* Transferred files from /mnt/jmaresca to /work/maurielm/working

cd /mnt/jmaresca

cp 180411\_Donofrio.tar /work/maurielm/working

cp 180812-Donofrio.tar /work/maurielm/working

cp ND-5837.tar.gz /work/maurielm/working

* Extracted the .tar files

cd /work/maurielm/working

tar -xvf 180411\_Donofrio.tar

tar -xvf 180812-Donofrio.tar

* This writes files to a new folder, /work/maurielm/working, in a .fastq.gz format
* Because the L001 files all unzipped into their respective folders, the files were moved into the /ND-5837 folder to have all files together. There should be 72 .fastq.gz files. Back up these files.
* Write FastQC script, saved to /work/maurielm/working/ND-5837

mkdir fastqc

nano fqc.sh

#!/bin/bash

#SBATCH --job-name=fastqc

#SBATCH --ntasks=8

#SBATCH --mem=16000

/usr/local/FastQC/fastqc -j /usr/bin/java /work/maurielm/working/ND-5837/\*.fastq -outdir=/work/maurielm/working/ND-5837/fastqc -t 4

* Ran the script:

sbatch fqc.sh

* Wrote a script to aggregate results, saved to work/maurielm/working/ND-5837/fastqc

#!/bin/bash

#SBATCH – job-name=fq\_agg

#SBATCH –ntasks=8

#SBATCH –mem=16000

zips=`ls \*.zip`

for i in $zips; do

unzip -o $i &>/dev/null;

done

fastq\_folders=$(zips/.zip/}

rm -rf fq\_aggregated

mkdir fq\_aggregated

for folder in $fastq\_folders; do

folder=${folder%.\*}

img\_files=`ls ${folder}/Images/\*png`;

for img in $img\_files; do

img\_name=$(basename "$img");

img\_name=${img\_name%.\*}

new\_name=${folder};

mkdir -p fq\_aggregated/${img\_name};

mv $img fq\_aggregated/${img\_name}/${folder/\_fastqc/}.png;

done;

done;

for folder in $fastq\_folders; do

folder=${folder%.\*}

cat ${folder}/summary.txt >> fq\_aggregated/summary.txt

done;

for folder in $fastq\_folders; do

folder=${folder%.\*}

head -n 10 ${folder}/fastqc\_data.txt | tail -n 7 | awk -v f=${folder/\_fastqc/} '{ print $0 "\t" f }' >> fq\_aggregated/statistics.txt

rm -rf ${folder}

done

* Ran the script:

sbatch fastqc\_agg.sh

* Results were saved to /work/maurielm/working/ND-5837/fastqc/fq\_aggregated
* Unzipped the fastqc outputs:

cd /work/maurielm/working/ND-5837/fastqc

unzip \\*zip

* Copied the .fastq.gz files (see step 4) to /work/maurielm/working/ND-5837/
* Write the trimming script, stored at /work/maurielm/working/ND-5837/fastqc:

#!/bin/bash

#SBATCH --job-name=cutadapt

#SBATCH --ntasks=8

#SBATCH --mem=16000

cd /work/maurielm/working/ND-5837

mkdir trim50

for R1 in \*\_L001\_R1\_001.fastq.gz

do

R2=${R1%%\_L001\_R1\_001.fastq.gz}"\_L001\_R2\_001.fastq.gz"

cutadapt -u 50 -U 50 --no-trim -q 0 -o /work/maurielm/working/ND-5837/trim50/$R1 -p /work/maurielm/working/ND-5837/trim50/$R2 $R1 $R2

done

* Ran the trimming script:

sbatch cutadapt.sh

* The reads were trimmed.  
  The output is stored at /work/maurielm/working/ND-5837/fastqc/slurm-56018.out and the trimmed reads are stored at /work/maurielm/working/ND-5837/trim50
* Wrote the trim galore script to quality filter and trim:

#!/bin/bash

#SBATCH --job-name=trim\_galore

#SBATCH --ntasks=8

#SBATCH --mem=16000

cd /work/maurielm/working/ND-5837

mkdir trim28

export PATH=$PATH:/usr/local/FastQC

for R1 in \*\_R1\_001.fastq.gz

do

R2=${R1%%\_R1\_001.fastq.gz}"\_R2\_001.fastq.gz"

trim\_galore -q 28 --fastqc --length 35 --paired --stringency 8 -o /work/maurielm/working/ND-5837/trim28 $R1 $R2

done

* Ran the trim galore script:

sbatch trim\_galore.sh

* Per the script, the outputs were saved to /work/maurielm/working/ND-5837/trim28

There are a total of 288 files - 4 for each input file.

* Wrote the script to combine technical replicates (same sample sequenced twice - once in each lane).

Saved as /work/maurielm/working/ND-5837/trim28/Combine\_Reps.sh

Note: the first concatenation (listed below) was done in a separate script to test the command:

cat 4091\_B\_1\_S19\_L002\_R1\_001\_val\_1.fq.gz 4091-B-1\_S4\_L001\_R1\_001\_val\_1.fq.gz > 4091\_B\_R1\_val1\_4091vB1R1.fq.gz

* Ran the script:

sbatch Combine\_Reps.sh

* Moved combined files to subfolder /work/maurielm/working/ND-5837/trim28/Combined
* Copy and edit the fastqc script to run on combined samples:

cp /work/maurielm/working/ND-5837/fqc.sh /work/maurielm/working/ND-5837/trim28/Combined

#!/bin/bash

#SBATCH --job-name=fastqc

#SBATCH --ntasks=8

#SBATCH --mem=16000

/usr/local/FastQC/fastqc -j /usr/bin/java /work/maurielm/working/ND-5837/trim28/Combined/\*fq.gz -outdir=/work/maurielm/working/ND-5837/trim28/Combined/FastQC -t 4

* Run fastqc:

sbatch fqc.sh

* Copy the aggregation script and run on combined samples:

cp /work/maurielm/working/ND-5837/fastqc/fastqc\_agg.sh /work/maurielm/working/ND-5837/trim28/Combined/FastQC

sbatch fastqc\_agg.sh

* An overview of the summary file:

egrep PASS -c /work/maurielm/working/ND-5837/trim28/Combined/FastQC/fq\_aggregated/summary.txt

181

egrep WARN -c /work/maurielm/working/ND-5837/trim28/Combined/FastQC/fq\_aggregated/summary.txt

117

egrep FAIL -c /work/maurielm/working/ND-5837/trim28/Combined/FastQC/fq\_aggregated/summary.txt

98

* Get the RefSeq genome assembly data from NCBI: <https://ftp.ncbi.nlm.nih.gov/genomes/refseq/fungi/Pyricularia_oryzae/latest_assembly_versions/GCF_000002495.2_MG8/> (22 March 2023)

~$ scp /home/binfcert/Desktop/GCF\_000002495.2\_MG8\_cds\_from\_genomic.fna.gz maurielm@biomix.dbi.udel.edu:/work/maurielm/working/ND-5837/GCF

~$ scp /home/binfcert/Desktop/GCF\_000002495.2\_MG8\_rna\_from\_genomic.fna.gz maurielm@biomix.dbi.udel.edu:/work/maurielm/working/ND-5837/GCF

~$ scp /home/binfcert/Desktop/GCF\_000002495.2\_MG8\_genomic.gff.gz

~$ scp /home/binfcert/Desktop/GCF\_000002495.2\_MG8\_genomic.gbff.gz

~$ scp /home/binfcert/Desktop/GCF\_000002495.2\_MG8\_genomic.fna.gz

* Unzip the files:

cd /work/maurielm/working/ND-5837/GCF

gunzip \*gz

* Copied the following scripts from cp /work/binf694/project\_1/scripts/ to /work/maurielm/working/ND-5837/

cp /work/binf694/project\_1/scripts/mapping.slurm /work/maurielm/working/ND-5837/

cp /work/binf694/project\_1/scripts/bamindex.slurm /work/maurielm/working/ND-5837/

cp /work/binf694/project\_1/scripts/counting.slurm /work/maurielm/working/ND-5837/

* Rename the .fna file to .fa format (fasta nucleic acid to fasta)

cp GCF\_000002495.2\_MG8\_cds\_from\_genomic.fna

cp GCF\_000002495.2\_MG8\_rna\_from\_genomic.fna

cp GCF\_000002495.2\_MG8\_genomic.fna

* Build HISAT2 index:

#Change the file names

mv GCF\_000002495.2\_MG8\_genomic.fa Mo\_genome.fa

mv GCF\_000002495.2\_MG8\_genomic.gff Mo\_genome.gff

#Version: HISAT2 2.2.1. Write a script to run hisat2:

#!/bin/bash

#SBATCH --job-name=hisat2

#SBATCH --ntasks=8

#SBATCH --mem=16000

PATH=$PATH:/usr/local/hisat2

hisat2-build -p 16 Mo\_genome.fa genome

* Ran the script:

sbatch hisat2.sh

* Modified the mapping script as follows:

