# Gene expression analysis

Contents

[Gene expression analysis 1](#_Toc533410699)

[Gene expression analysis 2](#_Toc533410700)

[1. Calculate gene length 2](#_Toc533410701)

[2. calculateFPKM 2](#_Toc533410702)

[3. Collect expressed genes of each treatment 2](#_Toc533410703)

[4. Differential analysis 4](#_Toc533410704)

[5. Calculate fold change 5](#_Toc533410705)

[6. Compare DEGs of different treatments 7](#_Toc533410706)

[7. Generate files for clustering 8](#_Toc533410707)

[8. GO analysis 8](#_Toc533410708)

# Gene expression analysis

## Calculate gene length

#R code

#calculate the length of genes (sum of the exon length for each gene gene)

#import the GTF-file that was used for htseq-count

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Reference/Curated/"**)**

library**(**GenomicFeatures**)**

txdb **<-** makeTxDbFromGFF**(**"V2.1\_iso1\_exon.gtf",format**=**"gtf"**)**

#collect the exons per gene id

exons.list.per.gene **<-** exonsBy**(**txdb,by**=**"gene"**)**

#for each gene, reduce all the exons to a set of non-overlapping exons, calculate their lengths (widths) and sum then

exonic.gene.sizes **<-** lapply**(**exons.list.per.gene,**function(**x**){**sum**(**width**(**reduce**(**x**)))})**

df**=**unlist**(**exonic.gene.sizes**)**

data**=**as.data.frame**(**df**)**

data**$**id**=**rownames**(**data**)**

data**=**data**[**,c**(**2,1**)]**

colnames**(**data**)=**c**(**"id","length"**)**

write.table**(**data, "V2.1\_iso1\_genelength.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## calculateFPKM

#R code

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/"**)**

count**=**read.table**(**"k100Mm\_GeneV2\_counttable.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

count**=**count**[!**grepl**(**"\_\_", row.names**(**count**))**,**]**

count**$**id**=**rownames**(**count**)**

rownames**(**count**)=NULL**

tmr**=**read.table**(**"../../../k100Mm/TotalMappedRead/TotalMappedRead.txt", header**=**T, sep**=**"\t"**)**

tmr**=**tmr**[**1,**]**

tmr**=**as.numeric**(**tmr**)**

length**=**read.table**(**"../../../Reference/Curated/V2.1\_iso1\_genelength.txt", header**=**T, sep**=**"\t"**)**

length**$**id\_temp**=**paste**(**length**$**id, ".1", sep**=**""**)**

count**=**merge**(**count, length, by.x**=**"id", by.y**=**"id\_temp"**)**

temp**=**data.frame**(**id**=**count**[**,1**])**

**for** **(**i **in** 1**:**39**){**

TMR**=**tmr**[**i**]**

xi**=(**count**[**,1**+**i**]\***10**^**9**)/(**TMR**\***count**[**,42**])**

temp**=**cbind**(**temp,xi**)**

**}**

colnames**(**temp**)=**c**(**"id","fpkm01","fpkm02","fpkm03","fpkm04","fpkm05","fpkm06","fpkm07","fpkm08","fpkm09","fpkm10","fpkm11","fpkm12","fpkm13","fpkm14","fpkm15","fpkm16","fpkm17","fpkm18","fpkm19","fpkm20","fpkm21","fpkm22","fpkm23","fpkm24","fpkm25","fpkm26","fpkm27","fpkm28","fpkm29","fpkm30","fpkm31","fpkm32","fpkm33","fpkm34","fpkm35","fpkm36","fpkm37","fpkm38","fpkm39"**)**

fpkm**=**cbind**(**id**=**count**[**,41**]**, temp**[**,**-**1**])**

write.table**(**fpkm, "k100Mm\_htseq\_geneFPKM.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Collect expressed genes of each treatment

#R code

setwd**(**"**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/"**)**

gene**=**read.table**(**'k100Mm\_htseq\_geneFPKM.txt', header**=**T, sep**=**"\t"**)**

