#Perseus\_Like\_Analysis

#log2-Impute(MNAR)-Subtract(Median):like a Perseus

#multcomp

#https://astatsa.com/OneWay\_Anova\_with\_TukeyHSD/\_Rcode\_tutorial/

#エクセル入力

#http://www.yujitakenoshita.org/post/read-excel-in-r/

#列削除

#http://byungdugjun.blogspot.com/2014/07/r-x-image-disease-1-1-1-2-2-0-3-0-1-4-1.html

#BH法

#https://stats.biopapyrus.jp/stats/fdr-bh.html

#THSD

#https://qiita.com/hfu62/items/f9f4803828fd7e1a5cec

#My anova loop prints out the same results in R

#https://stackoverflow.com/questions/50914023/my-anova-loop-prints-out-the-same-results-in-r

#############################################################

#############################################################

#############################################################

#log2-Impute(MNAR)-Subtract(Median):like a Perseus

#############################################################

#############################################################

#############################################################

#setwd("/Users/akira\_yoshimi/Dropbox/0\_Work/R/Perseus\_Like\_Analysis/AMY") #作業ディレクトリ設定

#setwd("/Users/akira\_yoshimi/Dropbox/0\_Work/R/Perseus\_Like\_Analysis/HIP") #作業ディレクトリ設定

#setwd("/Users/akira\_yoshimi/Dropbox/0\_Work/R/Perseus\_Like\_Analysis/NAc") #作業ディレクトリ設定

#setwd("/Users/akira\_yoshimi/Dropbox/0\_Work/R/Perseus\_Like\_Analysis/PFC") #作業ディレクトリ設定

setwd("/Users/akira\_yoshimi/Dropbox/0\_Work/R/Perseus\_Like\_Analysis/STR") #作業ディレクトリ設定

getwd()#作業ディレクトリ確認

dir() #作業ディレクトリ内のファイル表示

##############################################################

#パッケージインストール

#install.packages("multcomp") #多重比較検定

#install.packages("tidyverse")

#install.packages("dplyr")

#ライブラリ読み込み

library(DEP)

library(tidyverse) #ライブラリtidyverse(ggplot2,dplyr),gcookbook読み込み

library(dplyr)

library(readxl) #エクセル入力(read\_excel)

library(xlsx) #エクセル入力

library(openxlsx) #エクセル入出力(write.xlsx)

library(writexl) #xlsx出力

library(multcomp) #多重比較検定

#library(BH) #FDR

##############################################################

#xlsx入力

rm(list = ls(all.names = TRUE))

data <- read\_excel("SWATH.xlsx", 2) #シート2(エクセルにてデータ整形)入力

ExpDesign <- read\_excel("SWATH.xlsx", 3) #シート3(DEP.packcageのSEファイル)入力

dim(data) #The data.frame dimensions:

colnames(data) #The data.frame column names:

##############################################################

#分割

split <- str\_split(data$`Peak Name`, pattern = "\\|", simplify = TRUE)

colnames(split) <- c("sp", "Protein.IDs", "GeneName") #列名変更

class(split)

x <- data.frame(split)

#文字抽出

Protein.IDs <- str\_sub(x$`Protein.IDs`, start = 1, end = 6) #`Peak Name`列の1-6文字目(Protein.IDs)抽出

Gene.names <- str\_sub(x$`GeneName`, start = 1, end = -7) #`GeneName`列の1文字目〜-7文字目(GeneName)抽出

Species <- str\_sub(x$`GeneName`, start = -5, end = -1) #`GeneName`列の-5〜-1文字目(Species)抽出

#抽出文字結合

data <- cbind(data, Protein.IDs) #dataとProtein.IDsを列ベクトル単位で結合

data <- cbind(data, Gene.names) #dataとGene.namesを列ベクトル単位で結合

data <- cbind(data, Species) #dataとSpeciesを列ベクトル単位で結合

#Search Duplication

data$Protein.IDs %>% duplicated() %>% any()

data$Gene.names %>% duplicated() %>% any()

data$Species %>% duplicated() %>% any()

#Duplication table

data %>% group\_by(Protein.IDs) %>% summarize(frequency = n()) %>% arrange(desc(frequency)) %>% filter(frequency > 1)

data %>% group\_by(Gene.names) %>% summarize(frequency = n()) %>% arrange(desc(frequency)) %>% filter(frequency > 1)

data %>% group\_by(Species) %>% summarize(frequency = n()) %>% arrange(desc(frequency)) %>% filter(frequency > 1)

#Unique Uniprot ID

data\_unique <- make\_unique(data, "Gene.names", "Protein.IDs", delim = ";")

data$name %>% duplicated() %>% any() # Are there any duplicated names?

