# Analysing expression candidates by family

Contents

[Analysing expression candidates by family 1](#_Toc533363697)

[Analysing expression candidates by family 2](#_Toc533363698)

[1. Expression candidates of control (T=0) 2](#_Toc533363699)

[2. Expression candidates of mock treatment 3](#_Toc533363700)

[3. Expression candidates of yeast treatment 5](#_Toc533363701)

[4. Expression candidates of botrytis treatment 6](#_Toc533363702)

# Analysing expression candidates by family

## Expression candidates of control (T=0)

#R code

#Find full length candidates and categorize them by family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

coverage**=**read.table**(**"../BedCov\_OverlapBP\_sense\_breadthCov.txt", header**=**T, sep**=**"\t"**)**

treated**=**read.table**(**"AllExpeCandidate\_ctrl\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

cov**=**coverage**[**,c**(**4,7**:**9**)]**

data**=**merge**(**treated, cov, by.x**=**"id", by.y**=**"TEm"**)**

data**$**Mean\_cov00**=**rowMeans**(**data**[**,13**:**15**])**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

#Calculate expression candidates of each family

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "treated\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, data**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**df**[**,c**(**1,2**)]**, candidate**=**as.numeric**(**temp**[**,2**]))**

#Calculate TEs of zero expression and under-threshold TEs of each family

zero**=**cbind**(**cov, Mean**=**rowMeans**(**cov**[**,2**:**4**]))**

zero**=**subset**(**zero, zero**$**Mean**==**0**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "zeroExpr"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, zero**$**TEm**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**family, No\_expr**=**as.numeric**(**temp**[**,2**]))**

family**$**under\_threshold **<-** **(**family**$**copynumber **-** family**$**candidate **-** family**$**No\_expr**)**

#Calculate full-length candidates and those have breadth of coverage > 0.95

data2**=**subset**(**data, data**$**Lratio**>**0.9**)** #nrow=338

min\_cov **=** 0.95

data3 **<-** subset**(**data2, data2**$**Mean\_cov00 **>** min\_cov**)** #nrow=11

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthCandid"**=**numeric**(**0**)**, "fullLengthFullCov"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated1**=**data2**[**grepl**(**id, data2**$**id**)**,**]**

dftreated2**=**data3**[**grepl**(**id, data3**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated1**)**,nrow**(**dftreated2**))**

**}**

family**=**cbind**(**family, temp**)**

family**=**family**[**,**-**6**]**

write.table**(**family, "Candidates\_REfamily\_tag\_ctrl\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**data3, "CandidatesFullLenthFullCov\_tag\_ctrl\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Count trackable candidates and untrackable candidates

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"AllExpeCandidate\_ctrl\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_ctrl\_new.txt", header**=**T, sep**=**"\t"**)**

tracking**=**subset**(**data, data**$**C\_tracking**!=**" countdepthonly”**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, trackable\_candidate**=**temp**[**,2**])**

notracking**=**subset**(**data, data**$**C\_tracking**==**"countdepthonly"**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, untrackable\_candidate**=**temp**[**,2**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "FL\_trackable\_Candidate"**=**numeric**(**0**)**, "FG\_trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_trackable\_candidate**=**temp**[**,2**]**, FG\_trackable\_candidate**=**temp**[**,3**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**,"FL\_untrackable\_Candidate"**=**numeric**(**0**)**, "FG\_untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_untrackable\_candidate**=**temp**[**,2**]**, FG\_untrackable\_candidate**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_ctrl\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Calculate full-length and fragmented annotated TEs of each TE family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"../../Reference/Curated/AllTEsExpanded\_curated\_tags\_V4.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_ctrl\_new.txt", header**=**T, sep**=**"\t"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthTE"**=**numeric**(**0**)**, "fragmentTE"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**data**[**grepl**(**id, data**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, fullLengthTE**=**temp**[**,2**]**, fragmentTE**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_ctrl\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Expression candidates of mock treatment