#ctrl

#using fpkm

gene**=**gene**[**,1**:**4**]**

gene**$**mean**=**rowMeans**(**gene**[**,2**:**4**])**

expr**=**subset**(**gene, gene**$**mean**>**1**)**

noexpr**=**subset**(**gene, gene**$**mean**<=**1**)**

expr**$**FPKM\_tag**=**"expr"

noexpr**$**FPKM\_tag**=**"noexpr"

df1**=**rbind**(**expr, noexpr**)**

df1**=**df1**[**order**(**df1**$**id**)**,**]**

df1**$**temp\_id**=**paste**(**df1**$**id, ".1", sep**=**""**)**

#using raw count

gene**=**read.table**(**'k100Mm\_GeneV2\_counttable.txt', header**=**T, row.names**=**1**)**

gene**=**gene**[!**grepl**(**"\_\_", row.names**(**gene**))**,**]**

gene**$**temp\_id**=**row.names**(**gene**)**

gene**=**gene**[**,c**(**40,1**:**3**)]** #ctrl

gene**$**mean**=**rowMeans**(**gene**[**,2**:**4**])**

expr**=**subset**(**gene, gene**$**mean**>**10**)**

noexpr**=**subset**(**gene, gene**$**mean**<=**10**)**

expr**$**RowCount\_tag**=**"expr"

noexpr**$**RowCount\_tag**=**"noexpr"

df2**=**rbind**(**expr, noexpr**)**

df2**=**df2**[**order**(**df2**$**temp\_id**)**,**]**

df**=**merge**(**df1,df2, by**=**"temp\_id"**)**

write.table**(**df, "k100Mm\_Gene\_Ctrl\_exprTag.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#mock

#using fpkm

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount"**)**

gene**=**read.table**(**'k100Mm\_htseq\_geneFPKM.txt', header**=**T, sep**=**"\t"**)**

gene**=**gene**[**,c**(**1,5**:**16**)]** #mock

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minFPKM **=** 1

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minFPKM,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minFPKM,**]**

expr**$**FPKM\_tag**=**"expr"

noexpr**$**FPKM\_tag**=**"noexpr"

df1**=**rbind**(**expr, noexpr**)**

df1**=**df1**[**order**(**df1**$**id**)**,**]**

df1**$**temp\_id**=**paste**(**df1**$**id, ".1", sep**=**""**)**

#using raw count

gene**=**read.table**(**'k100Mm\_GeneV2\_counttable.txt', header**=**T, row.names**=**1**)**

gene**=**gene**[!**grepl**(**"\_\_", row.names**(**gene**))**,**]**

gene**$**temp\_id**=**row.names**(**gene**)**

gene**=**gene**[**,c**(**40,4**:**15**)]** #mock

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minRead **=** 10

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minRead,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minRead,**]**

expr**$**RowCount\_tag**=**"expr"

noexpr**$**RowCount\_tag**=**"noexpr"

df2**=**rbind**(**expr, noexpr**)**

df2**=**df2**[**order**(**df2**$**temp\_id**)**,**]**

df**=**merge**(**df1,df2, by**=**"temp\_id"**)**

write.table**(**df, "k100Mm\_Gene\_Mock\_exprTag.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#yeast

#using fpkm

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount"**)**

gene**=**read.table**(**'k100Mm\_htseq\_geneFPKM.txt', header**=**T, sep**=**"\t"**)**

gene**=**gene**[**,c**(**1,17**:**28**)]** #yeast

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minFPKM **=** 1

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minFPKM,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minFPKM,**]**

expr**$**FPKM\_tag**=**"expr"

noexpr**$**FPKM\_tag**=**"noexpr"

df1**=**rbind**(**expr, noexpr**)**

df1**=**df1**[**order**(**df1**$**id**)**,**]**

df1**$**temp\_id**=**paste**(**df1**$**id, ".1", sep**=**""**)**

#using raw count

gene**=**read.table**(**'k100Mm\_GeneV2\_counttable.txt', header**=**T, row.names**=**1**)**

gene**=**gene**[!**grepl**(**"\_\_", row.names**(**gene**))**,**]**

gene**$**temp\_id**=**row.names**(**gene**)**

gene**=**gene**[**,c**(**40,16**:**27**)]** #yeast

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minRead **=** 10

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minRead,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minRead,**]**

expr**$**RowCount\_tag**=**"expr"

noexpr**$**RowCount\_tag**=**"noexpr"

df2**=**rbind**(**expr, noexpr**)**

df2**=**df2**[**order**(**df2**$**temp\_id**)**,**]**

df**=**merge**(**df1,df2, by**=**"temp\_id"**)**

write.table**(**df, "k100Mm\_Gene\_Yeast\_exprTag.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#botrytis