#SummarizedExperiment

Sample\_columns <- grep("(SAL|PCP)", colnames(data\_unique)) # get Sample column numbers

experimental\_design <- ExpDesign #ExperimentalDesignSheet(label,condition,replicate)

###############################################################################

#Log2-transform

data\_se <- make\_se(data\_unique, Sample\_columns, experimental\_design) #columns=データ数, #Log2-transformation

data1 <- data.frame(data\_se@assays@data) #log2

#Impute:left-shifted Gaussian distribution (for MNAR)

data\_imp\_man <- impute(data\_se, fun = "man", shift = 1.8, scale = 0.3) #Perseus,imputation

data2 <- data.frame(data\_imp\_man@assays@data) #Subtract前log2imp

#Subtract(Median):Perseus

standardize <- function(z) {

colmed <- apply(z, 2, median) #Median of Each Sample's Protein Expression level

colmad <- apply(z, 2, mad) # median absolute deviation

rv <- sweep(z, 2, colmed,"-") #subtracting median expression

#rv <- sweep(rv, 2, colmad, "/") # dividing by median absolute deviation

return(rv)

}

data3 <- data2 #Subtract前log2impをコピー

Sample\_columns <- grep("(SC|PC)", colnames(data3)) # get Sample column numbers

data3[, Sample\_columns] <- standardize(data3[Sample\_columns]) #Subtract(Median),log2impsub

#############################################################

dat1 <- cbind(rownames(data1),data1) #log2

dat2 <- cbind(rownames(data2),data2) #log2imp

dat3 <- cbind(rownames(data3),data3) #log2impsub

#統合

dat <- cbind(data$Gene.names,data) #行名追加

dat4 <- left\_join(dat, dat1, by = c("Gene.names" = "rownames(data1)")) #raw+log2

dat4 <- left\_join(dat4, dat2, by = c("Gene.names" = "rownames(data2)")) #raw+log2+log2imp

dat4 <- left\_join(dat4, dat3, by = c("Gene.names" = "rownames(data3)")) #raw+log2+log2imp+log2impsub

#xlsx出力

smp <- list("raw"=dat,"log2"=dat1,"log2imp"=dat2,"log2impsub"=dat3,"integ"=dat4) #リスト作成,rawdata,log2,imputation,subtract,integration

write.xlsx(smp, "data.xlsx") #シート出力

#############################################################

#DEP packageによる解析

#data3のlog2impsubを2べき乗にしてimpsubにし、再度log2-transformationをDEP packageで処理

data5 <- data3

data5[Sample\_columns] <- 2^(data5[Sample\_columns]) #log2impsub→impsub

data3impsub <- data5 #log2impsub→impsub

#抽出文字結合

data3impsub <- cbind(data3impsub, Protein.IDs) #data3impsubとProtein.IDsを列ベクトル単位で結合

data3impsub <- cbind(data3impsub, Gene.names) #data3impsubとGene.namesを列ベクトル単位で結合

data3impsub <- cbind(data3impsub, Species) #data3impsubとSpeciesを列ベクトル単位で結合

#Unique Uniprot ID

data3impsub\_unique <- make\_unique(data3impsub, "Gene.names", "Protein.IDs", delim = ";")

#SummarizedExperiment

ExpDesign2 <- data.frame(cbind(ExpDesign$No,data.frame(list(colnames(data3impsub[Sample\_columns]))))) #ExpDesignにデータ解析後の情報を上書き

