**1.Difference analysis (Wilcoxon rank-sum test)**

library(WilcoxCV)

########## Read data file

data\_path = "CH.txt"

result\_path = "CH\_wilcoxon\_reslut\_sn s.txt"

index\_group\_1 =1:24

index\_group\_2 =25:75

data = read.delim(data\_path, row.names=1)

########## Run wilcoxon test and save results

feature\_list = c("ID", "Mean\_group1", "variance\_group1", "std\_group1", "Mean\_group2", "variance\_group2", "std\_group2", "p", "q")

matrix\_result <- array(0, dim=c((nrow(data) + 1), length(feature\_list)))

matrix\_result[1, ] = feature\_list

for (i in 1:nrow(data)){

group\_1 = as.numeric(data[i, index\_group\_1])

group\_2 = as.numeric(data[i, index\_group\_2])

result\_wilcoxon = wilcox.test(group\_1, group\_2,)

p\_wilcoxon = result\_wilcoxon$p.value

matrix\_result[i + 1, 1:length(feature\_list)-1] = c(rownames(data[i, ]), mean(group\_1), var(group\_1), sd(group\_1), mean(group\_2), var(group\_2), sd(group\_2), p\_wilcoxon)

}

matrix\_result[2:(nrow(data)+1),length(feature\_list)] = p.adjust(matrix\_result[2:(nrow(data)+1),length(feature\_list)-1], method = "fdr")

write(t(matrix\_result), file=result\_path, ncolumns=length(feature\_list), sep="\t")

**2. Correlation analysis (method 1)**

library(psych)

library(stringr)

library(Hmisc)

library(pheatmap)

z<-scale(read.delim(file="clipboard",header=T,row.names=1,sep="\t"))

correlation\_results<-corr.test(z,method="spearman",adjust="fdr")

r<-correlation\_results$r

p<-correlation\_results$p

write.csv(r,"D:/r.csv",row.names =T)

write.csv(p,"D:/p.csv",row.names =T)

**3. Correlation analysis (method 2)**

# Read data file

dat <- read.table('AB.txt', sep = '\t', row.names = 1, header = TRUE, stringsAsFactors = FALSE, check.names = FALSE)

dat\_env <- dat[1:5]

dat\_phylum <- dat[6:48]

phylum\_env\_spearman <- cor(dat\_phylum, dat\_env, method = 'spearman')

phylum\_env\_spearman [abs(phylum\_env\_spearman) <= 0.5] <- 0

write.csv(phylum\_env\_spearman, 'phylum\_env\_spearman.csv', quote = FALSE)

**4. Heatmap(method 1)**

# Read data file

spearman2 <- as.matrix(read.table('spearman2.txt', sep = '\t', row.names = 1, header = TRUE))

library(corrplot)

#Default style

corrplot(spearman2)

corrplot(spearman2, method = 'color')

# Adjust parameters and change the display style of correlation coefficients

corrplot(spearman2, method = 'color')

corrplot(spearman2, method = 'number')

corrplot(spearman2, method = 'pie')

**5. Heatmap(method 2)**

library(pheatmap)

z<-read.delim(file="clipboard",header=T,row.names=1,sep="\t")

pheatmap(z, scale="row",cluster\_cols=FALSE,

color = colorRampPalette(colors = c("Navy","white","firebrick3"))(100))

**6. Tax4Fun anlysis**

Library (Tax4Fun)

QIIMESingleData <- importQIIMEData("otu\_table.txt ")

folderReferenceData <- 'SILVA123'

Tax4FunOutput <- Tax4Fun(QIIMESingleData, folderReferenceData, fctProfiling = TRUE, refProfile = 'UProC', shortReadMode = TRUE, normCopyNo = TRUE)

tax4fun\_gene <- as.data.frame(t(Tax4FunOutput$Tax4FunProfile))

gene <- rownames(tax4fun\_gene)

write.table(cbind(gene, tax4fun\_gene), 'tax4fun.gene.txt', row.names = FALSE, sep = '\t', quote = FALSE)

#KEGG Abundance prediction of metabolic pathway (KO level 3)

Tax4FunOutput <- Tax4Fun(QIIMESingleData, folderReferenceData, fctProfiling = FALSE, refProfile = 'UProC', shortReadMode = TRUE, normCopyNo = TRUE)

tax4fun\_pathway <- as.data.frame(t(Tax4FunOutput$Tax4FunProfile))

pathway <- rownames(tax4fun\_pathway)

write.table(cbind(pathway, tax4fun\_pathway), 'tax4fun.pathway.txt', row.names = FALSE, sep = '\t', quote = FALSE)

# Map KO classification for KEGG metabolic pathway

kegg\_anno <- read.delim('pathway.anno.txt', sep = '\t', colClasses = 'character', check.names = FALSE)

tax4fun\_pathway$KO3\_id <- t(data.frame(strsplit(rownames(tax4fun\_pathway), ';')))[ ,1]

tax4fun\_pathway$KO3\_id <- gsub('ko', '', tax4fun\_pathway$KO3\_id)

tax4fun\_pathway <- merge(tax4fun\_pathway, kegg\_anno, by = 'KO3\_id')

write.table(tax4fun\_pathway, 'tax4fun.pathway.anno.txt', row.names = FALSE, sep = '\t', quote = FALSE)