#R code

#Find full length candidates and categorize them by family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

coverage**=**read.table**(**"../BedCov\_OverlapBP\_sense\_breadthCov.txt", header**=**T, sep**=**"\t"**)**

treated**=**read.table**(**"AllExpeCandidate\_mock\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

cov**=**coverage**[**,c**(**4,10**:**21**)]**

data**=**merge**(**treated, cov, by.x**=**"id", by.y**=**"TEm"**)**

data**$**Mean\_cov00**=**rowMeans**(**data**[**,13**:**24**])**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

#Calculate expression candidates of each family

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "treated\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, data**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**df**[**,c**(**1,2**)]**, candidate**=**as.numeric**(**temp**[**,2**]))**

#Calculate TEs of zero expression and under-threshold TEs of each family

zero**=**cbind**(**cov, Mean**=**rowMeans**(**cov**[**,2**:**13**]))**

zero**=**subset**(**zero, zero**$**Mean**==**0**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "zeroExpr"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, zero**$**TEm**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**family, No\_expr**=**as.numeric**(**temp**[**,2**]))**

family**$**under\_threshold **<-** **(**family**$**copynumber **-** family**$**candidate **-** family**$**No\_expr**)**

#Calculate full-length candidates and those have breadth of coverage > 0.95

data2**=**subset**(**data, data**$**Lratio**>**0.9**)** #nrow=497

min\_cov **=** 0.95

data3 **<-** subset**(**data2, data2**$**Mean\_cov00 **>** min\_cov**)** #nrow=12

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthCandid"**=**numeric**(**0**)**, "fullLengthFullCov"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated1**=**data2**[**grepl**(**id, data2**$**id**)**,**]**

dftreated2**=**data3**[**grepl**(**id, data3**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated1**)**,nrow**(**dftreated2**))**

**}**

family**=**cbind**(**family, temp**)**

family**=**family**[**,**-**6**]**

write.table**(**family, "Candidates\_REfamily\_tag\_mock\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**data3, "CandidatesFullLenthFullCov\_tag\_mock\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Count trackable candidates and untrackable candidates

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"AllExpeCandidate\_mock\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_mock\_new.txt", header**=**T, sep**=**"\t"**)**

tracking**=**subset**(**data, data**$**C\_tracking**!=**"countdepthonly"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, trackable\_candidate**=**temp**[**,2**])**

notracking**=**subset**(**data, data**$**C\_tracking**==**"countdepthonly"**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, untrackable\_candidate**=**temp**[**,2**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "FL\_trackable\_Candidate"**=**numeric**(**0**)**, "FG\_trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_trackable\_candidate**=**temp**[**,2**]**, FG\_trackable\_candidate**=**temp**[**,3**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**,"FL\_untrackable\_Candidate"**=**numeric**(**0**)**, "FG\_untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_untrackable\_candidate**=**temp**[**,2**]**, FG\_untrackable\_candidate**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_mock\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Calculate full-length and fragmented annotated TEs of each TE family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"../../Reference/Curated/AllTEsExpanded\_curated\_tags\_V4.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_mock\_new.txt", header**=**T, sep**=**"\t"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthTE"**=**numeric**(**0**)**, "fragmentTE"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**data**[**grepl**(**id, data**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, fullLengthTE**=**temp**[**,2**]**, fragmentTE**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_mock\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Expression candidates of yeast treatment

#R code

#Find full length candidates and categorize them by family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

coverage**=**read.table**(**"../BedCov\_OverlapBP\_sense\_breadthCov.txt", header**=**T, sep**=**"\t"**)**

treated**=**read.table**(**"AllExpeCandidate\_yeast\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

cov**=**coverage**[**,c**(**4,22**:**33**)]**

data**=**merge**(**treated, cov, by.x**=**"id", by.y**=**"TEm"**)**

data**$**Mean\_cov00**=**rowMeans**(**data**[**,13**:**24**])**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