#!/bin/bash

#SBATCH --job-name=mapping

#SBATCH --mem=32000

#SBATCH -c 16

#JOB LOG HEADER

perl -E 'say"="x80'; echo "JOB STARTED: `date`"; echo "NODE: `hostname`"; echo "SCRIPT ${0}:"; cat $0; perl -E 'say"="x80'

cd /work/maurielm/working/ND-5837/Mapping

#SOFTWARE REQUIREMENTS

export PATH=$PATH:/usr/local/hisat2

INFILE=/work/maurielm/working/ND-5837/trim28/Combined/4091\_B\_R1\_val1\_4091vB1R1.fq.gz

INDEX=/work/maurielm/working/ND-5837/GCF/genome

PREFIX=`basename -s ".fq.gz" ${INFILE}`

OUTFILE=/work/maurielm/working/ND-5837/Mapping/${PREFIX}.bam

# COMMAND(s) TO RUN

hisat2 -p ${SLURM\_CPUS\_PER\_TASK} -x ${INDEX} -U ${INFILE} \

| samtools view -Sbo ${OUTFILE} -

#JOB LOG FOOTER

perl -E 'say"="x80'; echo "JOB COMPLETED: `date`"; perl -E 'say"="x80'

* Ran the script:

sbatch mapping.slurm

#Run for each of the 36 input files in /work/maurielm/working/ND-5837/trim28/Combined/

* Edited the bamindex script:

#!/bin/bash

#SBATCH --job-name=bamindex

#SBATCH --mem=64000

#SBATCH -c 16

#JOB LOG HEADER

perl -E 'say"="x80'; echo "JOB STARTED: `date`"; echo "NODE: `hostname`"; echo "SCRIPT ${0}:"; cat $0; perl -E 'say"="x80'

cd /work/maurielm/working/ND-5837/Mapping

#SOFTWARE REQUIREMENTS

#VARIABLES

INFILE=/work/maurielm/working/ND-5837/Mapping/4091\_B\_R1\_val1\_4091vB1R1.bam

PREFIX=`basename -s ".bam" ${INFILE}`

# COMMAND(s) TO RUN

# THIS COMMAND HAS A SLIGHTLY UPDATED SYNTAX DUE TO A SOFTWARE UPDATE

samtools sort -@ 16 -m 3750M -o ${PREFIX}.bam ${INFILE}

samtools index ${PREFIX}.bam

#JOB LOG FOOTER

perl -E 'say"="x80'; echo "JOB COMPLETED: `date`"; perl -E 'say"="x80'

#JOB LOG FOOTER

perl -E 'say"="x80'; echo "JOB COMPLETED: `date`"; perl -E 'say"="x80'

* Ran the script:

sbatch bamindex.slurm

#Ran for each of 36 input files in /work/maurielm/working/ND-5837/Mapping/

* Edited the feature counting script:

#!/bin/bash

#SBATCH --job-name=counting

#SBATCH --mem=32000

#SBATCH -c 1

#JOB LOG HEADER

perl -E 'say"="x80'; echo "JOB STARTED: `date`"; echo "NODE: `hostname`"; echo "SCRIPT ${0}:"; cat $0; perl -E 'say"="x80'

cd /work/maurielm/working/ND-5837/Counting

#SOFTWARE REQUIREMENTS

#VARIABLES

INDIR=/work/maurielm/working/ND-5837/Mapping

GFF=/work/maurielm/working/ND-5837/GCF/Mo\_genome.gff

OUTDIR=/work/maurielm/working/ND-5837/Count\_Exons

#COMMAND(s) TO RUN

# THIS LOOP WILL RUN THE INDENTED PART ON EACH FILE ENDING IN .bam IN THE INPUT DIRECTORY

for i in `ls -1 ${INDIR}/\*.bam`

do

echo "PROCESSING FILE: ${i}"

PREFIX=`basename -s ".bam" ${i}`

htseq-count --mode intersection-nonempty \

--stranded no \

--format bam \

--type exon \

--idattr ID \

${i} \

${GFF} \

> ${OUTDIR}/${PREFIX}.count.txt

done

#JOB LOG FOOTER

perl -E 'say"="x80'; echo "JOB COMPLETED: `date`"; perl -E 'say"="x80'

* Ran the script:

Genes: 65821; Exons: 65822

* Copied count files to VirtualMachine (VM):

~/Desktop$ scp maurielm@biomix.dbi.udel.edu:/work/maurielm/working/ND-5837/Count\_Exons/\* ~/

/home/binfcert/Desktop/Counts

#pheatmap 1.1.12

#RColorBrewer 1.1-3

#BiocManager 1.30.20 (Bioconductor 3.16)

#DESeq2 1.38.3

#genefilter 1.80.3

> load("C:/Users/Megan/Desktop/Counts/Exploring\_Relationships.RData")

> library(pheatmap)

> library(RColorBrewer)

> library(BiocManager)

> library(genefilter)

> setwd("C:/Users/Megan/Desktop/Counts")

> count\_files<-c("4091\_B\_R1\_val1\_4091vB1R1.count.txt","4091\_B\_R2\_val2\_4091vB1R2.count.txt","4091\_B\_R1\_val1\_4091vB2R1.count.txt","4091\_B\_R2\_val2\_4091vB2R2.count.txt","4091\_B\_R1\_val1\_4091vB3R1.count.txt","4091\_B\_R2\_val2\_4091vB3R2.count.txt","4091\_C3\_R1\_val1\_4091vC1R1.count.txt","4091\_C3\_R2\_val2\_4091vC1R2.count.txt","4091\_C3\_R1\_val1\_4091vC2R1.count.txt","4091\_C3\_R2\_val2\_4091vC2R2.count.txt","4091\_C3\_R1\_val1\_4091vC3R1.count.txt","4091\_C3\_R2\_val2\_4091vC3R2.count.txt","4091\_S4\_R1\_val1\_4091vS1R1.count.txt","4091\_S4\_R2\_val2\_4091vS1R2.count.txt","4091\_S4\_R1\_val1\_4091vS2R1.count.txt","4091\_S4\_R2\_val2\_4091vS2R2.count.txt","4091\_S4\_R1\_val1\_4091vS3R1.count.txt","4091\_S4\_R2\_val2\_4091vS3R2.count.txt","pth11\_B\_R1\_val1\_pth11vB1R1.count.txt","pth11\_B\_R2\_val2\_pth11vB1R2.count.txt","pth11\_B\_R1\_val1\_pth11vB2R1.count.txt","pth11\_B\_R2\_val2\_pth11vB2R2.count.txt","pth11\_B\_R1\_val1\_pth11vB3R1.count.txt","pth11\_B\_R2\_val2\_pth11vB3R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC1R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC1R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC2R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC2R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC3R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC3R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS1R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS1R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS2R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS2R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS3R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS3R2.count.txt")

> strains<-c("4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4","Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("4091vB1R1","4091vB1R2","4091vB2R1","4091vB2R2","4091vB3R1","4091vB3R2","4091vC1R1","4091vC1R2","4091vC2R1","4091vC2R2","4091vC3R1","4091vC3R2","4091vS1R1","4091vS1R2","4091vS2R1","4091vS2R2","4091vS3R1","4091vS3R2","pth11vB1R1","pth11vB1R2","pth11vB2R1","pth11vB2R2","pth11vB3R1","pth11vB3R2","pth11vC1R1","pth11vC1R2","pth11vC2R1","pth11vC2R2","pth11vC3R1","pth11vC3R2","pth11vS1R1","pth11vS1R2","pth11vS2R1","pth11vS2R2","pth11vS3R1","pth11vS3R2")

> treatment\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=time)

> strains\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=strains)

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

> strain\_data<-DESeqDataSetFromHTSeqCount(sampleTable=strains\_table, design= ~ condition)

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> time\_rlog<- rlog(time\_data, blind=FALSE)

> strain\_rlog<- rlog(strain\_data, blind=FALSE)

> colors <- colorRampPalette (rev(brewer.pal(9, "Blues"))) (255)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> strain\_dist<-dist(t(assay(strain\_rlog)))

> strain\_matrix<-as.matrix(strain\_dist)

> pheatmap(strain\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_rlog, intgroup="condition")

> plotPCA(treat\_rlog, intgroup="condition")

> plotPCA(strain\_rlog, intgroup="condition")

treatVars<-rowVars(assay(treat\_rlog))

treatVarsOrdered<-order(treatVars, decreasing=TRUE)

treatTopVar<-head(treatVarsOrdered, 50)

treat\_matrix<-assay(treat\_rlog)[treatTopVar, ]

treat\_matrix<-treat\_matrix-rowMeans(treat\_matrix)

treat\_df<-as.data.frame(colData(treat\_rlog)[,c("condition"),drop=FALSE])

clear\_col\_names<-paste(treat\_rlog$condition,sep=".")

topGenesHeatmap <- pheatmap(treat\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

timeVars<-rowVars(assay(time\_rlog))

timeVarsOrdered<-order(timeVars, decreasing=TRUE)

timeTopVar<-head(timeVarsOrdered, 50)

time\_matrix<-assay(time\_rlog)[timeTopVar, ]

time\_matrix<-time\_matrix-rowMeans(time\_matrix)

time\_df<-as.data.frame(colData(time\_rlog)[,c("condition"),drop=FALSE])

clear\_col\_names<-paste(time\_rlog$condition,sep=".")

topGenesHeatmap <- pheatmap(time\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> strainVars<-rowVars(assay(strain\_rlog))

> strainVarsOrdered<-order(strainVars, decreasing=TRUE)

