# Appendix 5: Isolating DE Genes from Expression analysis

## Initial environment setup (Code Block 1)

knitr::opts\_chunk$set(echo = TRUE)

## Clear Global Environment (Code Block 2)

remove(list = ls())

## (Code Block 3)

#Set Library Directory  
#PC Path  
.libPaths(c("L:/RStudios/RPackRatLibLocations", "L:/RStudios/RPackRat\_2019\_04\_DESEQLibs"))  
  
#Set working directory  
#PC Path  
setwd("E:/Dropbox/Dropbox/Harrison Lab - Trevor Randall/RNASeq Analysis/RNASeqAnlyPkrat\_2020\_03/CF39vsCF39S/Pure Temperature Analysis Comparisons/")  
#Laptop Path  
#Set Library Directory  
#.libPaths(c(""))  
#setwd("")  
#sink(file = "./RSessionRawRun.txt")

## Libraries to install (Code Block 4)

install.packages('VennDiagram')

## Libraries to open (Code Block 5)

library(readr)  
library(VennDiagram)

## Opening Preprocessed Datasets (Code Block 6)

TDESEQList\_25vs25 <- read\_delim("../25 vs 25/DE\_Reg\_Genes\_Analysis/TDESEQ25vs25.txt", "\t", escape\_double = FALSE, trim\_ws = TRUE)  
colnames(TDESEQList\_25vs25) <- c("locus", "baseMean", "log2FoldChange\_25C", "lfcSE", "stat", "pvalue\_25C", "padj\_25C", "gene\_name", "gene\_discription", "raw\_locus")  
  
TDESEQList\_29vs29 <- read\_delim("../29 vs 29/DE\_Reg\_Genes\_Analysis/TDESEQ29vs29.txt", "\t", escape\_double = FALSE, trim\_ws = TRUE)  
colnames(TDESEQList\_29vs29) <- c("locus", "baseMean", "log2FoldChange\_29C", "lfcSE", "stat", "pvalue\_29C", "padj\_29C", "gene\_name", "gene\_discription", "raw\_locus")  
  
TDESEQList\_33vs33 <- read\_delim("../33 vs 33/DE\_Reg\_Genes\_Analysis/TDESEQ33vs33.txt", "\t", escape\_double = FALSE, trim\_ws = TRUE)  
colnames(TDESEQList\_33vs33) <- c("locus", "baseMean", "log2FoldChange\_33C", "lfcSE", "stat", "pvalue\_33C", "padj\_33C", "gene\_name", "gene\_discription", "raw\_locus")  
  
TDESEQList\_37vs37 <- read\_delim("../37 vs 37/DE\_Reg\_Genes\_Analysis/TDESEQ37vs37.txt", "\t", escape\_double = FALSE, trim\_ws = TRUE)  
colnames(TDESEQList\_37vs37) <- c("locus", "baseMean", "log2FoldChange\_37C", "lfcSE", "stat", "pvalue\_37C", "padj\_37C", "gene\_name", "gene\_discription", "raw\_locus")  
  
TDESEQList\_41vs41 <- read\_delim("../41 vs 41/DE\_Reg\_Genes\_Analysis/TDESEQ41vs41.txt", "\t", escape\_double = FALSE, trim\_ws = TRUE)  
colnames(TDESEQList\_41vs41) <- c("locus", "baseMean", "log2FoldChange\_41C", "lfcSE", "stat", "pvalue\_41C", "padj\_41C", "gene\_name", "gene\_discription", "raw\_locus")

## Convert Log2Fold Change to Fold Change (Code Block 7)

TDESEQList\_25vs25$FoldChange\_25 <- ifelse(((TDESEQList\_25vs25$log2FoldChange\_25C) >= 0), (2^TDESEQList\_25vs25$log2FoldChange\_25C), (-1/(2^TDESEQList\_25vs25$log2FoldChange\_25C)))  
  
ColOrder25 <- c("locus", "baseMean", "log2FoldChange\_25C", "FoldChange\_25", "lfcSE", "stat", "pvalue\_25C", "padj\_25C", "gene\_name", "gene\_discription", "raw\_locus")  
TDESEQList\_25vs25 <- TDESEQList\_25vs25[, ColOrder25]  
  