#Calculate expression candidates of each family

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "treated\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, data**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**df**[**,c**(**1,2**)]**, candidate**=**as.numeric**(**temp**[**,2**]))**

#Calculate TEs of zero expression and under-threshold TEs of each family

zero**=**cbind**(**cov, Mean**=**rowMeans**(**cov**[**,2**:**13**]))**

zero**=**subset**(**zero, zero**$**Mean**==**0**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "zeroExpr"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, zero**$**TEm**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**family, No\_expr**=**as.numeric**(**temp**[**,2**]))**

family**$**under\_threshold **<-** **(**family**$**copynumber **-** family**$**candidate **-** family**$**No\_expr**)**

#Calculate full-length candidates and those have breadth of coverage > 0.95

data2**=**subset**(**data, data**$**Lratio**>**0.9**)** #nrow=539

min\_cov **=** 0.95

data3 **<-** subset**(**data2, data2**$**Mean\_cov00 **>** min\_cov**)** #nrow=9

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthCandid"**=**numeric**(**0**)**, "fullLengthFullCov"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated1**=**data2**[**grepl**(**id, data2**$**id**)**,**]**

dftreated2**=**data3**[**grepl**(**id, data3**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated1**)**,nrow**(**dftreated2**))**

**}**

family**=**cbind**(**family, temp**)**

family**=**family**[**,**-**6**]**

write.table**(**family, "Candidates\_REfamily\_tag\_yeast\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**data3, "CandidatesFullLenthFullCov\_tag\_mock\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Count trackable candidates and untrackable candidates

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"AllExpeCandidate\_yeast\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_mock\_new.txt", header**=**T, sep**=**"\t"**)**

tracking**=**subset**(**data, data**$**C\_tracking**!=**"countdepthonly"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, trackable\_candidate**=**temp**[**,2**])**

notracking**=**subset**(**data, data**$**C\_tracking**==**"countdepthonly"**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, untrackable\_candidate**=**temp**[**,2**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "FL\_trackable\_Candidate"**=**numeric**(**0**)**, "FG\_trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_trackable\_candidate**=**temp**[**,2**]**, FG\_trackable\_candidate**=**temp**[**,3**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**,"FL\_untrackable\_Candidate"**=**numeric**(**0**)**, "FG\_untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_untrackable\_candidate**=**temp**[**,2**]**, FG\_untrackable\_candidate**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_yeast\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Calculate full-length and fragmented annotated TEs of each TE family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"../../Reference/Curated/AllTEsExpanded\_curated\_tags\_V4.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_yeast\_new.txt", header**=**T, sep**=**"\t"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthTE"**=**numeric**(**0**)**, "fragmentTE"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**data**[**grepl**(**id, data**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, fullLengthTE**=**temp**[**,2**]**, fragmentTE**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_yeast\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

## Expression candidates of botrytis treatment

#R code

#Find full length candidates and categorize them by family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

coverage**=**read.table**(**"../BedCov\_OverlapBP\_sense\_breadthCov.txt", header**=**T, sep**=**"\t"**)**

treated**=**read.table**(**"AllExpeCandidate\_botrytis\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

cov**=**coverage**[**,c**(**4,34**:**45**)]**

data**=**merge**(**treated, cov, by.x**=**"id", by.y**=**"TEm"**)**

data**$**Mean\_cov00**=**rowMeans**(**data**[**,13**:**24**])**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

#Calculate expression candidates of each family

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "treated\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, data**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**df**[**,c**(**1,2**)]**, candidate**=**as.numeric**(**temp**[**,2**]))**

#Calculate TEs of zero expression and under-threshold TEs of each family

zero**=**cbind**(**cov, Mean**=**rowMeans**(**cov**[**,2**:**13**]))**

zero**=**subset**(**zero, zero**$**Mean**==**0**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "zeroExpr"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated**=**data**[**grepl**(**id, zero**$**TEm**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated**))**