> strainTopVar<-head(strainVarsOrdered, 50)

> strain\_matrix<-assay(strain\_rlog)[strainTopVar, ]

> strain\_matrix<-strain\_matrix-rowMeans(strain\_matrix)

> strain\_df<-as.data.frame(colData(strain\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(strain\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(strain\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

#PCA and exon heatmaps both confirm clustering is by strain. Let's see what happens if we load in just the data for one strain or the other…

#for 4091

> counts\_4091<-c("4091\_B\_R1\_val1\_4091vB1R1.count.txt","4091\_B\_R2\_val2\_4091vB1R2.count.txt","4091\_B\_R1\_val1\_4091vB2R1.count.txt","4091\_B\_R2\_val2\_4091vB2R2.count.txt","4091\_B\_R1\_val1\_4091vB3R1.count.txt","4091\_B\_R2\_val2\_4091vB3R2.count.txt","4091\_C3\_R1\_val1\_4091vC1R1.count.txt","4091\_C3\_R2\_val2\_4091vC1R2.count.txt","4091\_C3\_R1\_val1\_4091vC2R1.count.txt","4091\_C3\_R2\_val2\_4091vC2R2.count.txt","4091\_C3\_R1\_val1\_4091vC3R1.count.txt","4091\_C3\_R2\_val2\_4091vC3R2.count.txt","4091\_S4\_R1\_val1\_4091vS1R1.count.txt","4091\_S4\_R2\_val2\_4091vS1R2.count.txt","4091\_S4\_R1\_val1\_4091vS2R1.count.txt","4091\_S4\_R2\_val2\_4091vS2R2.count.txt","4091\_S4\_R1\_val1\_4091vS3R1.count.txt","4091\_S4\_R2\_val2\_4091vS3R2.count.txt")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("4091vB1R1","4091vB1R2","4091vB2R1","4091vB2R2","4091vB3R1","4091vB3R2","4091vC1R1","4091vC1R2","4091vC2R1","4091vC2R2","4091vC3R1","4091vC3R2","4091vS1R1","4091vS1R2","4091vS2R1","4091vS2R2","4091vS3R1","4091vS3R2")

> treatment\_table<-data.frame(sampleName=samples, fileName=counts\_4091, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=counts\_4091, condition=time)

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

treat\_rlog<- rlog(treatment\_data, blind=FALSE)

time\_rlog<- rlog(time\_data, blind=FALSE)

treat\_dists<-dist(t(assay(treat\_rlog)))

treatment\_matrix<-as.matrix(treat\_dists)

pheatmap(treatment\_matrix, col=colors)

time\_dist<-dist(t(assay(time\_rlog)))

time\_matrix<-as.matrix(time\_dist)

pheatmap(time\_matrix, col=colors)

plotPCA(time\_rlog, intgroup="condition")

plotPCA(treat\_rlog, intgroup="condition")

> treatVars<-rowVars(assay(treat\_rlog))

> treatVarsOrdered<-order(treatVars, decreasing=TRUE)

> treatTopVar<-head(treatVarsOrdered, 50)

> treat\_matrix<-assay(treat\_rlog)[treatTopVar, ]

> treat\_matrix<-treat\_matrix-rowMeans(treat\_matrix)

> treat\_df<-as.data.frame(colData(treat\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(treat\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(treat\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> timeVars<-rowVars(assay(time\_rlog))

> timeVarsOrdered<-order(timeVars, decreasing=TRUE)

> timeTopVar<-head(timeVarsOrdered, 50)

> time\_matrix<-assay(time\_rlog)[timeTopVar, ]

> time\_matrix<-time\_matrix-rowMeans(time\_matrix)

> time\_df<-as.data.frame(colData(time\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(time\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(time\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

#forpth11

> counts\_pth11<-c("pth11\_B\_R1\_val1\_pth11vB1R1.count.txt","pth11\_B\_R2\_val2\_pth11vB1R2.count.txt","pth11\_B\_R1\_val1\_pth11vB2R1.count.txt","pth11\_B\_R2\_val2\_pth11vB2R2.count.txt","pth11\_B\_R1\_val1\_pth11vB3R1.count.txt","pth11\_B\_R2\_val2\_pth11vB3R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC1R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC1R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC2R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC2R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC3R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC3R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS1R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS1R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS2R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS2R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS3R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS3R2.count.txt")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("pth11vB1R1","pth11vB1R2","pth11vB2R1","pth11vB2R2","pth11vB3R1","pth11vB3R2","pth11vC1R1","pth11vC1R2","pth11vC2R1","pth11vC2R2","pth11vC3R1","pth11vC3R2","pth11vS1R1","pth11vS1R2","pth11vS2R1","pth11vS2R2","pth11vS3R1","pth11vS3R2")

> treatment\_table<-data.frame(sampleName=samples, fileName=counts\_pth11, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=counts\_pth11, condition=time)

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> time\_rlog<- rlog(time\_data, blind=FALSE)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_rlog, intgroup="condition")

> plotPCA(treat\_rlog, intgroup="condition")

R version 4.2.3 (2023-03-15 ucrt) -- "Shortstop Beagle"

Copyright (C) 2023 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Workspace loaded from C:/Users/Megan/Desktop/Counts/.RData]

Loading required package: DESeq2

Loading required package: S4Vectors

Loading required package: stats4

Loading required package: BiocGenerics

Attaching package: ‘BiocGenerics’

The following objects are masked from ‘package:stats’:

IQR, mad, sd, var, xtabs

The following objects are masked from ‘package:base’:

anyDuplicated, aperm, append, as.data.frame, basename, cbind,

colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,

get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,

match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,

Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,

table, tapply, union, unique, unsplit, which.max, which.min

Attaching package: ‘S4Vectors’

The following objects are masked from ‘package:base’:

expand.grid, I, unname

Loading required package: IRanges

Attaching package: ‘IRanges’

The following object is masked from ‘package:grDevices’:

windows

Loading required package: GenomicRanges

Loading required package: GenomeInfoDb

Loading required package: SummarizedExperiment

Loading required package: MatrixGenerics

Loading required package: matrixStats

Attaching package: ‘MatrixGenerics’

The following objects are masked from ‘package:matrixStats’:

colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,

colCounts, colCummaxs, colCummins, colCumprods, colCumsums,

colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,

colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,

colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,

colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,

colWeightedMeans, colWeightedMedians, colWeightedSds,

colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,

rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,

rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,

rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,

rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,

rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,

rowWeightedMads, rowWeightedMeans, rowWeightedMedians,

rowWeightedSds, rowWeightedVars

Loading required package: Biobase

Welcome to Bioconductor

Vignettes contain introductory material; view with

'browseVignettes()'. To cite Bioconductor, see

'citation("Biobase")', and for packages 'citation("pkgname")'.

Attaching package: ‘Biobase’

The following object is masked from ‘package:MatrixGenerics’:

rowMedians

The following objects are masked from ‘package:matrixStats’:

anyMissing, rowMedians

Loading required package: edgeR

Loading required package: limma

Attaching package: ‘limma’

The following object is masked from ‘package:DESeq2’:

plotMA

The following object is masked from ‘package:BiocGenerics’:

plotMA

> load("C:/Users/Megan/Desktop/Counts/Exploring\_Relationships.RData")

> library(pheatmap)

> library(RColorBrewer)

> library(BiocManager)

Bioconductor version 3.16 (BiocManager 1.30.20), R 4.2.3 (2023-03-15 ucrt)

> library(genefilter)

Attaching package: ‘genefilter’

The following objects are masked from ‘package:MatrixGenerics’:

rowSds, rowVars

The following objects are masked from ‘package:matrixStats’:

rowSds, rowVars

> setwd("C:/Users/Megan/Desktop/Counts")

> count\_files<-c("4091\_B\_R1\_val1\_4091vB1R1.count.txt","4091\_B\_R2\_val2\_4091vB1R2.count.txt","4091\_B\_R1\_val1\_4091vB2R1.count.txt","4091\_B\_R2\_val2\_4091vB2R2.count.txt","4091\_B\_R1\_val1\_4091vB3R1.count.txt","4091\_B\_R2\_val2\_4091vB3R2.count.txt","4091\_C3\_R1\_val1\_4091vC1R1.count.txt","4091\_C3\_R2\_val2\_4091vC1R2.count.txt","4091\_C3\_R1\_val1\_4091vC2R1.count.txt","4091\_C3\_R2\_val2\_4091vC2R2.count.txt","4091\_C3\_R1\_val1\_4091vC3R1.count.txt","4091\_C3\_R2\_val2\_4091vC3R2.count.txt","4091\_S4\_R1\_val1\_4091vS1R1.count.txt","4091\_S4\_R2\_val2\_4091vS1R2.count.txt","4091\_S4\_R1\_val1\_4091vS2R1.count.txt","4091\_S4\_R2\_val2\_4091vS2R2.count.txt","4091\_S4\_R1\_val1\_4091vS3R1.count.txt","4091\_S4\_R2\_val2\_4091vS3R2.count.txt","pth11\_B\_R1\_val1\_pth11vB1R1.count.txt","pth11\_B\_R2\_val2\_pth11vB1R2.count.txt","pth11\_B\_R1\_val1\_pth11vB2R1.count.txt","pth11\_B\_R2\_val2\_pth11vB2R2.count.txt","pth11\_B\_R1\_val1\_pth11vB3R1.count.txt","pth11\_B\_R2\_val2\_pth11vB3R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC1R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC1R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC2R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC2R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC3R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC3R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS1R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS1R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS2R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS2R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS3R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS3R2.count.txt")