ExpDesign2 <- cbind(ExpDesign2,ExpDesign[,3:4]) #新たなExpDesignを作成

colnames(ExpDesign2) <- colnames(ExpDesign) #列名を修正

#txt出力

write.table (ExpDesign2, file = "ExpDesign2.txt", sep = "\t") #保存

#txt入力

ExpDesign2 <- read.table("ExpDesign2.txt",header=T, sep="\t", stringsAsFactors = F) #再入力

experimental\_design2 <- ExpDesign2 #ExperimentalDesignSheet(label,condition,replicate)

###############################################################################

#Log2-transform

data3impsub\_se <- make\_se(data3impsub\_unique, Sample\_columns, experimental\_design2) #columns=データ数, #Log2-transformation

data3log2impsub <- data.frame(data3impsub\_se@assays@data) #log2impsubに戻す=data3

#############################################################

#############################################################

#############################################################

#multcomp

#############################################################

#############################################################

#############################################################

data\_rm <- data3

data\_rm[,1:2] <- NULL #列削除

#############################################################

group <- read\_excel("SWATH.xlsx", 4) #シート4(G)入力

PC <- factor(group$PC, levels = c("SC0", "SC10", "SC30", "PC0", "PC10", "PC30"))

P <- factor(group$P, levels = c("S", "P"))

C <- factor(group$C, levels = c("C0", "C10", "C30"))

g <- cbind(PC,P,C)

#############################################################

#1wANOVA function

#############################################################

aof <- function(x) {

m <- data.frame(PC, x);

anova(aov(x ~ PC, m))

}

# apply analysis to the data and get the pvalues.

onewayANOVA <- apply(data\_rm, 1, aof)

onewayANOVAp <- data.frame(lapply(onewayANOVA, function(x) { x["Pr(>F)"][1,] }))

onewayANOVAp2 <- data.frame(t(onewayANOVAp))

colnames(onewayANOVAp2) <- "p\_PC" #列名の変更

#############################################################

#2wANOVA function

#############################################################

aof2 <- function(x) {

n <- data.frame(P,C, x);

anova(aov(x ~ P + C + P\*C, n))

}

# apply analysis to the data and get the pvalues.

twowayANOVA <- apply(data\_rm, 1, aof2)

twowayANOVAp <- data.frame(lapply(twowayANOVA, function(x) { x["Pr(>F)"][1:3,] }))

twowayANOVAp2 <- data.frame(t(twowayANOVAp))

colnames(twowayANOVAp2) <- c("p\_P","p\_C","p\_PxC") #列名の取得

sdata <- cbind(data\_rm,c(onewayANOVAp2, twowayANOVAp2))

#############################################################

#2wANOVA BH-FDR

#############################################################

#p値

p\_PC <- sdata$p\_PC

p\_P <- sdata$p\_P

p\_C <- sdata$p\_C

p\_PxC <- sdata$p\_PxC

checkP <- data.frame(cbind(p\_PC, p\_P, p\_C, p\_PxC))

rownames(checkP) <- rownames(data3)

checkPr <- cbind(rownames(checkP),checkP)

#q値

q\_PC <- data.frame(p.adjust(p\_PC, method = "BH"))

q\_P <- data.frame(p.adjust(p\_P, method = "BH"))

q\_C <- data.frame(p.adjust(p\_C, method = "BH"))

q\_PxC <- data.frame(p.adjust(p\_PxC, method = "BH"))

checkQ <- data.frame(cbind(q\_PC, q\_P, q\_C, q\_PxC))

colnames(checkQ) <- c("q\_PC", "q\_P", "q\_C","q\_PxC") #列名の取得

rownames(checkQ) <- rownames(data3)

checkQr <- cbind(rownames(checkQ),checkQ)

sdata <- cbind(sdata, checkQ)

#############################################################

#TukeyHSD function

#diff群間の平均値の差(例)B-Aが-127.3であればデータBの平均がデータAの平均より-127.3大きい

#lwr,upr=下方信頼限界,情報信頼限界:信頼区間の下限値 (lower) と上限値 (upper)

#0を含まない場合 (例)B-A は含まず D-A は含む=2群間差は0ではないので有意差あり

#p.adj < 0.05=2群間に有意差あり(信頼区間内に0を含まない)

#############################################################

THSD <- function(x) {

nn <- data.frame(P,C, x);

TukeyHSD(aov(x ~ P + C + P\*C, nn))