  
TDESEQList\_29vs29$FoldChange\_29 <- ifelse(((TDESEQList\_29vs29$log2FoldChange\_29C) >= 0), (2^TDESEQList\_29vs29$log2FoldChange\_29C), (-1/(2^TDESEQList\_29vs29$log2FoldChange\_29C)))  
  
ColOrder29 <- c("locus", "baseMean", "log2FoldChange\_29C", "FoldChange\_29", "lfcSE", "stat", "pvalue\_29C", "padj\_29C", "gene\_name", "gene\_discription", "raw\_locus")  
TDESEQList\_29vs29 <- TDESEQList\_29vs29[, ColOrder29]  
  
  
TDESEQList\_33vs33$FoldChange\_33 <- ifelse(((TDESEQList\_33vs33$log2FoldChange\_33C) >= 0), (2^TDESEQList\_33vs33$log2FoldChange\_33C), (-1/(2^TDESEQList\_33vs33$log2FoldChange\_33C)))  
  
ColOrder33 <- c("locus", "baseMean", "log2FoldChange\_33C", "FoldChange\_33", "lfcSE", "stat", "pvalue\_33C", "padj\_33C", "gene\_name", "gene\_discription", "raw\_locus")  
TDESEQList\_33vs33 <- TDESEQList\_33vs33[, ColOrder33]  
  
  
TDESEQList\_37vs37$FoldChange\_37 <- ifelse(((TDESEQList\_37vs37$log2FoldChange\_37C) >= 0), (2^TDESEQList\_37vs37$log2FoldChange\_37C), (-1/(2^TDESEQList\_37vs37$log2FoldChange\_37C)))  
  
ColOrder37 <- c("locus", "baseMean", "log2FoldChange\_37C", "FoldChange\_37", "lfcSE", "stat", "pvalue\_37C", "padj\_37C", "gene\_name", "gene\_discription", "raw\_locus")  
TDESEQList\_37vs37 <- TDESEQList\_37vs37[, ColOrder37]  
  
  
TDESEQList\_41vs41$FoldChange\_41 <- ifelse(((TDESEQList\_41vs41$log2FoldChange\_41C) >= 0), (2^TDESEQList\_41vs41$log2FoldChange\_41C), (-1/(2^TDESEQList\_41vs41$log2FoldChange\_41C)))  
  
ColOrder41 <- c("locus", "baseMean", "log2FoldChange\_41C", "FoldChange\_41", "lfcSE", "stat", "pvalue\_41C", "padj\_41C", "gene\_name", "gene\_discription", "raw\_locus")  
TDESEQList\_41vs41 <- TDESEQList\_41vs41[, ColOrder41]

## Saving correct format for each table (Code Block 8)

write.table(x = TDESEQList\_25vs25, file="./OriginalFiles/TDESEQList\_25vs25\_Original.txt", sep="\t", row.names=FALSE, col.names=colnames(TDESEQList\_25vs25), quote=F)  
  
write.table(x = TDESEQList\_29vs29, file="./OriginalFiles/TDESEQList\_29vs29\_Original.txt", sep="\t", row.names=FALSE, col.names=colnames(TDESEQList\_29vs29), quote=F)  
  
write.table(x = TDESEQList\_33vs33, file="./OriginalFiles/TDESEQList\_33vs33\_Original.txt", sep="\t", row.names=FALSE, col.names=colnames(TDESEQList\_33vs33), quote=F)  
  
write.table(x = TDESEQList\_37vs37, file="./OriginalFiles/TDESEQList\_37vs37\_Original.txt", sep="\t", row.names=FALSE, col.names=colnames(TDESEQList\_37vs37), quote=F)  
  
write.table(x = TDESEQList\_41vs41, file="./OriginalFiles/TDESEQList\_41vs41\_Original.txt", sep="\t", row.names=FALSE, col.names=colnames(TDESEQList\_41vs41), quote=F)

## Parcing patterns from files

P designation is temperature pure (so for instance CF3925vsCF39S25 is P25 or CF3929vsCF39S29 is P29)

