Metabolome heatmaps

#LOAD PACKAGES

library(data.table) #for loading in data  
library(ComplexHeatmap) #for plotting heatmap  
library(dendextend) #for editing dendrograms in heatmap  
library(RColorBrewer) #for dendrogram colours  
library(viridis) #for heatmap colours

#LOAD DATA

#0dpi  
zero\_df <- read.csv("Data/0dpi\_ANOVA\_NO\_FDR.csv", header = TRUE) #load data fast with fread  
zero\_df <- data.frame(zero\_df, row.names = "id") #convert to df so metabolite can be set as index col  
  
#1dpi  
one\_df <- data.table::fread("Data/1dpi\_ANOVA\_NO\_FDR.csv", header = TRUE) #load data fast with fread  
one\_df <- data.frame(one\_df, row.names = "id") #convert to df so metabolite can be set as index col  
  
#2dpi  
two\_df <- data.table::fread("Data/2dpi\_ANOVA\_NO\_FDR.csv", header = TRUE) #load data fast with fread  
two\_df <- data.frame(two\_df, row.names = "id") #convert to df so metabolite can be set as index col

#PREPARE DATA

#0dpi  
zero\_df <- subset(zero\_df, zero\_df$ANOVApValue\_p.value <= 0.01) #select only data with p<=0.01  
  
zero\_df <- zero\_df[-c(1,2,19)] #remove rt mz and anova p-value columns as not needed for heatmap  
  
zero\_df[zero\_df==0] <- NA #convert 0 to NAs  
which(is.na(zero\_df), arr.ind=TRUE) #check which columns contain NAs/0s  
zero\_df <- na.omit(zero\_df) #remove NAs/0s  
  
#1dpi  
one\_df <- subset(one\_df, one\_df$ANOVApValue\_p.value <= 0.01) #select only data with p<=0.01  
  
one\_df <- one\_df[-c(1,2,35)] #remove rt mz and anova p-value columns as not needed for heatmap  
  
one\_df[one\_df==0] <- NA #convert 0 to NAs  
which(is.na(one\_df), arr.ind=TRUE) #check which columns contain NAs/0s  
one\_df <- na.omit(one\_df) #remove NAs/0s  
  
#2dpi  
two\_df <- subset(two\_df, two\_df$ANOVApValue\_p.value <= 0.01) #select only data with p<=0.01  
  
two\_df <- two\_df[-c(1,2,35)] #remove rt mz and anova p-value columns as not needed for heatmap  
  
two\_df[two\_df==0] <- NA #convert 0 to NAs  
which(is.na(two\_df), arr.ind=TRUE) #check which columns contain NAs/0s  
two\_df <- na.omit(two\_df) #remove NAs/0s

#SCALE DATA

#0dpi  
m = apply(zero\_df, 1, mean, na.rm = T) #calculate mean  
  
sd = apply(zero\_df, 1, sd, na.rm = T) #calculate sd  
  
scaled\_zero\_df <- (zero\_df - m)/sd #(auto)scale  
  
#1dpi  
m = apply(one\_df, 1, mean, na.rm = T) #calculate mean  
  
sd = apply(one\_df, 1, sd, na.rm = T) #calculate sd  
  
scaled\_one\_df <- (one\_df - m)/sd #(auto)scale  
  
#2dpi  
m = apply(two\_df, 1, mean, na.rm = T) #calculate mean  
  
sd = apply(two\_df, 1, sd, na.rm = T) #calculate sd  
  
scaled\_two\_df <- (two\_df - m)/sd #(auto)scale  
  
# #note scaling is based on the scale\_rows function within pheatmap which performs the following   
# scale\_rows = function(x){  
# m = apply(x, 1, mean, na.rm = T)  
# s = apply(x, 1, sd, na.rm = T)  
# return((x - m) / s)}  
  
#it is the equivalent of autoscaling the original data in metaboanalyst

#0dpi HEATMAP

#set up save for dendrograms  
dendrograms <- "Final\_figures/zero\_dpi\_dendrogram\_ANOVA\_0.01.pdf"  
pdf(file = dendrograms)  
  
metabo\_dend <- as.dendrogram(hclust(dist(scaled\_zero\_df))) #set up gene clustering dendrogram  
#dist calculates distances using euclidean distances as default   
#hclust performs hierarchical clustering using the complete linkage method by default  
  
metabo\_dend #height = 7.467744  
  
#try different heights to cut tree to find optimal number of clusters to colour code  
metabo\_dend1 <- color\_branches(metabo\_dend, h=7.4, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend1, main = "cut height 7.4") #2 clusters = definetely too few  
  
