**Supplemental S-2, The R-code used to perform the PCA analysis**

#based on http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/118-principal-component-analysis-in-r-prcomp-vs-princomp/

#install.packages

install.packages("factoextra")

install.packages("PerformanceAnalytics")

library("PerformanceAnalytics")

library(factoextra)

library(car)

library(rgl)

library(ggplot2)

library(RColorBrewer)

#data

data <- read.table(file = "Data location", header = TRUE, sep = "\t", as.is = TRUE, row.names=1)

#1:30 Liver

#31:60 CRLM

#61:74 colon

#75:88 CRC

groups <- c(rep("liver",30),rep("CRLM",30),rep("colon",14),rep("CRC",14)); data\_red <- data

groups <- c(rep("colon",14),rep("CRC",14)); data\_red <- data[c(61:88),]

groups <- c(rep("Liver",30),rep("CRLM",30)); data\_red <- data[c(1:60),]

groups <- c(rep("CRLM",30),rep("CRC",14)); data\_red <- data[c(31:60,75:88),]

groups <- c(rep("liver",30),rep("Colon",14)); data\_red <- data[c(1:30,61:74),]

groups <- c(rep("CRLM",30), rep("CRC",14)); data\_red <- data[c(31:60,75:88),]

groups <- c(rep("liver",30)); data\_red <- data[c(1:30),]

groups <- c(rep("CRLM",30)); data\_red <- data[c(31:60),]

groups <- c(rep("Colon",14)); data\_red <- data[c(61:74),]

groups <- c(rep("CRC",14)); data\_red <- data[c(75:88),]

#14 vs 14 matched

data <- read.table(file = "Data location", header = TRUE, sep = "\t", as.is = TRUE, row.names=1)

#1:14 Liver

#15:28 CRLM

#29:42 colon

#43:56 CRC

groups <- c(rep("liver",14),rep("CRLM",14),rep("colon",14),rep("CRC",14)); data\_red <- data

groups <- c(rep("colon",14),rep("CRC",14)); data\_red <- data[c(29:56),]

groups <- c(rep("Liver",14),rep("CRLM",14)); data\_red <- data[c(1:28),]

groups <- c(rep("CRLM",14),rep("CRC",14)); data\_red <- data[c(15:28,43:56),]

groups <- c(rep("liver",14),rep("Colon",14)); data\_red <- data[c(1:14,29:42),]

groups <- c(rep("CRLM",14), rep("CRC",14)); data\_red <- data[c(15:28,43:56),]

groups <- c(rep("colon",14), rep("CRLM",14)); data\_red <- data[c(29:42,15:28),]

#Preparing data

data\_red <- data\_red[,colSums(data\_red != 0) > 0] #remove columns that are empty.

data\_red.pca <- prcomp(data\_red, center = TRUE) #no scale. = TRUE, data is 0 or 1

#PCA results

summary(data\_red.pca)

fviz\_eig(data\_red.pca)

fviz\_contrib(data\_red.pca, choice="var", axes =1:2, invisible = "labels") #plots contribution of each protein to the PCA model axes = ? tells to which dimension

fviz\_contrib(data\_red.pca, choice="var", axes =1:2, top=25) + #plots contribution of each protein to the PCA model axes = ? tells to which dimension

 theme(axis.text.x = element\_text(angle = 90))

#contribution of var

fviz\_pca\_var(data\_red.pca,

 label="none",

 geom = "point",

 col.var="contrib")+

 scale\_color\_gradient2(low="white", mid="blue",high="red", midpoint=1.5)

#PCA

fviz\_pca\_ind(data\_red.pca,

 repel=TRUE,

 mean.point=FALSE,

 habillage = groups,

 addEllipses = TRUE,

 title = "Hydrox variation",

 label = "none",

 pointsize = 2)

#PCA w/o groups, with labels

fviz\_pca\_ind(data\_red.pca,

 repel=TRUE,

 invisible = "quali",

 addEllipses = TRUE,

 title = "Hydrox variation")