## Identify trends from P25 to P29 (Code Block 9)

P25vsP29\_SameGenes<- TDESEQList\_25vs25[(TDESEQList\_25vs25$locus %in% TDESEQList\_29vs29$locus), ]  
  
P25\_DiffGenesToP29 <- TDESEQList\_25vs25[!((TDESEQList\_25vs25$locus %in% TDESEQList\_29vs29$locus)& (!is.na(TDESEQList\_25vs25$padj\_25C))), ]  
  
P29\_DiffGenesToP25 <- TDESEQList\_29vs29[!((TDESEQList\_29vs29$locus %in% TDESEQList\_25vs25$locus)& (!is.na(TDESEQList\_29vs29$padj\_29C))),]

## Identify trends from P25 to P33 (Code Block 10)

P25vsP33\_SameGenes<- TDESEQList\_25vs25[(TDESEQList\_25vs25$locus %in% TDESEQList\_33vs33$locus), ]  
  
P25\_DiffGenesToP33 <- TDESEQList\_25vs25[!((TDESEQList\_25vs25$locus %in% TDESEQList\_33vs33$locus)& (!is.na(TDESEQList\_25vs25$padj\_25C))), ]  
  
P33\_DiffGenesToP25 <- TDESEQList\_33vs33[!((TDESEQList\_33vs33$locus %in% TDESEQList\_25vs25$locus)& (!is.na(TDESEQList\_33vs33$padj\_33C))),]

## Identify trends from P25 to P37 (Code Block 11)

P25vsP37\_SameGenes<- TDESEQList\_25vs25[(TDESEQList\_25vs25$locus %in% TDESEQList\_37vs37$locus), ]  
  
P25\_DiffGenesToP37 <- TDESEQList\_25vs25[!((TDESEQList\_25vs25$locus %in% TDESEQList\_37vs37$locus)& (!is.na(TDESEQList\_25vs25$padj\_25C))), ]  
  
P37\_DiffGenesToP25 <- TDESEQList\_37vs37[!((TDESEQList\_37vs37$locus %in% TDESEQList\_25vs25$locus)& (!is.na(TDESEQList\_37vs37$padj\_37C))),]

## Identify trends from P25 to P41 (Code Block 12)

P25vsP41\_SameGenes<- TDESEQList\_25vs25[(TDESEQList\_25vs25$locus %in% TDESEQList\_41vs41$locus), ]  
  
P25\_DiffGenesToP41 <- TDESEQList\_25vs25[!((TDESEQList\_25vs25$locus %in% TDESEQList\_41vs41$locus)& (!is.na(TDESEQList\_25vs25$padj\_25C))), ]  
  
P41\_DiffGenesToP25 <- TDESEQList\_41vs41[!((TDESEQList\_41vs41$locus %in% TDESEQList\_25vs25$locus)& (!is.na(TDESEQList\_41vs41$padj\_41C))),]

## Identify trends from P29 to P33 (Code Block 13)

P29vsP33\_SameGenes<- TDESEQList\_29vs29[(TDESEQList\_29vs29$locus %in% TDESEQList\_33vs33$locus), ]  
  
P29\_DiffGenesToP33 <- TDESEQList\_29vs29[!((TDESEQList\_29vs29$locus %in% TDESEQList\_33vs33$locus)& (!is.na(TDESEQList\_29vs29$padj\_29C))), ]  
  
P33\_DiffGenesToP29 <- TDESEQList\_33vs33[!((TDESEQList\_33vs33$locus %in% TDESEQList\_29vs29$locus)& (!is.na(TDESEQList\_33vs33$padj\_33C))),]

## Identify trends from P29 to P37 (Code Block 14)

P29vsP37\_SameGenes<- TDESEQList\_29vs29[(TDESEQList\_29vs29$locus %in% TDESEQList\_37vs37$locus), ]  
  
P29\_DiffGenesToP37 <- TDESEQList\_29vs29[!((TDESEQList\_29vs29$locus %in% TDESEQList\_37vs37$locus)& (!is.na(TDESEQList\_29vs29$padj\_29C))), ]  
  