#using fpkm

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount"**)**

gene**=**read.table**(**'k100Mm\_htseq\_geneFPKM.txt', header**=**T, sep**=**"\t"**)**

gene**=**gene**[**,c**(**1,29**:**40**)]** #botrytis

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minFPKM **=** 1

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minFPKM,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minFPKM,**]**

expr**$**FPKM\_tag**=**"expr"

noexpr**$**FPKM\_tag**=**"noexpr"

df1**=**rbind**(**expr, noexpr**)**

df1**=**df1**[**order**(**df1**$**id**)**,**]**

df1**$**temp\_id**=**paste**(**df1**$**id, ".1", sep**=**""**)**

#using raw count

gene**=**read.table**(**'k100Mm\_GeneV2\_counttable.txt', header**=**T, row.names**=**1**)**

gene**=**gene**[!**grepl**(**"\_\_", row.names**(**gene**))**,**]**

gene**$**temp\_id**=**row.names**(**gene**)**

gene**=**gene**[**,c**(**40,28**:**39**)]** #botrytis

gene**$**mean01**=**rowMeans**(**gene**[**,2**:**4**])**

gene**$**mean03**=**rowMeans**(**gene**[**,5**:**7**])**

gene**$**mean06**=**rowMeans**(**gene**[**,8**:**10**])**

gene**$**mean12**=**rowMeans**(**gene**[**,11**:**13**])**

minRead **=** 10

expr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **>** minRead,**]**

noexpr **<-** gene**[**apply**(**gene**[**,14**:**17**]**,1,**function(**x**){**max**(**x**)})** **<=** minRead,**]**

expr**$**RowCount\_tag**=**"expr"

noexpr**$**RowCount\_tag**=**"noexpr"

df2**=**rbind**(**expr, noexpr**)**

df2**=**df2**[**order**(**df2**$**temp\_id**)**,**]**

df**=**merge**(**df1,df2, by**=**"temp\_id"**)**

write.table**(**df, "k100Mm\_Gene\_Botrytis\_exprTag.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Differential analysis

library**(**DESeq2**)**

#Genes

#select expr Genes from htseq-count data

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/"**)**

data **=** read.table**(**'k100Mm\_GeneV2\_counttable.txt', header**=**T, row.names**=**1**)**

data**=**data**[!**grepl**(**"\_\_", row.names**(**data**))**,**]**

c**=**read.table**(**"k100Mm\_Gene\_Ctrl\_exprTag.txt", header**=**T, sep**=**"\t"**)**

m**=**read.table**(**"k100Mm\_Gene\_Mock\_exprTag.txt", header**=**T, sep**=**"\t"**)**

y**=**read.table**(**"k100Mm\_Gene\_Yeast\_exprTag.txt", header**=**T, sep**=**"\t"**)**

b**=**read.table**(**"k100Mm\_Gene\_Botrytis\_exprTag.txt", header**=**T, sep**=**"\t"**)**

c**=**subset**(**c, c**$**FPKM\_tag**==**"expr"**)**

m**=**subset**(**m, m**$**FPKM\_tag**==**"expr"**)**

y**=**subset**(**y, y**$**FPKM\_tag**==**"expr"**)**

b**=**subset**(**b, b**$**FPKM\_tag**==**"expr"**)**

c**=**data.frame**(**id**=**c**[**,1**])**

m**=**data.frame**(**id**=**m**[**,1**])**

y**=**data.frame**(**id**=**y**[**,1**])**

b**=**data.frame**(**id**=**b**[**,1**])**

temp**=**merge**(**c, m, by**=**"id", all**=TRUE)**

temp**=**merge**(**temp, y, by**=**"id", all**=TRUE)**

temp**=**merge**(**temp, b, by**=**"id", all**=TRUE)** #nrow=17462, expr. genes across all treatments

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/DESeq2\_Gene\_htseq\_FPKMover1/"**)**

data**$**id**=**rownames**(**data**)**

data**=**merge**(**data, temp, by**=**"id"**)**

rownames**(**data**)=**data**$**id

data**=**data**[**,**-**1**]**

groups **=** factor**(**c**(**rep**(**"Control",3**)**, rep**(**"Mock\_01hr",3**)**, rep**(**"Mock\_03hr",3**)**, rep**(**"Mock\_06hr",3**)**, rep**(**"Mock\_12hr",3**)**, rep**(**"Yeast\_01hr",3**)**, rep**(**"Yeast\_03hr",3**)**, rep**(**"Yeast\_06hr",3**)**, rep**(**"Yeast\_12hr",3**)**, rep**(**"Botrytis\_01hr",3**)**, rep**(**"Botrytis\_03hr",3**)**, rep**(**"Botrytis\_06hr",3**)**, rep**(**"Botrytis\_12hr",3**)))**

