integration\_code

## Load packages

#install.packages("usethis")   
  
#install.packages("devtools")   
  
#install.packages("plotly")   
  
#install.packages("processx")   
  
#download pRocessomic from github (see https://github.com/Valledor/pRocessomics)  
  
library("plotly") #for plotting  
  
library("usethis")   
  
library("devtools")

p <- plot\_ly(x = 1:10, y = 1:10, color = 1:10)   
  
orca(p, "plot.svg")   
  
load\_all()   
  
importfromexcel("priming")   
  
class(priming) <- "pRoDS"   
  
   
preprocess\_wizard(priming) #selecting the combination of treatments (timepoint+infection (treatment 1) and elicitor (treatment 2))   
  
#threshold 0.34 as I have 4 replicates, selection of RF using the avgIntensity   
  
#proccesing together metabolome and gene expression   
  
spls\_analysis\_wizard(priming\_preprocessed)   
  
#list to 'POL' before proceeding.   
  
mcia\_analysis\_wizard() #will guide you during the Multiple Coinertia Analysis (MCIA). This analysis links co-relationships between multiple high-dimensional datasets.   
  
spls\_analysis\_wizard() #will guide you during the sparse Partial Least Squares Analysis (SPLS) that allows you to identify the relationships between two types of omics data.   
  
bda\_analysis\_wizard() #will guide you during the DIABLO (Data Integration Analysis for Biomarker discovery using a Latent component method for Omics studies) analysis.   
  
export\_table(result\_list, filename = NULL)   
  
spls\_plot(priming\_preprocessed\_spls, treatment = 1, plottype = "Distance", variablespace = "xy", useannot = FALSE, fontsizes = c(14, 10, 16, 12), cutoff = 0.7, fortopscoring = c(1, 20, "abs"), confidence = 0.9)   
  
spls\_plot(priming\_preprocessed\_spls, treatment =2, plottype = "Network", variablespace = "xy", useannot = FALSE, fontsizes = c(14, 10, 16, 12), cutoff = 0.7, fortopscoring = c(1, 20, "abs"), confidence = 0.9)   
  
export\_plot(spls, filename = "priming\_new.pdf", width = NULL, height = NULL, cutoff = 0.7)

## Load all gene sequences

#read in fasta file of genes + their sequence as a list  
fasta <- seqinr::read.fasta("GO\_analysis/trinotate.transcripts.cdna.fa", as.string = TRUE)  
  
#convert fasta file to gene\_sequences df  
all\_gene\_sequences <- as.data.frame(fasta) #convert  
  
all\_gene\_sequences <- data.frame(sequence = unlist(all\_gene\_sequences)) #make each gene a row rather than separate cols  
  
all\_gene\_sequences$gene <- row.names(all\_gene\_sequences) #convert rownames to a column of genes

## Load genes from integration

integration\_genes <- data.frame(gene = c("Qrob\_T0061530.2",   
 "Qrob\_T0088100.2",  
 "Qrob\_T0245370.2",  
 "Qrob\_T0370930.2",  
 "Qrob\_T0157580.2",  
 "Qrob\_T0292030.2"))

## Merge sequences and integration genes for BLAST

#subset all genes to include only integration genes  
integration\_genes\_sequences <- subset(all\_gene\_sequences, gene %in% integration\_genes$gene)  
  
#save result to csv file  
write.csv(integration\_genes\_sequences, file = "integration/integration\_genes\_sequences.csv", row.names = FALSE)