#looking at scale on plot somewhere around 6.0 would be about right  
  
metabo\_dend2 <- color\_branches(metabo\_dend, h=6.0, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend2, main = "cut height 6.0") #6 clusters = looks good  
  
metabo\_dend3 <- color\_branches(metabo\_dend, h=5.8, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend3, main = "cut height 5.8") #6 clusters = looks good  
  
metabo\_dend4 <- color\_branches(metabo\_dend, h=5.6, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend4, main = "cut height 5.6") #8 clusters = looks good  
  
#cut height of 5.8 looks best  
  
#save optimal number of clusters as k  
cut <- cutree(metabo\_dend, h=5.8) #use h=5.6 to match best plot above  
k <- as.numeric(max(levels(as.factor(cut))))  
  
dev.off() #save dendrgrams  
  
#set up save heatmap as tiff  
heatmaps <- "Final\_figures/zero\_dpi\_heatmap\_anova\_0.01.tiff"  
tiff(filename = heatmaps)  
  
#set up sample annotations for heatmap  
sample\_annotations <- data.frame(Condition = c(rep("Mock Water", 4),  
 rep("Mock BABA", 4),  
 rep("Mock JA", 4),  
 rep("Mock SA", 4)))  
   
conditions <- HeatmapAnnotation(Condition = as.vector(sample\_annotations$Condition), col = list(Condition = c("Mock Water" = "blue", "Mock BABA" = "#117733", "Mock JA" = "#AA4499", "Mock SA" = "#DDCC77")), gp = gpar(col = "black"))  
  
  
htmp1 <- ComplexHeatmap::Heatmap(scaled\_zero\_df, cluster\_columns = FALSE,   
 cluster\_rows = metabo\_dend3, #use optimal cluster dendrogram for colour coding   
 split = k, #split up dendrogram to number clusters  
 col = viridis(7), show\_row\_names = FALSE, show\_column\_names = FALSE,  
 heatmap\_legend\_param = list(title = NULL, direction = "horizontal"),   
 top\_annotation = conditions)  
  
draw(htmp1, heatmap\_legend\_side = "bottom")  
  
dev.off() #save heatmap

#1dpi HEATMAP

#set up save for dendrograms  
dendrograms <- "Final\_figures/one\_dpi\_dendrogram\_ANOVA\_0.01.pdf"  
pdf(file = dendrograms)  
  
metabo\_dend <- as.dendrogram(hclust(dist(scaled\_one\_df))) #set up gene clustering dendrogram  
#dist calculates distances using euclidean distances as default   
#hclust performs hierarchical clustering using the complete linkage method by default  
  
metabo\_dend #height = 10.20959  
  
#try different heights to cut tree to find optimal number of clusters to colour code  
metabo\_dend1 <- color\_branches(metabo\_dend, h=10.2, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend1, main = "cut height 10.2") #2 clusters = definetely too few  
  
#looking at scale on plot somewhere around 9.5 would be about right  
  
metabo\_dend2 <- color\_branches(metabo\_dend, h=9.5, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend2, main = "cut height 9.5") #3 clusters = too few  
  
metabo\_dend3 <- color\_branches(metabo\_dend, h=9.3, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend3, main = "cut height 9.3") #5 clusters = about right, but one really small  
  
metabo\_dend4 <- color\_branches(metabo\_dend, h=9.1, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend4, main = "cut height 9.1") #5 clusters = about right, but one really small  
  
metabo\_dend2 <- color\_branches(metabo\_dend, h=9.4, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend2, main = "cut height 9.4") #3 clusters = too few  
  
metabo\_dend5 <- color\_branches(metabo\_dend, h=9.35, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend5, main = "cut height 9.35") #4 clusters = looks good, gets rid of really small one  
  
#cut height of 9.35 looks best  
  
#save optimal number of clusters as k  
cut <- cutree(metabo\_dend, h=9.35) #use h=8.9 to match best plot above  
k <- as.numeric(max(levels(as.factor(cut))))  
  
dev.off() #save dendrgrams  
  
#set up save heatmap as tiff  
heatmaps <- "Final\_figures/one\_dpi\_heatmap\_anova\_0.01.tiff"  
tiff(filename = heatmaps)  
  
#set up sample annotations for heatmap  
sample\_annotations <- data.frame(Condition = c(rep("Mock Water", 4),  
 rep("Mock BABA", 4),  
 rep("Mock JA", 4),  
 rep("Mock SA", 4),  
 rep("PM Water", 4),  
 rep("PM BABA", 4),  
 rep("PM JA", 4),  
 rep("PM SA", 4)))  
  
conditions <- HeatmapAnnotation(Condition = as.vector(sample\_annotations$Condition), col = list(Condition = c("Mock Water" = "blue", "Mock BABA" = "#117733", "Mock JA" = "#AA4499", "Mock SA" = "#DDCC77", "PM Water" = "#33bbee", "PM BABA" = "#44AA99", "PM JA" = "#882255", "PM SA" = "#999933")), gp = gpar(col = "black"))  
  