times **=** factor**(**c**(**rep**(**"00",3**)**, rep**(**"01",3**)**, rep**(**"03",3**)**, rep**(**"06",3**)**, rep**(**"12",3**)**, rep**(**"01",3**)**, rep**(**"03",3**)**, rep**(**"06",3**)**, rep**(**"12",3**)**, rep**(**"01",3**)**, rep**(**"03",3**)**, rep**(**"06",3**)**, rep**(**"12",3**)))**

treatments **=** factor**(**c**(**rep**(**"Mock",15**)**, rep**(**"Yeast",12**)**, rep**(**"Botrytis",12**)))**

colData **=** data.frame**(**times **=** times, treatments **=** treatments, row.names**=**colnames**(**data**))**

colData**$**treatments **<-** relevel**(**colData**$**treatments, "Mock"**)**

cdsFull **<-** DESeqDataSetFromMatrix**(**countData **=** data, colData **=** colData, design **=** **~** 1**)**

#set design formula (use model.matrix) and run DESeq2

mm **<-** model.matrix**(~**times **+** treatments**:**times, colData**(**cdsFull**))**

mm.full **<-** mm**[**,**-**c**(**6,11**)]**

mm.reduced **<-** mm.full**[**,**-**c**(**6**:**13**)]**

cdsFull **<-** DESeq**(**cdsFull, full**=**mm.full, reduced**=**mm.reduced, test**=**"LRT"**)**

res **<-** results**(**cdsFull**)**

head**(**res**[**order**(**res**$**padj**)**,**]**, 4**)**

resSig **=** res**[**which**(**res**$**padj **<** 0.05**)**,**]**

resSigdf**=**as.data.frame**(**resSig**)**

resSigdf**=**resSigdf**[**,**-**2**]**

write.table**(**resSigdf, file**=**"k100Mm\_DESeq2\_SigGeneRes.txt", sep**=**"\t", quote**=**F, row.names**=**T, col.names**=**T**)**

#estimate size factors (normalise)

cdsFull **<-** estimateSizeFactors**(**cdsFull**)**

#Write Count tables (all features)

normGeneCounts **=** counts**(**cdsFull,normalized**=**T**)**

write.table**(**normGeneCounts, file**=**"k100Mm\_DESeq2NormCounts\_Gene.txt",sep**=**"\t", quote**=**F, row.names**=**T, col.names**=**T**)**

#Logarithm transformation

#varianceStabilizingTransformation

vsd **<-** varianceStabilizingTransformation**(**cdsFull, blind**=FALSE)**

head**(**assay**(**vsd**)**,6**)** #take a look

tail**(**assay**(**vsd**)**,6**)**

vsddf **<-** assay**(**vsd**)**

write.table**(**vsddf, file**=**"k100Mm\_DESeq2NormCounts\_VST\_Gene.txt",sep**=**"\t", quote**=**F, row.names**=**T, col.names**=**T**)**

## Calculate fold change

#use vst normalized readcount (tables SenseGenerated by argument "normTransform" in DESeq)

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/DESeq2\_Gene\_htseq\_FPKMover1/"**)**

ref**<-**read.table**(**"k100Mm\_DESeq2NormCounts\_VST\_Gene.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

#log2FC to Ctrl

#mock

data**<-**read.table**(**"k100Mm\_DESeq2\_SigGeneRes.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

ref**=**cbind**(**id**=**rownames**(**ref**)**,ref**)**

SenseGenep**=**data.frame**(**id**=**rownames**(**data**))**

df**=**merge**(**SenseGenep,ref, by**=**"id"**)**

rownames**(**df**)=**df**[**,1**]**

df**=**df**[**,**-**1**]**

df**=(**df**[**,**-**c**(**1**:**3**)]-**rowMeans**(**df**[**,1**:**3**]))**

df**<-**df**[**,c**(**1**:**12**)]**

dfmean**<-**data.frame**(**x01h**=**rowMeans**(**df**[**,c**(**1**:**3**)])**, x03h**=**rowMeans**(**df**[**,c**(**4**:**6**)])**, x06h**=**rowMeans**(**df**[**,c**(**7**:**9**)])**, x12h**=**rowMeans**(**df**[**,c**(**10**:**12**)]))**

min\_FC **=** 1

dfmean **<-** dfmean**[**apply**(**dfmean,1,**function(**x**){**max**(**abs**(**x**))})** **>** min\_FC,**]**

df **<-** df**[**apply**(**dfmean,1,**function(**x**){**max**(**abs**(**x**))})** **>** min\_FC,**]**