> strains<-c("4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4","Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("4091vB1R1","4091vB1R2","4091vB2R1","4091vB2R2","4091vB3R1","4091vB3R2","4091vC1R1","4091vC1R2","4091vC2R1","4091vC2R2","4091vC3R1","4091vC3R2","4091vS1R1","4091vS1R2","4091vS2R1","4091vS2R2","4091vS3R1","4091vS3R2","pth11vB1R1","pth11vB1R2","pth11vB2R1","pth11vB2R2","pth11vB3R1","pth11vB3R2","pth11vC1R1","pth11vC1R2","pth11vC2R1","pth11vC2R2","pth11vC3R1","pth11vC3R2","pth11vS1R1","pth11vS1R2","pth11vS2R1","pth11vS2R2","pth11vS3R1","pth11vS3R2")

> treatment\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=time)

> strains\_table<-data.frame(sampleName=samples, fileName=count\_files, condition=strains)

> head (treatment\_table)

sampleName fileName condition

1 4091vB1R1 4091\_B\_R1\_val1\_4091vB1R1.count.txt Buffer

2 4091vB1R2 4091\_B\_R2\_val2\_4091vB1R2.count.txt Buffer

3 4091vB2R1 4091\_B\_R1\_val1\_4091vB2R1.count.txt Buffer

4 4091vB2R2 4091\_B\_R2\_val2\_4091vB2R2.count.txt Buffer

5 4091vB3R1 4091\_B\_R1\_val1\_4091vB3R1.count.txt Buffer

6 4091vB3R2 4091\_B\_R2\_val2\_4091vB3R2.count.txt Buffer

> head (treatment\_table,10)

sampleName fileName condition

1 4091vB1R1 4091\_B\_R1\_val1\_4091vB1R1.count.txt Buffer

2 4091vB1R2 4091\_B\_R2\_val2\_4091vB1R2.count.txt Buffer

3 4091vB2R1 4091\_B\_R1\_val1\_4091vB2R1.count.txt Buffer

4 4091vB2R2 4091\_B\_R2\_val2\_4091vB2R2.count.txt Buffer

5 4091vB3R1 4091\_B\_R1\_val1\_4091vB3R1.count.txt Buffer

6 4091vB3R2 4091\_B\_R2\_val2\_4091vB3R2.count.txt Buffer

7 4091vC1R1 4091\_C3\_R1\_val1\_4091vC1R1.count.txt C3

8 4091vC1R2 4091\_C3\_R2\_val2\_4091vC1R2.count.txt C3

9 4091vC2R1 4091\_C3\_R1\_val1\_4091vC2R1.count.txt C3

10 4091vC2R2 4091\_C3\_R2\_val2\_4091vC2R2.count.txt C3

> head(time\_table, 10)

sampleName fileName condition

1 4091vB1R1 4091\_B\_R1\_val1\_4091vB1R1.count.txt 0\_hpi

2 4091vB1R2 4091\_B\_R2\_val2\_4091vB1R2.count.txt 0\_hpi

3 4091vB2R1 4091\_B\_R1\_val1\_4091vB2R1.count.txt 3\_hpi

4 4091vB2R2 4091\_B\_R2\_val2\_4091vB2R2.count.txt 3\_hpi

5 4091vB3R1 4091\_B\_R1\_val1\_4091vB3R1.count.txt 9\_hpi

6 4091vB3R2 4091\_B\_R2\_val2\_4091vB3R2.count.txt 9\_hpi

7 4091vC1R1 4091\_C3\_R1\_val1\_4091vC1R1.count.txt 0\_hpi

8 4091vC1R2 4091\_C3\_R2\_val2\_4091vC1R2.count.txt 0\_hpi

9 4091vC2R1 4091\_C3\_R1\_val1\_4091vC2R1.count.txt 3\_hpi

10 4091vC2R2 4091\_C3\_R2\_val2\_4091vC2R2.count.txt 3\_hpi

> head(strain\_table, 10)

Error in h(simpleError(msg, call)) :

error in evaluating the argument 'x' in selecting a method for function 'head': object 'strain\_table' not found

> head(strains\_table, 10)

sampleName fileName condition

1 4091vB1R1 4091\_B\_R1\_val1\_4091vB1R1.count.txt 4091

2 4091vB1R2 4091\_B\_R2\_val2\_4091vB1R2.count.txt 4091

3 4091vB2R1 4091\_B\_R1\_val1\_4091vB2R1.count.txt 4091

4 4091vB2R2 4091\_B\_R2\_val2\_4091vB2R2.count.txt 4091

5 4091vB3R1 4091\_B\_R1\_val1\_4091vB3R1.count.txt 4091

6 4091vB3R2 4091\_B\_R2\_val2\_4091vB3R2.count.txt 4091

7 4091vC1R1 4091\_C3\_R1\_val1\_4091vC1R1.count.txt 4091

8 4091vC1R2 4091\_C3\_R2\_val2\_4091vC1R2.count.txt 4091

9 4091vC2R1 4091\_C3\_R1\_val1\_4091vC2R1.count.txt 4091

10 4091vC2R2 4091\_C3\_R2\_val2\_4091vC2R2.count.txt 4091

> Data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> Data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> Data<-DESeqDataSetFromHTSeqCount(sampleTable=strains\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> strain\_data<-DESeqDataSetFromHTSeqCount(sampleTable=strain\_table, design= ~ condition)

Error in h(simpleError(msg, call)) :

error in evaluating the argument 'x' in selecting a method for function 'as.data.frame': object 'strain\_table' not found

> strain\_data<-DESeqDataSetFromHTSeqCount(sampleTable=strains\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> head (time\_data, 10)

class: DESeqDataSet

dim: 10 36

metadata(1): version

assays(1): counts

rownames(10): exon-MGG\_20000-1 exon-MGG\_20001-1 ... exon-MGG\_20008-1

exon-MGG\_20009-1

rowData names(0):

colnames(36): 4091vB1R1 4091vB1R2 ... pth11vS3R1 pth11vS3R2

colData names(1): condition

> head (strain\_data, 10)

class: DESeqDataSet

dim: 10 36

metadata(1): version

assays(1): counts

rownames(10): exon-MGG\_20000-1 exon-MGG\_20001-1 ... exon-MGG\_20008-1

exon-MGG\_20009-1

rowData names(0):

colnames(36): 4091vB1R1 4091vB1R2 ... pth11vS3R1 pth11vS3R2

colData names(1): condition

> head (treatment\_data, 10)

class: DESeqDataSet

dim: 10 36

metadata(1): version

assays(1): counts

rownames(10): exon-MGG\_20000-1 exon-MGG\_20001-1 ... exon-MGG\_20008-1

exon-MGG\_20009-1

rowData names(0):

colnames(36): 4091vB1R1 4091vB1R2 ... pth11vS3R1 pth11vS3R2

colData names(1): condition

> time\_filt<- data[rowSums(counts(time\_data)) >1]

Error in data[rowSums(counts(time\_data)) > 1] :

object of type 'closure' is not subsettable

> time\_filt<-data[rowSums(counts(time\_data)) >1]

Error in data[rowSums(counts(time\_data)) > 1] :

object of type 'closure' is not subsettable

> treat\_filt<-data[rowSums(counts(treatment\_data)) >1]

Error in data[rowSums(counts(treatment\_data)) > 1] :

object of type 'closure' is not subsettable

> strain\_filt<-data[rowSums(counts(strain\_data)) >1]

Error in data[rowSums(counts(strain\_data)) > 1] :

object of type 'closure' is not subsettable

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

rlog() may take a few minutes with 30 or more samples,

vst() is a much faster transformation

> time\_rlog<- rlog(time\_data, blind=FALSE)

rlog() may take a few minutes with 30 or more samples,

vst() is a much faster transformation

> strain\_rlog<- rlog(strain\_data, blind=FALSE)

rlog() may take a few minutes with 30 or more samples,

vst() is a much faster transformation

> treat\_dist<- dist(t(assay(treatment\_data)))

> time\_dist<- dist(t(assay(time\_data)))

> strain\_dist<- dist(t(assay(strain\_data)))

> strain\_matrix<- as.matrix(strain\_dist)

> time\_matrix<- as.matrix(time\_dist)

> treatment\_matrix<- as.matrix(treatment\_dist)

Error in h(simpleError(msg, call)) :

error in evaluating the argument 'x' in selecting a method for function 'as.matrix': object 'treatment\_dist' not found

> treatment\_matrix<- as.matrix(treat\_dist)