P37\_DiffGenesToP29 <- TDESEQList\_37vs37[!((TDESEQList\_37vs37$locus %in% TDESEQList\_29vs29$locus)& (!is.na(TDESEQList\_37vs37$padj\_37C))),]

## Identify trends from P29 to P41 (Code Block 15)

P29vsP41\_SameGenes<- TDESEQList\_29vs29[(TDESEQList\_29vs29$locus %in% TDESEQList\_41vs41$locus), ]  
  
P29\_DiffGenesToP41 <- TDESEQList\_29vs29[!((TDESEQList\_29vs29$locus %in% TDESEQList\_41vs41$locus)& (!is.na(TDESEQList\_29vs29$padj\_29C))), ]  
  
P41\_DiffGenesToP29 <- TDESEQList\_41vs41[!((TDESEQList\_41vs41$locus %in% TDESEQList\_29vs29$locus)& (!is.na(TDESEQList\_41vs41$padj\_41C))),]

## Identify trends from P33 to P37 (Code Block 16)

P33vsP37\_SameGenes<- TDESEQList\_33vs33[(TDESEQList\_33vs33$locus %in% TDESEQList\_37vs37$locus), ]  
  
P33\_DiffGenesToP37 <- TDESEQList\_33vs33[!((TDESEQList\_33vs33$locus %in% TDESEQList\_37vs37$locus)& (!is.na(TDESEQList\_33vs33$padj\_33C))), ]  
  
P37\_DiffGenesToP33 <- TDESEQList\_37vs37[!((TDESEQList\_37vs37$locus %in% TDESEQList\_33vs33$locus)& (!is.na(TDESEQList\_37vs37$padj\_37C))),]

## Identify trends from P33 to P41 (Code Block 17)

P33vsP41\_SameGenes<- TDESEQList\_33vs33[(TDESEQList\_33vs33$locus %in% TDESEQList\_41vs41$locus), ]  
  
P33\_DiffGenesToP41 <- TDESEQList\_33vs33[!((TDESEQList\_33vs33$locus %in% TDESEQList\_41vs41$locus)& (!is.na(TDESEQList\_33vs33$padj\_33C))), ]  
  
P41\_DiffGenesToP33 <- TDESEQList\_41vs41[!((TDESEQList\_41vs41$locus %in% TDESEQList\_33vs33$locus)& (!is.na(TDESEQList\_41vs41$padj\_41C))),]

## Identify trends from P37 to P41 (Code Block 18)

P37vsP41\_SameGenes<- TDESEQList\_37vs37[(TDESEQList\_37vs37$locus %in% TDESEQList\_41vs41$locus), ]  
  
P37\_DiffGenesToP41 <- TDESEQList\_37vs37[!((TDESEQList\_37vs37$locus %in% TDESEQList\_41vs41$locus)& (!is.na(TDESEQList\_37vs37$padj\_37C))), ]  
  
P41\_DiffGenesToP37 <- TDESEQList\_41vs41[!((TDESEQList\_41vs41$locus %in% TDESEQList\_37vs37$locus)& (!is.na(TDESEQList\_41vs41$padj\_41C))),]

## Combine full files (Code Block 19)

LocusGeneAssociation <- read\_csv("../SourceRNASEQCountsForDESEQ/Locus\_GeneNamesAndPredictedFunction.csv")  
  
  
Combined\_List <- 0  
Combined\_List <- merge(TDESEQList\_25vs25, y = TDESEQList\_29vs29, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List <- merge(Combined\_List, y = TDESEQList\_33vs33, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List <- merge(Combined\_List, y = TDESEQList\_37vs37, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List <- merge(Combined\_List, y = TDESEQList\_41vs41, by.x= "locus", by.y = "locus", all.x = T, all.y = T)

## Reformat Files

## Shrink All (Code Block 20)

ColToRemove <- c("baseMean", "lfcSE", "stat", "gene\_name", "gene\_discription", "raw\_locus")  
  