dfmean**=**cbind**(**id**=**rownames**(**dfmean**)**, dfmean**)**

df**=**cbind**(**id**=**rownames**(**df**)**, df**)**

df**=**merge**(**df, dfmean, by**=**"id"**)**

df**=**df**[**,**-**c**(**14**:**17**)]**

exprC**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprCtrl05.txt", header**=**T, sep**=**"\t"**)**

exprM**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprMock05.txt", header**=**T, sep**=**"\t"**)**

exprC**=**data.frame**(**id**=**exprC**[**,1**])**

exprM**=**data.frame**(**id**=**exprM**[**,1**])**

expr**=**merge**(**exprC, exprM, by**=**"id", all**=TRUE)**

diff.id**<-(!**df**$**id %in% expr**$**id**)**

temp**=**df**[**diff.id,**]** #nrow=0, means all DE TE found in Mock are either expr. candidates of Ctrl or Mock

write.table**(**dfmean, "k100Mmlog2VSTFCtoCtrl\_mean\_DESeq2\_sigMockGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**df, "k100Mmlog2VSTFCtoCtrl\_DESeq2\_sigMockGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#count log2FC to Mock

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/DESeq2\_Gene\_htseq\_FPKMover1/"**)**

y **=** read.table**(**"k100Mm\_DESeq2\_SigGeneRes.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

b **=** read.table**(**"k100Mm\_DESeq2\_SigGeneRes.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

ref**<-**read.table**(**"k100Mm\_DESeq2NormCounts\_VST\_Gene.txt", header**=**T, row.names**=**1, sep**=**"\t"**)**