> rownames(strain\_matrix)<-paste(strain\_data$ WTvB1R1, strain\_data$ WTvB1R2, strain\_data$ WTvB2R1, strain\_data$ WTvB2R2, strain\_data$ WTvB3R1, strain\_data$ WTvB3R2, strain\_data$ WTvC1R1, strain\_data$ WTvC1R2, strain\_data$ WTvC2R1, strain\_data$ WTvC2R2, strain\_data$ WTvC3R1, strain\_data$ WTvC3R2, strain\_data$ WTvS1R1, strain\_data$ WTvS1R2, strain\_data$ WTvS2R1, strain\_data$ WTvS2R2, strain\_data$ WTvS3R1, strain\_data$ WTvS3R2, strain\_data$pth11vB1R1, strain\_data$pth11vB1R2, strain\_data$pth11vB2R1, strain\_data$pth11vB2R2, strain\_data$pth11vB3R1, strain\_data$pth11vB3R2, strain\_data$pth11vC1R1, strain\_data$pth11vC1R2, strain\_data$pth11vC2R1, strain\_data$pth11vC2R2, strain\_data$pth11vC3R1, strain\_data$pth11vC3R2, strain\_data$pth11vS1R1, strain\_data$pth11vS1R2, strain\_data$pth11vS2R1, strain\_data$pth11vS2R2, strain\_data$pth11vS3R1, strain\_data$pth11vS3R2)

> rownames(strain\_matrix)<-paste(strain\_data$ WTvB1R1, strain\_data$ WTvB1R2, strain\_data$ WTvB2R1, strain\_data$ WTvB2R2, strain\_data$ WTvB3R1, strain\_data$ WTvB3R2, strain\_data$ WTvC1R1, strain\_data$ WTvC1R2, strain\_data$ WTvC2R1, strain\_data$ WTvC2R2, strain\_data$ WTvC3R1, strain\_data$ WTvC3R2, strain\_data$ WTvS1R1, strain\_data$ WTvS1R2, strain\_data$ WTvS2R1, strain\_data$ WTvS2R2, strain\_data$ WTvS3R1, strain\_data$ WTvS3R2, strain\_data$pth11vB1R1, strain\_data$pth11vB1R2, strain\_data$pth11vB2R1, strain\_data$pth11vB2R2, strain\_data$pth11vB3R1, strain\_data$pth11vB3R2, strain\_data$pth11vC1R1, strain\_data$pth11vC1R2, strain\_data$pth11vC2R1, strain\_data$pth11vC2R2, strain\_data$pth11vC3R1, strain\_data$pth11vC3R2, strain\_data$pth11vS1R1, strain\_data$pth11vS1R2, strain\_data$pth11vS2R1, strain\_data$pth11vS2R2, strain\_data$pth11vS3R1, strain\_data$pth11vS3R2, sep="-")

> colnames(strain\_matrix)<-paste(strain\_data$ WTvB1R1, strain\_data$ WTvB1R2, strain\_data$ WTvB2R1, strain\_data$ WTvB2R2, strain\_data$ WTvB3R1, strain\_data$ WTvB3R2, strain\_data$ WTvC1R1, strain\_data$ WTvC1R2, strain\_data$ WTvC2R1, strain\_data$ WTvC2R2, strain\_data$ WTvC3R1, strain\_data$ WTvC3R2, strain\_data$ WTvS1R1, strain\_data$ WTvS1R2, strain\_data$ WTvS2R1, strain\_data$ WTvS2R2, strain\_data$ WTvS3R1, strain\_data$ WTvS3R2, strain\_data$pth11vB1R1, strain\_data$pth11vB1R2, strain\_data$pth11vB2R1, strain\_data$pth11vB2R2, strain\_data$pth11vB3R1, strain\_data$pth11vB3R2, strain\_data$pth11vC1R1, strain\_data$pth11vC1R2, strain\_data$pth11vC2R1, strain\_data$pth11vC2R2, strain\_data$pth11vC3R1, strain\_data$pth11vC3R2, strain\_data$pth11vS1R1, strain\_data$pth11vS1R2, strain\_data$pth11vS2R1, strain\_data$pth11vS2R2, strain\_data$pth11vS3R1, strain\_data$pth11vS3R2, sep="-")

> View(strain\_matrix)

> rownames(time\_matrix)<-paste(time\_data$ WTvB1R1, time\_data$ WTvB1R2, time\_data$ WTvB2R1, time\_data$ WTvB2R2, time\_data$ WTvB3R1, time\_data$ WTvB3R2, time\_data$ WTvC1R1, time\_data$ WTvC1R2, time\_data$ WTvC2R1, time\_data$ WTvC2R2, time\_data$ WTvC3R1, time\_data$ WTvC3R2, time\_data$ WTvS1R1, time\_data$ WTvS1R2, time\_data$ WTvS2R1, time\_data$ WTvS2R2, time\_data$ WTvS3R1, time\_data$ WTvS3R2, time\_data$pth11vB1R1, time\_data$pth11vB1R2, time\_data$pth11vB2R1, time\_data$pth11vB2R2, time\_data$pth11vB3R1, time\_data$pth11vB3R2, time\_data$pth11vC1R1, time\_data$pth11vC1R2, time\_data$pth11vC2R1, time\_data$pth11vC2R2, time\_data$pth11vC3R1, time\_data$pth11vC3R2, time\_data$pth11vS1R1, time\_data$pth11vS1R2, time\_data$pth11vS2R1, time\_data$pth11vS2R2, time\_data$pth11vS3R1, time\_data$pth11vS3R2, sep="-")

> colnames(time\_matrix)<-paste(time\_data$ WTvB1R1, time\_data$ WTvB1R2, time\_data$ WTvB2R1, time\_data$ WTvB2R2, time\_data$ WTvB3R1, time\_data$ WTvB3R2, time\_data$ WTvC1R1, time\_data$ WTvC1R2, time\_data$ WTvC2R1, time\_data$ WTvC2R2, time\_data$ WTvC3R1, time\_data$ WTvC3R2, time\_data$ WTvS1R1, time\_data$ WTvS1R2, time\_data$ WTvS2R1, time\_data$ WTvS2R2, time\_data$ WTvS3R1, time\_data$ WTvS3R2, time\_data$pth11vB1R1, time\_data$pth11vB1R2, time\_data$pth11vB2R1, time\_data$pth11vB2R2, time\_data$pth11vB3R1, time\_data$pth11vB3R2, time\_data$pth11vC1R1, time\_data$pth11vC1R2, time\_data$pth11vC2R1, time\_data$pth11vC2R2, time\_data$pth11vC3R1, time\_data$pth11vC3R2, time\_data$pth11vS1R1, time\_data$pth11vS1R2, time\_data$pth11vS2R1, time\_data$pth11vS2R2, time\_data$pth11vS3R1, time\_data$pth11vS3R2, sep="-")

> rownames(treatment\_matrix)<-paste(treatment\_data$ WTvB1R1, treatment\_data$ WTvB1R2, treatment\_data$ WTvB2R1, treatment\_data$ WTvB2R2, treatment\_data$ WTvB3R1, treatment\_data$ WTvB3R2, treatment\_data$ WTvC1R1, treatment\_data$ WTvC1R2, treatment\_data$ WTvC2R1, treatment\_data$ WTvC2R2, treatment\_data$ WTvC3R1, treatment\_data$ WTvC3R2, treatment\_data$ WTvS1R1, treatment\_data$ WTvS1R2, treatment\_data$ WTvS2R1, treatment\_data$ WTvS2R2, treatment\_data$ WTvS3R1, treatment\_data$ WTvS3R2, treatment\_data$pth11vB1R1, treatment\_data$pth11vB1R2, treatment\_data$pth11vB2R1, treatment\_data$pth11vB2R2, treatment\_data$pth11vB3R1, treatment\_data$pth11vB3R2, treatment\_data$pth11vC1R1, treatment\_data$pth11vC1R2, treatment\_data$pth11vC2R1, treatment\_data$pth11vC2R2, treatment\_data$pth11vC3R1, treatment\_data$pth11vC3R2, treatment\_data$pth11vS1R1, treatment\_data$pth11vS1R2, treatment\_data$pth11vS2R1, treatment\_data$pth11vS2R2, treatment\_data$pth11vS3R1, treatment\_data$pth11vS3R2)

> rownames(treatment\_matrix)<-paste(treatment\_data$ WTvB1R1, treatment\_data$ WTvB1R2, treatment\_data$ WTvB2R1, treatment\_data$ WTvB2R2, treatment\_data$ WTvB3R1, treatment\_data$ WTvB3R2, treatment\_data$ WTvC1R1, treatment\_data$ WTvC1R2, treatment\_data$ WTvC2R1, treatment\_data$ WTvC2R2, treatment\_data$ WTvC3R1, treatment\_data$ WTvC3R2, treatment\_data$ WTvS1R1, treatment\_data$ WTvS1R2, treatment\_data$ WTvS2R1, treatment\_data$ WTvS2R2, treatment\_data$ WTvS3R1, treatment\_data$ WTvS3R2, treatment\_data$pth11vB1R1, treatment\_data$pth11vB1R2, treatment\_data$pth11vB2R1, treatment\_data$pth11vB2R2, treatment\_data$pth11vB3R1, treatment\_data$pth11vB3R2, treatment\_data$pth11vC1R1, treatment\_data$pth11vC1R2, treatment\_data$pth11vC2R1, treatment\_data$pth11vC2R2, treatment\_data$pth11vC3R1, treatment\_data$pth11vC3R2, treatment\_data$pth11vS1R1, treatment\_data$pth11vS1R2, treatment\_data$pth11vS2R1, treatment\_data$pth11vS2R2, treatment\_data$pth11vS3R1, treatment\_data$pth11vS3R2, sep="-")