**}**

family**=**cbind**(**family, No\_expr**=**as.numeric**(**temp**[**,2**]))**

family**$**under\_threshold **<-** **(**family**$**copynumber **-** family**$**candidate **-** family**$**No\_expr**)**

#Calculate full-length candidates and those have breadth of coverage > 0.95

data2**=**subset**(**data, data**$**Lratio**>**0.9**)** #nrow=481

min\_cov **=** 0.95

data3 **<-** subset**(**data2, data2**$**Mean\_cov00 **>** min\_cov**)** #nrow=10

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthCandid"**=**numeric**(**0**)**, "fullLengthFullCov"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dftreated1**=**data2**[**grepl**(**id, data2**$**id**)**,**]**

dftreated2**=**data3**[**grepl**(**id, data3**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dftreated1**)**,nrow**(**dftreated2**))**

**}**

family**=**cbind**(**family, temp**)**

family**=**family**[**,**-**6**]**

write.table**(**family, "Candidates\_REfamily\_tag\_botrytis\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

write.table**(**data3, "CandidatesFullLenthFullCov\_tag\_botrytis\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Count trackable candidates and untrackable candidates

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"AllExpeCandidate\_botrytis\_tag\_new.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_botrytis\_new.txt", header**=**T, sep**=**"\t"**)**

tracking**=**subset**(**data, data**$**C\_tracking**!=**"countdepthonly"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, trackable\_candidate**=**temp**[**,2**])**

notracking**=**subset**(**data, data**$**C\_tracking**==**"countdepthonly"**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**dfnew**))**

**}**

family**=**cbind**(**family, untrackable\_candidate**=**temp**[**,2**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "FL\_trackable\_Candidate"**=**numeric**(**0**)**, "FG\_trackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**tracking**[**grepl**(**id, tracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_trackable\_candidate**=**temp**[**,2**]**, FG\_trackable\_candidate**=**temp**[**,3**])**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**,"FL\_untrackable\_Candidate"**=**numeric**(**0**)**, "FG\_untrackable\_Candidate"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**notracking**[**grepl**(**id, notracking**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, FL\_untrackable\_candidate**=**temp**[**,2**]**, FG\_untrackable\_candidate**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_botrytis\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**

#Calculate full-length and fragmented annotated TEs of each TE family

setwd**(**"/media/ting-hsuan/ExtraDrive1/PhD/analysis/ECstress\_TEalignment/ExprCandidate"**)**

df**=**read.table**(**"../../Reference/Curated/AllRefTEsExpanded\_copyNumber.txt", header**=**T, sep**=**"\t"**)**

data**=**read.table**(**"../../Reference/Curated/AllTEsExpanded\_curated\_tags\_V4.txt", header**=**T, sep**=**"\t"**)**

family**=**read.table**(**"Candidates\_REfamily\_tag\_botrytis\_new.txt", header**=**T, sep**=**"\t"**)**

df**$**V3**=**paste**(**df**$**id, "\_", sep**=**""**)**

temp**=**data.frame**(**"id"**=**numeric**(**0**)**, "fullLengthTE"**=**numeric**(**0**)**, "fragmentTE"**=**numeric**(**0**))**

**for** **(**i **in** 1**:**nrow**(**df**)){**

re**=**df**[**i,**]**

id**=**re**[**,3**]**

dfnew**=**data**[**grepl**(**id, data**$**id**)**,**]**

FL**=**subset**(**dfnew, dfnew**$**Lratio**>**0.9**)**

FG**=**subset**(**dfnew, dfnew**$**Lratio**<=**0.9**)**

temp**[**nrow**(**temp**)+**1,**]<-**c**(**as.character**(**re**[**,1**])**,nrow**(**FL**)**, nrow**(**FG**))**

**}**

family**=**cbind**(**family, fullLengthTE**=**temp**[**,2**]**, fragmentTE**=**temp**[**,3**])**

write.table**(**family, "Candidates\_REfamily\_tag\_botrytis\_new.txt", col.names**=**T, row.names**=**F, sep**=**"\t", quote**=**F**)**