembl' IDs is 12525.

| select all none | pathway name | set size | candidates contained | p-value | q-value | pathway source |
|--------------------------|---|----------|----------------------|----------|----------|----------------|
| <input type="checkbox"/> | Innate Immune System | 1309 | 189 (14.8%) | 6.79e-31 | 1.43e-27 | Reactome |
| <input type="checkbox"/> | Immune System | 1950 | 241 (12.7%) | 3.16e-29 | 3.32e-26 | Reactome |
| <input type="checkbox"/> | Neutrophil degranulation | 497 | 93 (19.2%) | 8.51e-23 | 5.96e-20 | Reactome |
| <input type="checkbox"/> | TCR | 247 | 50 (20.7%) | 4.78e-14 | 2.51e-11 | NetPath |
| <input type="checkbox"/> | Kit receptor signaling pathway | 59 | 22 (37.3%) | 3.17e-12 | 1.33e-09 | WikiPathways |
| <input type="checkbox"/> | IL-3 Signaling Pathway | 49 | 20 (40.8%) | 4.26e-12 | 1.49e-09 | WikiPathways |
| <input type="checkbox"/> | EGFR1 | 458 | 69 (15.2%) | 6.83e-12 | 2.05e-09 | NetPath |
| <input type="checkbox"/> | Fc-epsilon receptor 1 signaling in mast cells | 62 | 22 (35.5%) | 1.01e-11 | 2.65e-09 | PID |
| <input type="checkbox"/> | IL6 | 77 | 24 (32.0%) | 1.47e-11 | 3.44e-09 | NetPath |
| <input type="checkbox"/> | Tuberculosis - Homo sapiens (human) | 179 | 37 (20.8%) | 9.42e-11 | 1.95e-08 | KEGG |
| <input type="checkbox"/> | Cytoplasmic Ribosomal Proteins | 88 | 25 (28.4%) | 1.02e-10 | 1.95e-08 | WikiPathways |
| <input type="checkbox"/> | Peptide chain elongation | 103 | 26 (27.1%) | 1.39e-10 | 2.44e-08 | Reactome |

Similar tools:

- GOrilla (<http://cbl-gorilla.cs.technion.ac.il/>); only gene ontology
- PlantGSEA (<http://structuralbiology.cau.edu.cn/PlantGSEA/>); gene ontology, pathways
- DAVID (<https://david.ncifcrf.gov/>); universal tool
- TopGO (R: <http://www.bioconductor.org/packages/release/bioc/html/topGO.html>)
- GSEAServer (plantgrn.noble.org/GSEAServer); non-model organisms
- KAAS (<http://www.genome.jp/tools/kaas/>); transcriptome annotation

Addendum

What are the requirements for RNA-Seq experiments?

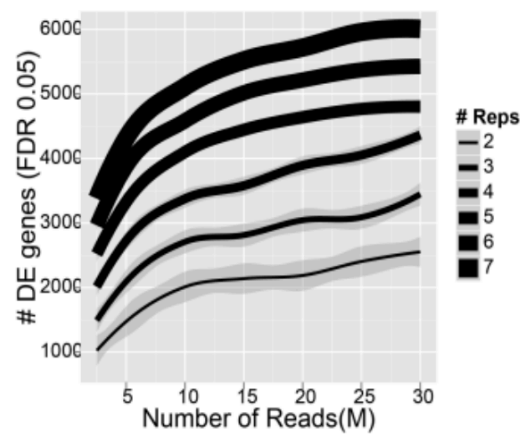
| Category | Detection or Application | Recommended Coverage (x) or Reads (millions) | References |
|---------------------------------|-----------------------------------|--|--|
| Transcriptome Sequencing | Differential expression profiling | 10-25M | Liu Y. et al., 2014; ENCODE 2011 RNA-Seq |
| | Alternative splicing | 50-100M | Liu Y. et al., 2013; ENCODE 2011 RNA-Seq |
| | Allele specific expression | 50-100M | Liu Y. et al., 2013; ENCODE 2011 RNA-Seq |
| | De novo assembly | >100M | Liu Y. et al., 2013; ENCODE 2011 RNA-Seq |
| Small RNA (microRNA) Sequencing | Differential Expression | ~1-2M | Metpally RPR et al., 2013; Campbell et al., 2015 |
| | Discovery | ~5-8M | Metpally RPR et al., 2013; Campbell et al., 2015 |

For comparison:

| Category | Detection or Application | Recommended Coverage (x) or Reads (millions) | References |
|-------------------------|--------------------------|--|--|
| Whole genome sequencing | Homozygous SNVs | 15x | Bentley et al., 2008 |
| | Heterozygous SNVs | 33x | Bentley et al., 2008 |
| | INDELs | 60x | Feng et al., 2014 |
| | Genotype calls | 35x | Ajay et al., 2011 |
| | CNV | 1-8x | Xie et al., 2009; Medvedev et al., 2010 |
| Whole exome sequencing | Homozygous SNVs | 100x (3x local depth) | Clark et al., 2011; Meynert et al., 2013 |
| | Heterozygous SNVs | 100x (13x local depth) | Clark et al., 2011; Meynert et al., 2013 |
| | INDELs | not recommended | Feng et al., 2014 |

Source: <https://genohub.com/recommended-sequencing-coverage-by-application/>

Number of replicates vs coverage / sequencing depth:



Source: Liu Y, Zhou J, White KP. *RNA-seq differential expression studies: more sequence or more replication?* Bioinformatics. 2014