ref**=**cbind**(**id**=**rownames**(**ref**)**,ref**)**

SenseGenep**=**data.frame**(**id**=**rownames**(**y**))**

df**=**merge**(**SenseGenep,ref, by**=**"id"**)**

rownames**(**df**)=**df**[**,1**]**

df**=**df**[**,**-**1**]**

yFC**=**data.frame**(**id**=**rownames**(**df**)**,

Yeast\_1\_a**=**df**[**,16**]-**rowMeans**(**df**[**,4**:**6**])**,

Yeast\_1\_b**=**df**[**,17**]-**rowMeans**(**df**[**,4**:**6**])**,

Yeast\_1\_c**=**df**[**,18**]-**rowMeans**(**df**[**,4**:**6**])**,

Yeast\_3\_a**=**df**[**,19**]-**rowMeans**(**df**[**,7**:**9**])**,

Yeast\_3\_b**=**df**[**,20**]-**rowMeans**(**df**[**,7**:**9**])**,

Yeast\_3\_c**=**df**[**,21**]-**rowMeans**(**df**[**,7**:**9**])**,

Yeast\_6\_a**=**df**[**,22**]-**rowMeans**(**df**[**,10**:**12**])**,

Yeast\_6\_b**=**df**[**,23**]-**rowMeans**(**df**[**,10**:**12**])**,

Yeast\_6\_c**=**df**[**,24**]-**rowMeans**(**df**[**,10**:**12**])**,

Yeast\_12\_a**=**df**[**,25**]-**rowMeans**(**df**[**,13**:**15**])**,

Yeast\_12\_b**=**df**[**,26**]-**rowMeans**(**df**[**,13**:**15**])**,

Yeast\_12\_c**=**df**[**,27**]-**rowMeans**(**df**[**,13**:**15**]))**

SenseGenep**=**data.frame**(**id**=**rownames**(**b**))**

df**=**merge**(**SenseGenep,ref, by**=**"id"**)**

rownames**(**df**)=**df**[**,1**]**

df**=**df**[**,**-**1**]**

bFC**=**data.frame**(**id**=**rownames**(**b**)**,

Botrytis\_1\_a**=**df**[**,28**]-**rowMeans**(**df**[**,4**:**6**])**,

Botrytis\_1\_b**=**df**[**,29**]-**rowMeans**(**df**[**,4**:**6**])**,

Botrytis\_1\_c**=**df**[**,30**]-**rowMeans**(**df**[**,4**:**6**])**,

Botrytis\_3\_a**=**df**[**,31**]-**rowMeans**(**df**[**,7**:**9**])**,

Botrytis\_3\_b**=**df**[**,32**]-**rowMeans**(**df**[**,7**:**9**])**,

Botrytis\_3\_c**=**df**[**,33**]-**rowMeans**(**df**[**,7**:**9**])**,

Botrytis\_6\_a**=**df**[**,34**]-**rowMeans**(**df**[**,10**:**12**])**,

Botrytis\_6\_b**=**df**[**,35**]-**rowMeans**(**df**[**,10**:**12**])**,

Botrytis\_6\_c**=**df**[**,36**]-**rowMeans**(**df**[**,10**:**12**])**,

Botrytis\_12\_a**=**df**[**,37**]-**rowMeans**(**df**[**,13**:**15**])**,

Botrytis\_12\_b**=**df**[**,38**]-**rowMeans**(**df**[**,13**:**15**])**,

Botrytis\_12\_c**=**df**[**,39**]-**rowMeans**(**df**[**,13**:**15**]))**

rownames**(**yFC**)=NULL**

rownames**(**bFC**)=NULL**

yFCmean**<-**data.frame**(**id**=**yFC**[**,1**]**,y01h**=**rowMeans**(**yFC**[**,c**(**2**:**4**)])**, y03h**=**rowMeans**(**yFC**[**,c**(**5**:**7**)])**, y06h**=**rowMeans**(**yFC**[**,c**(**8**:**10**)])**, y12h**=**rowMeans**(**yFC**[**,c**(**11**:**13**)]))**

bFCmean**<-**data.frame**(**id**=**bFC**[**,1**]**,b01h**=**rowMeans**(**bFC**[**,c**(**2**:**4**)])**, b03h**=**rowMeans**(**bFC**[**,c**(**5**:**7**)])**, b06h**=**rowMeans**(**bFC**[**,c**(**8**:**10**)])**, b12h**=**rowMeans**(**bFC**[**,c**(**11**:**13**)]))**

min\_FC **=** 1

y**=**data.frame**(**id**=**rownames**(**y**))**

y1**=**merge**(**yFC,y,by**=**"id", all**=**F**)**

y2**=**merge**(**yFCmean,y,by**=**"id", all**=**F**)**

y2 **<-** y2**[**apply**(**y2**[**,**-**1**]**,1,**function(**x**){**max**(**abs**(**x**))})** **>** min\_FC,**]**

y1**=**merge**(**y1, y2, by**=**"id"**)**

y1**=**y1**[**,**-**c**(**14**:**17**)]**

exprC**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprCtrl05.txt", header**=**T, sep**=**"\t"**)**

exprY**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprYeast05.txt", header**=**T, sep**=**"\t"**)**

exprC**=**data.frame**(**id**=**exprC**[**,1**])**

exprY**=**data.frame**(**id**=**exprY**[**,1**])**

expr**=**merge**(**exprC, exprY, by**=**"id", all**=TRUE)**

diff.id**<-(!**y1**$**id %in% expr**$**id**)**

temp**=**y1**[**diff.id,**]** #nrow=10

same.id**<-(**y1**$**id %in% expr**$**id**)**

y1**=**y1**[**same.id,**]**

y2**=**y2**[**same.id,**]**

write.table**(**y1, "k100Mmlog2VSTFCtoM1\_DESeq2\_sigYeastGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**y2, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

b**=**data.frame**(**id**=**rownames**(**b**))**

b1**=**merge**(**bFC,b,by**=**"id", all**=**F**)**

b2**=**merge**(**bFCmean,b,by**=**"id", all**=**F**)**

b2 **<-** b2**[**apply**(**b2**[**,**-**1**]**,1,**function(**x**){**max**(**abs**(**x**))})** **>** min\_FC,**]**

b1**=**merge**(**b1, b2, by**=**"id"**)**

b1**=**b1**[**,**-**c**(**14**:**17**)]**

exprC**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprCtrl05.txt", header**=**T, sep**=**"\t"**)**

exprB**=**read.table**(**"../Htseq-count/Gene\_k100Mm/GeneV2GtfReadCount/k100Mm\_GeneV2\_exprBotrytis05.txt", header**=**T, sep**=**"\t"**)**

exprC**=**data.frame**(**id**=**exprC**[**,1**])**

exprB**=**data.frame**(**id**=**exprB**[**,1**])**

expr**=**merge**(**exprC, exprB, by**=**"id", all**=TRUE)**

diff.id**<-(!**y1**$**id %in% expr**$**id**)**

temp**=**b1**[**diff.id,**]** #nrow=31

same.id**<-(**b1**$**id %in% expr**$**id**)**

b1**=**b1**[**same.id,**]**

b2**=**b2**[**same.id,**]**

write.table**(**b1, "k100Mmlog2VSTFCtoM1\_DESeq2\_sigBotrytisGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**b2, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigBotrytisGene.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

yb**=**merge**(**y2,b2, by**=**"id"**)**

yb\_y**=**yb**[**,1**:**5**]**

yb\_b**=**yb**[**,c**(**1,6**:**9**)]**

dif.id**<-(!**y2**$**id %in% b2**$**id**)**

y\_uniq**<-**y2**[**dif.id,**]** #unique yeast section

dif.id**<-(!**b2**$**id %in% y2**$**id**)**

b\_uniq**<-**b2**[**dif.id,**]** #unique botrytis section

write.table**(**yb, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastBotrytisGeneYB.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**yb\_y, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastBotrytisGeneY.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**yb\_b, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastBotrytisGeneB.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**y\_uniq, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastGeneYuniq.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**b\_uniq, "k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigBotrytisGeneBuniq.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Compare DEGs of different treatments

