**#1 Install edge R**

cd  
 git clone <https://github.com/ctb/2017-ucdavis-igg201b.git>  
  
sudo Rscript --no-save ~/2017-ucdavis-igg201b/lab7/install-edgeR.R

**#2 Copy salmon**

cd  
 curl -L -O <https://github.com/COMBINE-lab/salmon/releases/download/v0.8.0/Salmon-0.8.0_linux_x86_64.tar.gz>  
 tar xzf Salmon-0.8.0\_linux\_x86\_64.tar.gz  
 export PATH=$PATH:$HOME/Salmon-latest\_linux\_x86\_64/bin

**#3 make directory named yeast**

mkdir yeast  
 cd yeast

**# 4 Load data files for both mutant and wild type (6 files from 2 different biological replicates for mutant and same for wild type)**

#Mutant Data Files:

#From Biological Rep 1 (3 technical replicates)

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458500/ERR458500.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458501/ERR458501.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458502/ERR458502.fastq.gz>

#From Biological Rep 2 (3 technical replicates)

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458507/ERR458507.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458508/ERR458508.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458509/ERR458509.fastq.gz>

#Wild Data Files:

#From Biological Rep 1

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458493/ERR458493.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458494/ERR458494.fastq.gz>

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458495/ERR458495.fastq.gz>

#From Biological Rep 2

curl -O [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458878/ERR458878.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458948/ERR458948.fastq.gz)

curl -O [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458879/ERR458879.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458949/ERR458949.fastq.gz)

curl -O <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR458/ERR458880/ERR458880.fastq.gz>

#(Type ls to make sure all 12 files have been uploaded)

**#5 Download the yeast reference transcriptome:**

curl -O <http://downloads.yeastgenome.org/sequence/S288C_reference/orf_dna/orf_coding.fasta.gz>

**#6 Index the yeast transcriptome:**

salmon index --index yeast\_orfs --type quasi --transcripts orf\_coding.fasta.gz

**#7 Run salmon on all the samples:**

for i in \*.fastq.gz

do

salmon quant -i yeast\_orfs --libType U -r $i -o $i.quant --seqBias --gcBias

done

**#8 Collect all of the sample counts using**[**this Python script**](https://github.com/ngs-docs/2016-aug-nonmodel-rnaseq/blob/master/files/gather-counts.py):

curl -L -O <https://github.com/ngs-docs/2016-aug-nonmodel-rnaseq/raw/master/files/gather-counts.py>  
 python2 gather-counts.py

**#9 Run edgeR using the new script (more biological replicates are added for both wild type and mutant type)**

git clone <https://github.com/northstr/GGG201B2017LabHW.git>

Rscript --no-save ~/GGG201B2017LabHW/H.txt

**#10 Viewing the plot**

Jupyter console >>yeast >> yeast-MA-plot.pdf and yeast-MDS-plot.pdf

#

git clone <https://github.com/ctb/2017-ucdavis-igg201b.git>

git clone <https://github.com/northstr/GGG201B2017LabHW.git>

#

cd ~/2017-ucdavis-igg201b/lab9/

cd ~/GGG201B2017LabHW

mkdir functional-analysis

cd functional-analysis

mkdir functional-analysis

cd functional-analysis

**#**

**R**

**R**

**#**

gene.data <- read.csv(file='../yeast-edgeR.csv', row.names=1)

gene.data.new <- read.csv(file='../HW3.csv', row.names=1)

**#**

de.genes <- subset(gene.data, FDR < 0.05)

up.genes <-subset(de.genes, logFC > 1)

up.genes.names <-row.names(up.genes)

length(up.genes.names)

#800

de.genes <- subset(gene.data, FDR < 0.05)

up.genes <-subset(de.genes, logFC > 1)

up.genes.names <-row.names(up.genes)

length(up.genes.names)

182

Adding more files for different biological replicates, we lost gene 618 genes

182 genes are differentially expressed in two biological replicates for mutants that were not expressed in 2 biological replicates of wild types at an FDR 0.05 for both datasets use in this homework an in Lab 8

MDS plot for new dataset has clearly clustered the technical replicates within chosen biological replicates , in both wild type and mutants. Thus we can trust our samples

Try the whole analysis again chosin respective replicates for both wild type and mutant type

**#Comments:**