# Statistical summation at the second level of KO, and add sample grouping information

group <- read.delim('group1.txt', sep = '\t', stringsAsFactors = FALSE, check.names = FALSE)

names(group)[1] <- 'variable'

tax4fun\_pathway <- tax4fun\_pathway[c(group$variable, 'KO2\_id')]

tax4fun\_pathway <- reshape2::melt(tax4fun\_pathway, id = 'KO2\_id')

tax4fun\_pathway <- doBy::summaryBy(value~variable+KO2\_id, tax4fun\_pathway, FUN = sum)

tax4fun\_pathway <- merge(tax4fun\_pathway, group, by = 'variable')

#Calculate the mean (mean) ± standard deviation (sd), or standard error (se)

se <- function(x) sd(x) / (length(x))^0.5

pathway\_stat <- doBy::summaryBy(value.sum~group+KO2\_id, tax4fun\_pathway, FUN = c(mean, sd, se))

##add notes

kegg\_anno\_2 <- kegg\_anno[!duplicated(kegg\_anno$KO2), ][-c(5:7)]

pathway\_stat <- merge(pathway\_stat, kegg\_anno\_2, by = 'KO2\_id', all.x = TRUE)

write.table(pathway\_stat, 'tax4fun.KO2.anno.txt', row.names = FALSE, sep = '\t', quote = FALSE)

# Significant difference analysis, here directly use non-parametric wilcoxon to test the difference between the two groups

KO2\_id <- unique(tax4fun\_pathway$KO2\_id)

for (i in KO2\_id) {

tax4fun\_pathway\_2\_i <- subset(tax4fun\_pathway, KO2\_id == i)

test <- wilcox.test(value.sum~group, tax4fun\_pathway\_2\_i)

line\_t <- which(pathway\_stat$KO2\_id == i & pathway\_stat$group == 'treat')

pathway\_stat[line\_t,'p.value'] <- test$p.value

if (test$p.value < 0.05 & test$p.value >= 0.01) {

pathway\_stat[line\_t,'sign'] <- '\*'

}

if (test$p.value < 0.01 & test$p.value >= 0.001) {

pathway\_stat[line\_t,'sign'] <- '\*\*'

}

if (test$p.value < 0.001) {

pathway\_stat[line\_t,'sign'] <- '\*\*\*'

}

}

write.table(pathway\_stat, 'tax4fun.KO2.anno\_stat.txt', row.names = FALSE, sep = '\t', quote = FALSE, na = '')

#Use ggplot2 to plot

library(ggplot2)

pathway\_stat <- pathway\_stat[order(pathway\_stat$value.sum.mean), ]

pathway\_stat$KO2 <- factor(pathway\_stat$KO2, levels = unique(pathway\_stat$KO2))

pathway\_stat$KO1 <- factor(pathway\_stat$KO1, levels = rev(unique(pathway\_stat$KO1)))

pathway\_stat$value.sum.mean <- 100 \* pathway\_stat$value.sum.mean

pathway\_stat$value.sum.sd <- 100 \* pathway\_stat$value.sum.sd

ko2\_plot <- ggplot(pathway\_stat, aes(x = KO2, y = value.sum.mean, fill = group)) +

geom\_col(position = 'dodge', width = 0.8, colour = 'black', size = 0.05) + #“dodge Histogram" style

scale\_fill\_manual(values = c('red', 'blue')) + # Fill color

geom\_errorbar(aes(ymin = value.sum.mean - value.sum.sd, ymax = value.sum.mean + value.sum.sd),

size = 0.05, width = 0.35, position = position\_dodge(width = 0.8)) + # Add error bars (mean ± standard deviation)

geom\_text(aes(label = sign, y = value.sum.mean + value.sum.sd + 0.5), size = 4, position = position\_dodge(0.8)) + # Add a distinctive mark “\*”

facet\_grid(KO1~., space = 'free', scale = 'free\_y') +

coord\_flip() + # The horizontal and vertical axes are reversed

theme(panel.grid = element\_blank(), panel.background = element\_rect(fill = 'transparent', color = 'black'),

legend.title = element\_blank(), legend.position = 'top') +

labs(x = 'KEGG Orthology (KO)', y = 'Relative Abundance (%)')

ggsave('ko2\_stat.pdf', ko2\_plot, width = 8, height = 10)

ggsave('ko2\_stat.png', ko2\_plot, width = 8, height = 10)

**7. ######Alpha Diversity####**

library(spaa)

library(vegan)

z<-read.delim(file="clipboard",header=T,row.names=1,sep="\t")

shannon.Wiener <- diversity(z, index = "shannon")

Simpson <- diversity(z, index = "simpson")

Inverse.Simpson <- diversity(z, index = "inv")

S <- specnumber(z)

plot(S)

J <- shannon.Wiener/log(S)

write.csv(shannon.Wiener,"D:/shannon.Wiener.csv",row.names =T)

write.csv(Simpson,"D:/Simpson.csv",row.names =T)

write.csv(J,"D:/J.csv",row.names =T)