> colnames(treatment\_matrix)<-paste(treatment\_data$ WTvB1R1, treatment\_data$ WTvB1R2, treatment\_data$ WTvB2R1, treatment\_data$ WTvB2R2, treatment\_data$ WTvB3R1, treatment\_data$ WTvB3R2, treatment\_data$ WTvC1R1, treatment\_data$ WTvC1R2, treatment\_data$ WTvC2R1, treatment\_data$ WTvC2R2, treatment\_data$ WTvC3R1, treatment\_data$ WTvC3R2, treatment\_data$ WTvS1R1, treatment\_data$ WTvS1R2, treatment\_data$ WTvS2R1, treatment\_data$ WTvS2R2, treatment\_data$ WTvS3R1, treatment\_data$ WTvS3R2, treatment\_data$pth11vB1R1, treatment\_data$pth11vB1R2, treatment\_data$pth11vB2R1, treatment\_data$pth11vB2R2, treatment\_data$pth11vB3R1, treatment\_data$pth11vB3R2, treatment\_data$pth11vC1R1, treatment\_data$pth11vC1R2, treatment\_data$pth11vC2R1, treatment\_data$pth11vC2R2, treatment\_data$pth11vC3R1, treatment\_data$pth11vC3R2, treatment\_data$pth11vS1R1, treatment\_data$pth11vS1R2, treatment\_data$pth11vS2R1, treatment\_data$pth11vS2R2, treatment\_data$pth11vS3R1, treatment\_data$pth11vS3R2, sep="-")

> colors <- colorRampPalette (rev(brewer.pal(9, "Blues"))) (255)

> pheatmap(strain\_matrix, clustering\_distance\_rows=strain\_dist,clustering\_distance\_cols=strain\_dist,col=colors)

> rownames(strain\_data)<-rownames(colData(strain\_rlog))

> colnames(strain\_data)<-rownames(colData(strain\_rlog))

> pheatmap(strain\_matrix, clustering\_distance\_rows=strain\_dist,clustering\_distance\_cols=strain\_dist,col=colors)

> pheatmap(treatment\_matrix, clustering\_distance\_rows = treat\_dist, clustering\_distance\_cols = treat\_dist, col=colors)

> d <- matrix(rnorm(25), 5, 5)

> colnames(d) = paste ("one", 1:5, sep ="")

> rownames(d) = paste ("two", 1:5, sep ="")

> pheatmap(d)xwds/weseszwswew9q5816+ 336+3 123wqzsedw]9-+\*

Q

]A[

] A a-\* A-\*9-6> head(time\_rlog)

class: DESeqTransform

dim: 6 36

metadata(1): version

assays(1): ''

rownames(6): exon-MGG\_20000-1 exon-MGG\_20001-1 ... exon-MGG\_20004-1 exon-MGG\_20005-1

rowData names(7): baseMean baseVar ... dispFit rlogIntercept

colnames(36): 4091vB1R1 4091vB1R2 ... pth11vS3R1 pth11vS3R2

colData names(2): condition sizeFactor

> rownames(time\_matrix)<-paste(time\_rlog$WTvB1R1, time\_rlog$WTvB1R2, time\_rlog$WTvB2R1, time\_rlog$WTvB2R2, time\_rlog$WTvB3R1, time\_rlog$WTvB3R2, time\_rlog$WTvC1R1, time\_rlog$WTvC1R2, time\_rlog$WTvC2R1, time\_rlog$WTvC2R2, time\_rlog$WTvC3R1, time\_rlog$WTvC3R2, time\_rlog$WTvS1R1, time\_rlog$WTvS1R2, time\_rlog$WTvS2R1, time\_rlog$WTvS2R2, time\_rlog$WTvS3R1, time\_rlog$WTvS3R2, time\_rlog$pth11vB1R1, time\_rlog$pth11vB1R2, time\_rlog$pth11vB2R1, time\_rlog$pth11vB2R2, time\_rlog$pth11vB3R1, time\_rlog$pth11vB3R2, time\_rlog$pth11vC1R1, time\_rlog$pth11vC1R2, time\_rlog$pth11vC2R1, time\_rlog$pth11vC2R2, time\_rlog$pth11vC3R1, time\_rlog$pth11vC3R2, time\_rlog$pth11vS1R1, time\_rlog$pth11vS1R2, time\_rlog$pth11vS2R1, time\_rlog$pth11vS2R2, time\_rlog$pth11vS3R1, time\_rlog$pth11vS3R2, sep="-")

> colnames(time\_matrix)<-paste(time\_rlog$WTvB1R1, time\_rlog$WTvB1R2, time\_rlog$WTvB2R1, time\_rlog$WTvB2R2, time\_rlog$WTvB3R1, time\_rlog$WTvB3R2, time\_rlog$WTvC1R1, time\_rlog$WTvC1R2, time\_rlog$WTvC2R1, time\_rlog$WTvC2R2, time\_rlog$WTvC3R1, time\_rlog$WTvC3R2, time\_rlog$WTvS1R1, time\_rlog$WTvS1R2, time\_rlog$WTvS2R1, time\_rlog$WTvS2R2, time\_rlog$WTvS3R1, time\_rlog$WTvS3R2, time\_rlog$pth11vB1R1, time\_rlog$pth11vB1R2, time\_rlog$pth11vB2R1, time\_rlog$pth11vB2R2, time\_rlog$pth11vB3R1, time\_rlog$pth11vB3R2, time\_rlog$pth11vC1R1, time\_rlog$pth11vC1R2, time\_rlog$pth11vC2R1, time\_rlog$pth11vC2R2, time\_rlog$pth11vC3R1, time\_rlog$pth11vC3R2, time\_rlog$pth11vS1R1, time\_rlog$pth11vS1R2, time\_rlog$pth11vS2R1, time\_rlog$pth11vS2R2, time\_rlog$pth11vS3R1, time\_rlog$pth11vS3R2, sep="-")

> pheatmap(time\_matrix, clusering\_distance\_rows=time\_dist, clustering\_distance\_cols=time\_dist, col=colors)

> pheatmap(time\_matrix,col=colors)

> pheatmap(DistMatrix, clustering\_distance\_rows=sampleDists, clustering\_distance\_cols=sampleDists, col=colors)

> head (time\_matrix)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> rownames(treatment\_matrix)<-paste(treatment\_rlog$ WTvB1R1, treatment\_rlog$ WTvB1R2, treatment\_rlog$ WTvB2R1, treatment\_rlog$ WTvB2R2, treatment\_rlog$ WTvB3R1, treatment\_rlog$ WTvB3R2, treatment\_rlog$ WTvC1R1, treatment\_rlog$ WTvC1R2, treatment\_rlog$ WTvC2R1, treatment\_rlog$ WTvC2R2, treatment\_rlog$ WTvC3R1, treatment\_rlog$ WTvC3R2, treatment\_rlog$ WTvS1R1, treatment\_rlog$ WTvS1R2, treatment\_rlog$ WTvS2R1, treatment\_rlog$ WTvS2R2, treatment\_rlog$ WTvS3R1, treatment\_rlog$ WTvS3R2, treatment\_rlog$pth11vB1R1, treatment\_rlog$pth11vB1R2, treatment\_rlog$pth11vB2R1, treatment\_rlog$pth11vB2R2, treatment\_rlog$pth11vB3R1, treatment\_rlog$pth11vB3R2, treatment\_rlog$pth11vC1R1, treatment\_rlog$pth11vC1R2, treatment\_rlog$pth11vC2R1, treatment\_rlog$pth11vC2R2, treatment\_rlog$pth11vC3R1, treatment\_rlog$pth11vC3R2, treatment\_rlog$pth11vS1R1, treatment\_rlog$pth11vS1R2, treatment\_rlog$pth11vS2R1, treatment\_rlog$pth11vS2R2, treatment\_rlog$pth11vS3R1, treatment\_rlog$pth11vS3R2, sep="-")

Error in eval(quote(list(...)), env) : object 'treatment\_rlog' not found

> rownames(treatment\_matrix)<-paste(treat\_rlog$ WTvB1R1, treat\_rlog$ WTvB1R2, treat\_rlog$ WTvB2R1, treat\_rlog$ WTvB2R2, treat\_rlog$ WTvB3R1, treat\_rlog$ WTvB3R2, treat\_rlog$ WTvC1R1, treat\_rlog$ WTvC1R2, treat\_rlog$ WTvC2R1, treat\_rlog$ WTvC2R2, treat\_rlog$ WTvC3R1, treat\_rlog$ WTvC3R2, treat\_rlog$ WTvS1R1, treat\_rlog$ WTvS1R2, treat\_rlog$ WTvS2R1, treat\_rlog$ WTvS2R2, treat\_rlog$ WTvS3R1, treat\_rlog$ WTvS3R2, treat\_rlog$pth11vB1R1, treat\_rlog$pth11vB1R2, treat\_rlog$pth11vB2R1, treat\_rlog$pth11vB2R2, treat\_rlog$pth11vB3R1, treat\_rlog$pth11vB3R2, treat\_rlog$pth11vC1R1, treat\_rlog$pth11vC1R2, treat\_rlog$pth11vC2R1, treat\_rlog$pth11vC2R2, treat\_rlog$pth11vC3R1, treat\_rlog$pth11vC3R2, treat\_rlog$pth11vS1R1, treat\_rlog$pth11vS1R2, treat\_rlog$pth11vS2R1, treat\_rlog$pth11vS2R2, treat\_rlog$pth11vS3R1, treat\_rlog$pth11vS3R2 sep="-")