#R code

library**(**VennDiagram**)**

cal.overlap**=function** **(**x**)** **{**

**if** **(**1 **==** length**(**x**))** **{**

overlap **<-** x

**}**

**else** **if** **(**2 **==** length**(**x**))** **{**

A **<-** x**[[**1**]]**

B **<-** x**[[**2**]]**

nab **<-** intersect**(**A, B**)**

a1 **=** A**[!** A %in% nab**]**

a2 **=** B**[!** B %in% nab**]**

overlap **<-** list**(**a1 **=** a1, a2 **=** a2, a3 **=** nab**)**

**}**

**else** **if** **(**3 **==** length**(**x**))** **{**

A **<-** x**[[**1**]]**

B **<-** x**[[**2**]]**

C **<-** x**[[**3**]]**

nab **<-** intersect**(**A, B**)**

nbc **<-** intersect**(**B, C**)**

nac **<-** intersect**(**A, C**)**

nabc **<-** intersect**(**nab, C**)**

a5 **=** nabc

a2 **=** nab**[!** nab %in% a5**]**

a4 **=** nac**[-**which**(**nac %in% a5**)]**

a6 **=** nbc**[-**which**(**nbc %in% a5**)]**

a1 **=** A**[-**which**(**A %in% c**(**a2, a4, a5**))]**

a3 **=** B**[-**which**(**B %in% c**(**a2, a5, a6**))]**

a7 **=** C**[-**which**(**C %in% c**(**a4, a5, a6**))]**

overlap **<-** list**(**a5 **=** a5, a2 **=** a2, a4 **=** a4, a6 **=** a6, a1 **=** a1,

a3 **=** a3, a7 **=** a7**)**

**}**

**else** **if** **(**4 **==** length**(**x**))** **{**

A **<-** x**[[**1**]]**

B **<-** x**[[**2**]]**

C **<-** x**[[**3**]]**

D **<-** x**[[**4**]]**

nab **<-** intersect**(**A, B**)**

nbc **<-** intersect**(**B, C**)**

nac **<-** intersect**(**A, C**)**

nbd **<-** intersect**(**B, D**)**

nad **<-** intersect**(**A, D**)**

ncd **<-** intersect**(**C, D**)**

nabc **<-** intersect**(**nab, C**)**

nabd **<-** intersect**(**nab, D**)**

nacd **<-** intersect**(**nac, D**)**

nbcd **<-** intersect**(**nbc, D**)**

nabcd **<-** intersect **(**nabc, D**)**

a6 **=** nabcd

a12 **=** nabc**[-**which**(**nabc %in% a6**)]**

a11 **=** nabd**[-**which**(**nabd %in% a6**)]**

a5 **=** nacd**[-**which**(**nacd %in% a6**)]**

a7 **=** nbcd**[-**which**(**nbcd %in% a6**)]**

a15 **=** nab**[-**which**(**nab %in% c**(**a6,a11,a12**))]**

a4 **=** nac**[-**which**(**nac %in% c**(**a6,a5,a12**))]**

a10 **=** nad**[-**which**(**nad %in% c**(**a6,a5,a11**))]**

a13 **=** nbc**[-**which**(**nbc %in% c**(**a6,a7,a12**))]**

a8 **=** nbd**[-**which**(**nbd %in% c**(**a6,a7,a11**))]**

a2 **=** ncd**[-**which**(**ncd %in% c**(**a6,a5,a7**))]**

a9 **=** A**[-**which**(**A %in% c**(**a4,a5,a6,a10,a11,a12,a15**))]**

a14 **=** B**[-**which**(**B %in% c**(**a6,a7,a8,a11,a12,a13,a15**))]**

a1 **=** C**[-**which**(**C %in% c**(**a2,a4,a5,a6,a7,a12,a13**))]**

a3 **=** D**[-**which**(**D %in% c**(**a2,a5,a6,a7,a8,a10,a11**))]**

overlap **<-** list**(**a6**=**a6, a12**=**a12, a11**=**a11, a5**=**a5, a7**=**a7, a15**=**a15,

a4**=**a4, a10**=**a10, a13**=**a13, a8**=**a8, a2**=**a2, a9**=**a9, a14**=**a14, a1**=**a1, a3**=**a3**)**