#3d PCA

fviz\_contrib(data\_red.pca, choice="var", axes =1:3, top=25) + #plots contribution of each protein to the PCA model axes = ? tells to which dimension

 theme(axis.text.x = element\_text(angle = 90))

data\_red.3dpca <- get\_pca\_ind(data\_red.pca)$coord

x <- data\_red.3dpca[,1]

y <- data\_red.3dpca[,2]

z <- data\_red.3dpca[,3]

scatter3d(x,y,z,

 groups = as.factor(groups),

 surface.col = c("purple", "blue", "green", "orange"),

 surface = FALSE,

 ellipsoid = TRUE,

 grid=TRUE,

 xlab = "dim.1", ylab= "dim.2", zlab="dim.3")

rgl.viewpoint(0, 20)

for( i in seq(1,720, by=1.5)){

 rgl.viewpoint(i,i\*0.04)

}

#Contribution of in the individuals

fviz\_pca\_ind(data\_red.pca,

 repel=TRUE,

 col.ind="cos2",

 title = "Hydrox variation",

 label = "none")+

 scale\_color\_gradient2(low="white", mid="blue",high="red", midpoint=0.3)

#Mean relation

cor\_m <- read.table(file ="clipboard", sep="\t", header = TRUE)

chart.Correlation(cor\_m, histogram = TRUE, pch=19)

plot(cor\_m[,1], cor\_m[,2], col = "green")

points(cor\_m[,1], cor\_m[,3], col = "blue")

points(cor\_m[,1], cor\_m[,4], col = "red")

hist\_m <- read.table(file = "clipboard", sep ="\t", header = TRUE)

hist(hist\_m[,1], col = rgb(1,0,0,0.5), ylim = c(0,1), freq = F)$count

hist(hist\_m[,2], col = rgb(0,1,0,0.5), add = T, freq = F)$count

hist(hist\_m[,3], col = rgb(0,0,1,0.5), add = T, freq = F)$count

hist(hist\_m[,4], col = rgb(1,1,0,0.5), add = T, freq = F)$count

#Pretty figure of PCA result

#Liver/CRLM data

age <- c(74,74,62,68,65,77,74,76,43,86,73,73,76,60,64,70,36,77,49,73,71,66,61,58,64,69,78,54,78,59)

gender <- c("M","M","M","F","M","M","F","M","M","M","M","M","F","M","M","F","M","M","F","F","M","M","F","M","F","M","M","F","F","M")

location <- c("Sigmoïd","Sigmoïd" ,"Rectum" ,"Rectum" ,"Rectum" ,"Rectum" ,"Rectum" ,"Colon descendant" ,

 "Rectum" ,"Rectum" ,"Colon ascendant" ,"Colon ascendant" ,"Sigmoïd" ,"Rectum" ,"Rectum" ,

 "Sigmoïd" ,"Sigmoïd" ,"Rectum" ,"Rectum" ,"Rectum" ,"Rectum" ,"Sigmoïd" ,"Sigmoïd" ,"Sigmoïd" ,

 "Colon descendant" ,"Sigmoïd" ,"Rectum" ,"Rectum" ,"Rectum" ,"Sigmoïd")

#Colon/CRC data

age <- c(61, 52, 44, 63, 75, 56, 61, 84, 74, 74, 64, 69, 69, 43)

gender <- c("M", "F", "F", "M", "M", "M", "M", "M", "F", "F", "M", "F", "M", "M")

location <- c("Rectum","Rectum" ,"Rectum" ,"Rectum" ,"Rectum" ,"Sigmoïd" ,"Rectum" ,"Rectum" ,"Sigmoïd" ,"Rectum" ,"Sigmoïd" ,"Sigmoïd" ,"Sigmoïd" ,"Rectum")

#x,y originate from 3d PCA section

dataGG <- data.frame(x,y,location,gender,age) #Tumor data frame is required for ggplot

dataGG <- data.frame(x,y,gender,age) #Healthy data frame is required for ggplot

#healthy

ggplot(data = dataGG, aes(x = x, y = y)) +

 geom\_point(aes(col = age, size = gender)) +

 scale\_size\_manual(values = c(3,6)) +

 scale\_color\_continuous(low = "yellow", high = "red") +

 xlim(-4,4) + ylim(-3,3)

#tumor

ggplot(data = dataGG, aes(x = x, y = y)) +

 geom\_point(aes(col = age, size = gender, shape = location)) + #remove location

 scale\_size\_manual(values = c(3,6)) +

 scale\_color\_continuous(low = "yellow", high = "red") +

 xlim(-2,6) + ylim(-2,3)