TDESEQList\_25vs25\_Shrunk <- TDESEQList\_25vs25[, !names(TDESEQList\_25vs25) %in% ColToRemove, drop = F]  
TDESEQList\_29vs29\_Shrunk <- TDESEQList\_29vs29[, !names(TDESEQList\_29vs29) %in% ColToRemove, drop = F]  
TDESEQList\_33vs33\_Shrunk <- TDESEQList\_33vs33[, !names(TDESEQList\_33vs33) %in% ColToRemove, drop = F]  
TDESEQList\_37vs37\_Shrunk <- TDESEQList\_37vs37[, !names(TDESEQList\_37vs37) %in% ColToRemove, drop = F]  
TDESEQList\_41vs41\_Shrunk <- TDESEQList\_41vs41[, !names(TDESEQList\_41vs41) %in% ColToRemove, drop = F]

## Combine shrunk files (Code Block 21)

LocusGeneAssociation <- read\_csv("../SourceRNASEQCountsForDESEQ/Locus\_GeneNamesAndPredictedFunction.csv")  
  
  
Combined\_List\_Shrunk <- 0  
Combined\_List\_Shrunk <- merge(TDESEQList\_25vs25\_Shrunk, y = TDESEQList\_29vs29\_Shrunk, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List\_Shrunk <- merge(Combined\_List\_Shrunk, y = TDESEQList\_33vs33\_Shrunk, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List\_Shrunk <- merge(Combined\_List\_Shrunk, y = TDESEQList\_37vs37\_Shrunk, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List\_Shrunk <- merge(Combined\_List\_Shrunk, y = TDESEQList\_41vs41\_Shrunk, by.x= "locus", by.y = "locus", all.x = T, all.y = T)  
Combined\_List\_Shrunk <- merge(Combined\_List\_Shrunk, LocusGeneAssociation, by = "locus")  
  
write.table(x = Combined\_List\_Shrunk, file="./Combined List Analysis Shrunk.txt", sep="\t", row.names=FALSE, col.names=colnames(Combined\_List\_Shrunk), quote=F)

## (Code Block 22)

Remove Igbot/Igtop form analysis Remove igtop/igbot from analysis Remove Unwanted Columns

ColToRemove <- c("baseMean", "lfcSE", "stat", "gene", "product", "locusNumber")  
  
Combined\_List\_Shrunk2\_Full <- Combined\_List\_Shrunk[, !names(Combined\_List\_Shrunk) %in% ColToRemove, drop = F]  
Combined\_List\_Shrunk2TailListSense <- Combined\_List\_Shrunk2\_Full[((grep((glob2rx('\*\_sense')), Combined\_List\_Shrunk2\_Full$locus))), ]  
Combined\_List\_Shrunk2TailListAntiSense <- Combined\_List\_Shrunk2\_Full[((grep((glob2rx('\*\_antis')), Combined\_List\_Shrunk2\_Full$locus))), ]  
Combined\_List\_Shrunk2TailListMerged <- rbind(Combined\_List\_Shrunk2TailListSense, Combined\_List\_Shrunk2TailListAntiSense)  
Combined\_List\_Shrunk2 <- Combined\_List\_Shrunk2TailListMerged[order(Combined\_List\_Shrunk2TailListMerged$locus), ]  
Combined\_List\_Shrunk2 <- merge(Combined\_List\_Shrunk2, LocusGeneAssociation, by = "locus")  
  
write.table(x = Combined\_List\_Shrunk2, file="./Combined List Analysis Shrunk2.txt", sep="\t", row.names=FALSE, col.names=colnames(Combined\_List\_Shrunk2), quote=F)

## Session Info (Code Block 23)

sink(file = "./SessionInfo.txt")  
sessionInfo()  
sink(file = NULL)

## Citations (Code Block 24)

sink(file = "./SessionCitations.txt")  
  
citation("VennDiagram")  
citation("readr")  
citation("gridExtra")  
  
sink(file = NULL)

Venn Diagram Page: <https://rstudio-pubs-static.s3.amazonaws.com/13301_6641d73cfac741a59c0a851feb99e98b.html?fbclid=IwAR1h2pL9V1tblTk4z9E1JX9udrel2ILE_WKKp2M8Kjd6HB0ZC6wwctFu61Y>