}

THSDresults <- apply(data\_rm, 1, THSD)

THSD\_PC <- data.frame(lapply(THSDresults, function(x) {x["P:C"]}))

#############################################################

#############################################################

#############################################################

#エラーが出るのでtidyverse再インストール,再読み込み

install.packages("tidyverse") #←←←←←←←←←←←←←←←←←←←←←←stop←←←←←

#############################################################

#############################################################

#############################################################

library(tidyverse) #ライブラリtidyverse(ggplot2,dplyr),gcookbook読み込み

THSDp\_PC <- select(THSD\_PC, ends\_with("p.adj")) #p値抽出

THSDd\_PC <- select(THSD\_PC, ends\_with(".diff")) #diff値抽出

#transpose

THSDp\_PC2 <- data.frame(t(THSDp\_PC))

THSDd\_PC2 <- data.frame(t(THSDd\_PC))

#列名変更

colnames(THSDp\_PC2) <- str\_c("THSDp", colnames(THSDp\_PC2), sep="\_")

colnames(THSDd\_PC2) <- str\_c("diff", colnames(THSDd\_PC2), sep="\_")

#結合

THSDpd <- cbind(THSDp\_PC2, THSDd\_PC2)

rownames(THSDpd) <- rownames(data3)

THSDpd <- cbind(rownames(THSDpd),THSDpd)

sdata <- cbind(sdata, THSDpd)

#############################################################

#Annotation

#############################################################

sdata2 <- cbind(rownames(sdata),sdata)

#抽出文字結合

sdata2 <- cbind(sdata2, Protein.IDs) #sdata2とProtein.IDsを列ベクトル単位で結合

sdata2 <- cbind(sdata2, Gene.names) #sdata2とGene.namesを列ベクトル単位で結合

sdata2 <- cbind(sdata2, Species) #sdata2とSpeciesを列ベクトル単位で結合

sdata2 <- left\_join(sdata2, data.frame(data[,grep("(Protein.IDs|Group)", colnames(data))]), by = "Protein.IDs")

#############################################################

#xlsx出力

library(writexl) #xlsx出力

sheets <- list("integ" = sdata2, "anovap" = checkPr, "anovaq" = checkQr, "THSDpd" = THSDpd) #assume sheet1-4 are data frames

write\_xlsx(sheets, "stat.xlsx", format\_headers = FALSE)

#txt出力

#write.table (sdata2, file = "integ.txt", sep = "\t") #保存

#############################################################

# ココで中断した場合以下から再開

#############################################################

#txt入力

#sdata2 <- read.table("integ.txt",header=T, sep="\t", stringsAsFactors = F)

#dim(sdata2) #The data.frame dimensions:

#colnames(sdata2) #The data.frame column names:

#############################################################

#DEPリスト作成

#############################################################

twANOVA\_Pq005 <- sdata2 %>% filter(Species == "MOUSE") %>% filter(q\_P < 0.05)

twANOVA\_Cq005 <- sdata2 %>% filter(Species == "MOUSE") %>% filter(q\_C < 0.05)

twANOVA\_PxCq005 <- sdata2 %>% filter(Species == "MOUSE") %>% filter(q\_PxC < 0.05)

sheets2 <- list("Pq005"=twANOVA\_Pq005[,grep("(rownames.sdata.|p\_P$|p\_C$|p\_PxC$|q\_P$|q\_C$|q\_PxC$|Protein.IDs|Gene.names|Species|Group)", colnames(twANOVA\_Pq005))],

"Cq005"=twANOVA\_Cq005[,grep("(rownames.sdata.|p\_P$|p\_C$|p\_PxC$|q\_P$|q\_C$|q\_PxC$|Protein.IDs|Gene.names|Species|Group)", colnames(twANOVA\_Cq005))],

"PxCq005"=twANOVA\_PxCq005[,grep("(rownames.sdata.|p\_P$|p\_C$|p\_PxC$|q\_P$|q\_C$|q\_PxC$|Protein.IDs|Gene.names|Species|Group)", colnames(twANOVA\_PxCq005))])

write\_xlsx(sheets2, "DEPtwANOVA.xlsx", format\_headers = FALSE)

#############################################################