Error: unexpected symbol in "treat\_rlog$pth11vC1R2, treat\_rlog$pth11vC2R1, treat\_rlog$pth11vC2R2, treat\_rlog$pth11vC3R1, treat\_rlog$pth11vC3R2, treat\_rlog$pth11vS1R1, treat\_rlog$pth11vS1R2, treat\_rlog$pth11vS2R1, treat\_rl"

> rownames(treatment\_matrix)<-paste(treat\_rlog$ WTvB1R1, treat\_rlog$ WTvB1R2, treat\_rlog$ WTvB2R1, treat\_rlog$ WTvB2R2, treat\_rlog$ WTvB3R1, treat\_rlog$ WTvB3R2, treat\_rlog$ WTvC1R1, treat\_rlog$ WTvC1R2, treat\_rlog$ WTvC2R1, treat\_rlog$ WTvC2R2, treat\_rlog$ WTvC3R1, treat\_rlog$ WTvC3R2, treat\_rlog$ WTvS1R1, treat\_rlog$ WTvS1R2, treat\_rlog$ WTvS2R1, treat\_rlog$ WTvS2R2, treat\_rlog$ WTvS3R1, treat\_rlog$ WTvS3R2, treat\_rlog$pth11vB1R1, treat\_rlog$pth11vB1R2, treat\_rlog$pth11vB2R1, treat\_rlog$pth11vB2R2, treat\_rlog$pth11vB3R1, treat\_rlog$pth11vB3R2, treat\_rlog$pth11vC1R1, treat\_rlog$pth11vC1R2, treat\_rlog$pth11vC2R1, treat\_rlog$pth11vC2R2, treat\_rlog$pth11vC3R1, treat\_rlog$pth11vC3R2, treat\_rlog$pth11vS1R1, treat\_rlog$pth11vS1R2, treat\_rlog$pth11vS2R1, treat\_rlog$pth11vS2R2, treat\_rlog$pth11vS3R1, treat\_rlog$pth11vS3R2, sep="-")

> colnames(treatment\_matrix)<-paste(treat\_rlog$ WTvB1R1, treat\_rlog$ WTvB1R2, treat\_rlog$ WTvB2R1, treat\_rlog$ WTvB2R2, treat\_rlog$ WTvB3R1, treat\_rlog$ WTvB3R2, treat\_rlog$ WTvC1R1, treat\_rlog$ WTvC1R2, treat\_rlog$ WTvC2R1, treat\_rlog$ WTvC2R2, treat\_rlog$ WTvC3R1, treat\_rlog$ WTvC3R2, treat\_rlog$ WTvS1R1, treat\_rlog$ WTvS1R2, treat\_rlog$ WTvS2R1, treat\_rlog$ WTvS2R2, treat\_rlog$ WTvS3R1, treat\_rlog$ WTvS3R2, treat\_rlog$pth11vB1R1, treat\_rlog$pth11vB1R2, treat\_rlog$pth11vB2R1, treat\_rlog$pth11vB2R2, treat\_rlog$pth11vB3R1, treat\_rlog$pth11vB3R2, treat\_rlog$pth11vC1R1, treat\_rlog$pth11vC1R2, treat\_rlog$pth11vC2R1, treat\_rlog$pth11vC2R2, treat\_rlog$pth11vC3R1, treat\_rlog$pth11vC3R2, treat\_rlog$pth11vS1R1, treat\_rlog$pth11vS1R2, treat\_rlog$pth11vS2R1, treat\_rlog$pth11vS2R2, treat\_rlog$pth11vS3R1, treat\_rlog$pth11vS3R2, sep="-")

> pheatmap(treatment\_matrix, clustering\_distance\_rows=treat\_dists, clustering\_distance\_cols =treat\_dists, col=colors)

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> strain\_dist<-dist(t(assay(strain\_rlog)))

> strain\_matrix<-as.matrix(strain\_dist)

> pheatmap(strain\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_dist)

> pheatmap(time\_dist, col=colors)

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_data, intgroup="condition")

Error in (function (classes, fdef, mtable) :

unable to find an inherited method for function ‘plotPCA’ for signature ‘"DESeqDataSet"’

> plotPCA(time\_rlog, intgroup="condition")

> plotPCA(treat\_rlog, intgroup="condition")

> plotPCA(strain\_rlog, intgroup="condition")

> treatVars<-rowVars(assay(treat\_rlog))

> treatVarsOrdered<-order(treatVars, decreasing=TRUE)

> treatTopVar<-order(treatVarsOrdered, 50)

Error in order(...) : argument lengths differ

> treatTopVar<-head(treatVarsOrdered, 50)

> matrix<-assay(treat\_rlog)[treatTopVar, ]

> treat\_matrix<-assay(treat\_rlog)[treatTopVar, ]

> treat\_matrix<-treat\_matrix-rowmeans(treat\_matrix)

Error in rowmeans(treat\_matrix) : could not find function "rowmeans"

> treat\_matrix<-treat\_matrix-rowMeans(treat\_matrix)

> treat\_df<-as.data.frame(colData(treat\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(treat\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(treat\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> timeVars<-rowVars(assay(time\_rlog))

> timeVarsOrdered<-order(timeVars, decreasing=TRUE

+ )

> timeVarsOrdered<-order(timeVars, decreasing=TRUE)

> timeTopVar<-head(timeVarsOrdered, 50)

> matrix<-assay(time\_rlog)[timeTopVar, ]

> time\_matrix<-assay(time\_rlog)[timeTopVar, ]

> time\_matrix<-time\_matrix-rowMeans(time\_matrix)

> time\_df<-as.data.frame(colData(time\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(time\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(time\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> strainVars<-rowVars(assay(strain\_rlog))

> strainVarsOrdered<-order(strainVars, decreasing=TRUE)

> strainTopVar<-head(strainVarsOrdered, 50)

> strain\_matrix<-assay(strain\_rlog)[strainTopVar, ]

> strain\_matrix<-strain\_matrix-rowMeans(strain\_matrix)

> strain\_df<-as.data.frame(colData(strain\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(strain\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(strain\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> counts\_4091<-c("4091\_B\_R1\_val1\_4091vB1R1.count.txt","4091\_B\_R2\_val2\_4091vB1R2.count.txt","4091\_B\_R1\_val1\_4091vB2R1.count.txt","4091\_B\_R2\_val2\_4091vB2R2.count.txt","4091\_B\_R1\_val1\_4091vB3R1.count.txt","4091\_B\_R2\_val2\_4091vB3R2.count.txt","4091\_C3\_R1\_val1\_4091vC1R1.count.txt","4091\_C3\_R2\_val2\_4091vC1R2.count.txt","4091\_C3\_R1\_val1\_4091vC2R1.count.txt","4091\_C3\_R2\_val2\_4091vC2R2.count.txt","4091\_C3\_R1\_val1\_4091vC3R1.count.txt","4091\_C3\_R2\_val2\_4091vC3R2.count.txt","4091\_S4\_R1\_val1\_4091vS1R1.count.txt","4091\_S4\_R2\_val2\_4091vS1R2.count.txt","4091\_S4\_R1\_val1\_4091vS2R1.count.txt","4091\_S4\_R2\_val2\_4091vS2R2.count.txt","4091\_S4\_R1\_val1\_4091vS3R1.count.txt","4091\_S4\_R2\_val2\_4091vS3R2.count.txt")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c(time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

+ )

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("4091vB1R1","4091vB1R2","4091vB2R1","4091vB2R2","4091vB3R1","4091vB3R2","4091vC1R1","4091vC1R2","4091vC2R1","4091vC2R2","4091vC3R1","4091vC3R2","4091vS1R1","4091vS1R2","4091vS2R1","4091vS2R2","4091vS3R1","4091vS3R2")

> > treatment\_table<-data.frame(sampleName=samples, fileName=counts\_4091, condition=treatment)

Error: unexpected '>' in ">"

> treatment\_table<-data.frame(sampleName=samples, fileName=counts\_4091, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=counts\_4091, condition=time)

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> time\_rlog<- rlog(time\_data, blind=FALSE)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_matrix, col=colors)

>

> plotPCA(time\_rlog, intgroup="condition")

> plotPCA(treat\_rlog, intgroup="condition")

> treatVars<-rowVars(assay(treat\_rlog))

> treatVarsOrdered<-order(treatVars, decreasing=TRUE)

> treatTopVar<-head(treatVarsOrdered, 50)

> treat\_matrix<-assay(treat\_rlog)[treatTopVar, ]

> treat\_matrix<-treat\_matrix-rowMeans(treat\_matrix)

> treat\_df<-as.data.frame(colData(treat\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(treat\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(treat\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

> > timeVars<-rowVars(assay(time\_rlog))

Error: unexpected '>' in ">"

> timeVars<-rowVars(assay(time\_rlog))