**}**

**else** **{**

flog.error**(**"Invalid size of input object", name **=** "VennDiagramLogger"**)**

stop**(**"Invalid size of input object"**)**

**}**

return**(**overlap**)**

**}**

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/DESeq2\_Gene\_htseq\_FPKMover1/"**)**

y **=** read.table**(**"k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastGene.txt", header**=**T, sep**=**"\t"**)**

b **=** read.table**(**"k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigBotrytisGene.txt", header**=**T, sep**=**"\t"**)**

m **=** read.table**(**"k100Mmlog2VSTFCtoCtrl\_mean\_DESeq2\_sigMockGene.txt", header**=**T, sep**=**"\t"**)**

y**=**y**[**order**(**y**$**id**)**,**]**

b**=**b**[**order**(**b**$**id**)**,**]**

m**=**m**[**order**(**m**$**id**)**,**]**

y**=**data.frame**(**id**=**y**[**,1**])**

b**=**data.frame**(**id**=**b**[**,1**])**

m**=**data.frame**(**id**=**m**[**,1**])**

v**<-**venn.diagram**(**list**(**"Yeast"**=**y**[**,1**]**, "Botrytis"**=**b**[**,1**]**,"Mock"**=**m**[**,1**])**, filename**=NULL)**

grid.newpage**()**

pdf**(**file **=** "VennDiagram\_log2VSTFCtoM1\_DEG\_YBM.pdf", width**=**7, height**=**7**)**

pushViewport**(**viewport**(**width**=**unit**(**0.75, "npc"**)**, height **=** unit**(**0.75, "npc"**)))**

grid.draw**(**v**)**

dev.off**()**

overlap**<-**cal.overlap**(**x**=**list**(**"Yeast"**=**y**[**,1**]**, "Botrytis"**=**b**[**,1**]**,"Mock"**=**m**[**,1**]))**

capture.output**(**print**(**overlap**)**, file **=** "OverlapPrint\_log2VSTFCtoM1\_DEG\_YBM.txt"**)**

capture.output**(**summary**(**overlap**)**, file **=** "OverlapSummary\_log2VSTFCtoM1\_DEG\_YBM.txt"**)**

v**<-**venn.diagram**(**list**(**"Yeast"**=**y**[**,1**]**, "Botrytis"**=**b**[**,1**])**, filename**=NULL)**

grid.newpage**()**

pdf**(**file **=** "VennDiagram\_log2VSTFCtoM1\_DEG\_YB.pdf", width**=**7, height**=**7**)**

pushViewport**(**viewport**(**width**=**unit**(**0.75, "npc"**)**, height **=** unit**(**0.75, "npc"**)))**

grid.draw**(**v**)**

dev.off**()**

overlap**<-**cal.overlap**(**x**=**list**(**"Yeast"**=**y**[**,1**]**, "Botrytis"**=**b**[**,1**]))**

capture.output**(**print**(**overlap**)**, file **=** "OverlapPrint\_log2VSTFCtoM1\_DEG\_YB.txt"**)**

capture.output**(**summary**(**overlap**)**, file **=** "OverlapSummary\_log2VSTFCtoM1\_DEG\_YB.txt"**)**

## Generate files for clustering

#R code

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/DESeq2\_Gene\_htseq\_FPKMover1/"**)**

df **=** read.table**(**"k100Mmlog2VSTFCtoCtrl\_mean\_DESeq2\_sigMockGene.txt", header**=**T, sep**=**"\t"**)**

df**$**m00h **=** "0"

df **=** df**[**,c**(**1,6, 2**:**5**)]**

write.table**(**df, "Genesis\_Mock\_DEG.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

df **=** read.table**(**"k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigYeastGene.txt", header**=**T, sep**=**"\t"**)**

df**$**y00h **=** "0"

df **=** df**[**,c**(**1,6, 2**:**5**)]**

write.table**(**df, "Genesis\_Yeast\_DEG.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

df **=** read.table**(**"k100Mmlog2VSTFCtoM1\_mean\_DESeq2\_sigBotrytisGene.txt", header**=**T, sep**=**"\t"**)**

df**$**b00h **=** "0"

df **=** df**[**,c**(**1,6, 2**:**5**)]**

write.table**(**df, "Genesis\_Botrytis\_DEG.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## GO analysis