#set up column split to separate mock and infected in plots  
column\_split = rep("Mock", 32)  
column\_split[17:32] = "PM"  
  
htmp1 <- ComplexHeatmap::Heatmap(scaled\_one\_df, cluster\_columns = FALSE,   
 cluster\_rows = metabo\_dend5, #use optimal cluster dendrogram for colour coding   
 split = k, #split up dendrogram to number clusters  
 col = viridis(7), show\_row\_names = FALSE, show\_column\_names = FALSE,  
 heatmap\_legend\_param = list(title = NULL, direction = "horizontal"),   
 top\_annotation = conditions, column\_split = column\_split)  
  
draw(htmp1, heatmap\_legend\_side = "bottom")  
  
dev.off() #save heatmap

#2dpi HEATMAP

#set up save for dendrograms  
dendrograms <- "Final\_figures/two\_dpi\_dendrogram\_ANOVA\_0.01.pdf"  
pdf(file = dendrograms)  
  
metabo\_dend <- as.dendrogram(hclust(dist(scaled\_two\_df))) #set up gene clustering dendrogram  
#dist calculates distances using euclidean distances as default   
#hclust performs hierarchical clustering using the complete linkage method by default  
  
metabo\_dend #height = 10.01602   
  
#try different heights to cut tree to find optimal number of clusters to colour code  
metabo\_dend1 <- color\_branches(metabo\_dend, h=10.0, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend1, main = "cut height 10.0") #2 clusters = too few  
  
#looking at scale on plot somewhere around 8.5 would be about right  
  
metabo\_dend2 <- color\_branches(metabo\_dend, h=8.5, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend2, main = "cut height 8.5")   
#9 clusters = too many small ones  
  
metabo\_dend3 <- color\_branches(metabo\_dend, h=8.7, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend3, main = "cut height 8.7") #9 clusters = too many small ones  
  
metabo\_dend4 <- color\_branches(metabo\_dend, h=8.9, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend4, main = "cut height 8.9") #8 clusters = too many small ones  
  
metabo\_dend5 <- color\_branches(metabo\_dend, h=9.1, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend5, main = "cut height 9.1")   
#5 clusters = about right  
  
metabo\_dend4 <- color\_branches(metabo\_dend, h=9.0, col = brewer.pal(8,"Dark2"))  
plot(metabo\_dend4, main = "cut height 9.0") #6 clusters = looks perfect  
  
#cut height of 9.0 looks best  
  
#save optimal number of clusters as k  
cut <- cutree(metabo\_dend, h=9.0) #use h=9.3 to match best plot above  
k <- as.numeric(max(levels(as.factor(cut))))  
  
dev.off() #save dendrgrams  
  
#set up save heatmap as tiff  
heatmaps <- "Final\_figures/two\_dpi\_heatmaps\_anova\_0.01.tiff"  
tiff(filename = heatmaps)  
  
#set up sample annotations for heatmap  
sample\_annotations <- data.frame(Condition = c(rep("Mock Water", 4),  
 rep("Mock BABA", 4),  
 rep("Mock JA", 4),  
 rep("Mock SA", 4),  
 rep("PM Water", 4),  
 rep("PM BABA", 4),  
 rep("PM JA", 4),  
 rep("PM SA", 4)))  
  
  
conditions <- HeatmapAnnotation(Condition = as.vector(sample\_annotations$Condition), col = list(Condition = c("Mock Water" = "blue", "Mock BABA" = "#117733", "Mock JA" = "#AA4499", "Mock SA" = "#DDCC77", "PM Water" = "#33bbee", "PM BABA" = "#44AA99", "PM JA" = "#882255", "PM SA" = "#999933")), gp = gpar(col = "black"))  
  
#set up column split to separate mock and infected in plots  
column\_split = rep("Mock", 32)  
column\_split[17:32] = "PM"  
  
htmp1 <- ComplexHeatmap::Heatmap(scaled\_two\_df, cluster\_columns = FALSE,   
 cluster\_rows = metabo\_dend4, #use optimal cluster dendrogram for colour coding   
 split = k, #split up dendrogram to number clusters  
 col = viridis(7), show\_row\_names = FALSE, show\_column\_names = FALSE,  
 heatmap\_legend\_param = list(title = NULL, direction = "horizontal"),   
 top\_annotation = conditions, column\_split = column\_split)  
  
draw(htmp1, heatmap\_legend\_side = "bottom")  
  
dev.off() #save heatmap