> timeVarsOrdered<-order(timeVars, decreasing=TRUE)

> timeTopVar<-head(timeVarsOrdered, 50)

> time\_matrix<-assay(time\_rlog)[timeTopVar, ]

> time\_matrix<-time\_matrix-rowMeans(time\_matrix)

> time\_df<-as.data.frame(colData(time\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(time\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(time\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

>

> counts\_pth11<-c("pth11\_B\_R1\_val1\_pth11vB1R1.count.txt","pth11\_B\_R2\_val2\_pth11vB1R2.count.txt","pth11\_B\_R1\_val1\_pth11vB2R1.count.txt","pth11\_B\_R2\_val2\_pth11vB2R2.count.txt","pth11\_B\_R1\_val1\_pth11vB3R1.count.txt","pth11\_B\_R2\_val2\_pth11vB3R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC1R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC1R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC2R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC2R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC3R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC3R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS1R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS1R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS2R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS2R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS3R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS3R2.count.txt")

> treatment<-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

> time<-c("0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi","0\_hpi","0\_hpi","3\_hpi","3\_hpi","9\_hpi","9\_hpi")

> samples<-c("pth11vB1R1","pth11vB1R2","pth11vB2R1","pth11vB2R2","pth11vB3R1","pth11vB3R2","pth11vC1R1","pth11vC1R2","pth11vC2R1","pth11vC2R2","pth11vC3R1","pth11vC3R2","pth11vS1R1","pth11vS1R2","pth11vS2R1","pth11vS2R2","pth11vS3R1","pth11vS3R2")

> > treatment\_table<-data.frame(sampleName=samples, fileName=counts\_pth11, condition=treatment)

Error: unexpected '>' in ">"

> treatment\_table<-data.frame(sampleName=samples, fileName=counts\_pth11, condition=treatment)

> time\_table<-data.frame(sampleName=samples, fileName=counts\_4pth11, condition=time)

Error in data.frame(sampleName = samples, fileName = counts\_4pth11, condition = time) :

object 'counts\_4pth11' not found

> time\_table<-data.frame(sampleName=samples, fileName=counts\_pth11, condition=time)

> treatment\_data<-DESeqDataSetFromHTSeqCount(sampleTable=treatment\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> time\_data<-DESeqDataSetFromHTSeqCount(sampleTable=time\_table, design= ~ condition)

Warning message:

In DESeqDataSet(se, design = design, ignoreRank) :

some variables in design formula are characters, converting to factors

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> time\_rlog<- rlog(time\_data, blind=FALSE)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_rlog, intgroup="condition")

> plotPCA(treat\_rlog, intgroup="condition")

>

>

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> treat\_rlog<- rlog(treatment\_data, blind=FALSE)

> time\_rlog<- rlog(time\_data, blind=FALSE)

> treat\_dists<-dist(t(assay(treat\_rlog)))

> treatment\_matrix<-as.matrix(treat\_dists)

> pheatmap(treatment\_matrix, col=colors)

> time\_dist<-dist(t(assay(time\_rlog)))

> time\_matrix<-as.matrix(time\_dist)

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_rlog, intgroup="condition")

> pheatmap(time\_matrix, col=colors)

> plotPCA(time\_rlog, intgroup="condition")

> pheatmap(treatment\_matrix, col=colors)

> pheatmap(time\_matrix, col=colors)

> plotPCA(treat\_rlog, intgroup="condition")

> > treatVars<-rowVars(assay(treat\_rlog))

Error: unexpected '>' in ">"

> treatVars<-rowVars(assay(treat\_rlog))

> treatVarsOrdered<-order(treatVars, decreasing=TRUE)

> treatTopVar<-head(treatVarsOrdered, 50)

> treat\_matrix<-assay(treat\_rlog)[treatTopVar, ]

> treat\_matrix<-treat\_matrix-rowMeans(treat\_matrix)

> treat\_df<-as.data.frame(colData(treat\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(treat\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(treat\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

>

> timeVars<-rowVars(assay(time\_rlog))

> timeVarsOrdered<-order(timeVars, decreasing=TRUE)

> timeTopVar<-head(timeVarsOrdered, 50)

> time\_matrix<-assay(time\_rlog)[timeTopVar, ]

> time\_matrix<-time\_matrix-rowMeans(time\_matrix)

> time\_df<-as.data.frame(colData(time\_rlog)[,c("condition"),drop=FALSE])

> clear\_col\_names<-paste(time\_rlog$condition,sep=".")

> topGenesHeatmap <- pheatmap(time\_matrix, annotation\_col=treat\_df, labels\_col=clear\_col\_names)

library(edgeR)

library(statmod)

if (!require("BiocManager", quietly = TRUE))

install.packages("BiocManager")

BiocManager::install("edgeR")

library(edgeR)

if (!require("BiocManager", quietly = TRUE))

install.packages("BiocManager")

BiocManager::install("statmod")

library(statmod)

library(statmod)

edgeR

files <-c("4091\_B\_R1\_val1\_4091vB1R1.count.txt","4091\_B\_R2\_val2\_4091vB1R2.count.txt","4091\_B\_R1\_val1\_4091vB2R1.count.txt","4091\_B\_R2\_val2\_4091vB2R2.count.txt","4091\_B\_R1\_val1\_4091vB3R1.count.txt","4091\_B\_R2\_val2\_4091vB3R2.count.txt","4091\_C3\_R1\_val1\_4091vC1R1.count.txt","4091\_C3\_R2\_val2\_4091vC1R2.count.txt","4091\_C3\_R1\_val1\_4091vC2R1.count.txt","4091\_C3\_R2\_val2\_4091vC2R2.count.txt","4091\_C3\_R1\_val1\_4091vC3R1.count.txt","4091\_C3\_R2\_val2\_4091vC3R2.count.txt","4091\_S4\_R1\_val1\_4091vS1R1.count.txt","4091\_S4\_R2\_val2\_4091vS1R2.count.txt","4091\_S4\_R1\_val1\_4091vS2R1.count.txt","4091\_S4\_R2\_val2\_4091vS2R2.count.txt","4091\_S4\_R1\_val1\_4091vS3R1.count.txt","4091\_S4\_R2\_val2\_4091vS3R2.count.txt","pth11\_B\_R1\_val1\_pth11vB1R1.count.txt","pth11\_B\_R2\_val2\_pth11vB1R2.count.txt","pth11\_B\_R1\_val1\_pth11vB2R1.count.txt","pth11\_B\_R2\_val2\_pth11vB2R2.count.txt","pth11\_B\_R1\_val1\_pth11vB3R1.count.txt","pth11\_B\_R2\_val2\_pth11vB3R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC1R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC1R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC2R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC2R2.count.txt","pth11\_C3\_R1\_val1\_pth11vC3R1.count.txt","pth11\_C3\_R2\_val2\_pth11vC3R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS1R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS1R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS2R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS2R2.count.txt","pth11\_S4\_R1\_val1\_pth11vS3R1.count.txt","pth11\_S4\_R2\_val2\_pth11vS3R2.count.txt")

strains <-c("4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","4091","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11","pth11")

treatment <-c("Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4","Buffer","Buffer","Buffer","Buffer","Buffer","Buffer","C3","C3","C3","C3","C3","C3","S4","S4","S4","S4","S4","S4")

samples <-c("4091vB1R1","4091vB1R2","4091vB2R1","4091vB2R2","4091vB3R1","4091vB3R2","4091vC1R1","4091vC1R2","4091vC2R1","4091vC2R2","4091vC3R1","4091vC3R2","4091vS1R1","4091vS1R2","4091vS2R1","4091vS2R2","4091vS3R1","4091vS3R2","pth11vB1R1","pth11vB1R2","pth11vB2R1","pth11vB2R2","pth11vB3R1","pth11vB3R2","pth11vC1R1","pth11vC1R2","pth11vC2R1","pth11vC2R2","pth11vC3R1","pth11vC3R2","pth11vS1R1","pth11vS1R2","pth11vS2R1","pth11vS2R2","pth11vS3R1","pth11vS3R2")

table <-data.frame(sampleName=samples, fileName=files, condition=treatment)

DGE <-readDGE(files, group=treatment, labels=samples)

DGE <-calcNormFactors(DGE)

DGE$samples

plotMDS(DGE)

DGE <- extimateDisp(DGE)

DGE <- estimateDisp(DGE)

plotBCV(DGE)

exact\_test <- exactTest(DGE, pair="Buffer","S4")

exact\_test <- exactTest(DGE, pair=c("Buffer","S4"))

summary <- decideTestsDGE(exact\_test)

summary

DEtags <- rownames(DGE)[as.logical(DE)]

DEtags <- rownames(DGE)[as.logical(de)]

plotSmear(exact\_test, de.tags=DEtags)

DEtags <- rownames(DGE)

plotSmear(exact\_test, de.tags=DEtags)

abline(h=c(-1,1),col="blue")

plotSmear(exact\_test, de.tags=DEtags)

abline(h=c(-1,1),col="blue")

DEgenes <- topTags(exact\_test, n=1000, p.value=0.05)

head(diffExpGenes$table)

head(DEgenes$table)

write.table(DEgenes$table, file="B\_v\_S4\_exactTest.txt",sep="\t",row.names=TRUE, col.names=NA)