Loading library

```{r}
library(biomaRt)
library(ComplexHeatmap)
library(factoextra)
library(FactoMineR)
library(ggplot2)
library(ggpubr)
```

First, download RNA-seq dataset from the ImmPort repository, accession number SDY939 (https://www.immport.org/immport-open/public/study/study/displayStudyDetail/SDY939)
Name of RNA-seq dataset is "kallisto_ensembl83.612606.tsv"

```{r}

rawdata <- read.delim("/path/to/kallisto_ensembl83.612606.tsv", header=TRUE)
head(rawdata)
gene <- rawdata$Gene



head(listMarts())
db <- useMart("ensembl")
head(listDatasets(db))
hg <- useDataset("hsapiens_gene_ensembl", mart = db)

res <- getBM(attributes = c("ensembl_gene_id", "start_position", "end_position","hgnc_symbol"),
                filters = "ensembl_gene_id",
                values = gene,
                mart = hg)

df <- merge(res,rawdata,by.x="ensembl_gene_id", by.y="Gene")
bool <- duplicated(df$ensembl_gene_id) #removing genes overlapping ensembl_gene_id
```

```
df2 <- df[! bool,]
rownames(df2) <- df2$ensembl_gene_id
df2 <- data.frame(t(df2[,-(1:4)]))
df2[,1:5]
cell3 <-
rep(c("CXCR5[-]PD1[-]","Tph","Tph","Tph","Tph","CXCR5[+]PD1[-]","Tfh","Tfh"),4)
df2 <- data.frame(cell3,df2)
df2 <- filter(df2, cell3 %in% c("Tph","Tfh"))
```

PLS-DA

```{r}

set.seed(71)
opls_res <- opls(df2[,-1],
 as.character(df2$cell3),
 predI=NA,
 orthoI=0,
 permI=500,
 crossvalI=6,
 scaleC="standard",
 printL=FALSE,
 plotL=FALSE
)

print(opls_res)
layout(matrix(1:4, nrow=2, byrow=TRUE))
for(typeC in c("x-score", "overview","permutation","outlier"))
  plot(opls_res, typeVc=typeC,
       parDevNewL=FALSE
  )


```

Subtracting genes with highly variable importance score (>1.5) and divide Tph-signature genes and Tfh-signature genes

```{r}

df_VIP <- data.frame(t(df2[, -which (colnames(df2) %in% c("cell3"))]),opls_res@vipVn)
df_VIP_1.5 <- as.data.frame(t(subset(df_VIP, opls_res.vipVn > 1.5))) # ubtracting genes with VIP score > 1.5
df_VIP_1.5 <- df_VIP_1.5[-which (rownames(df_VIP_1.5) %in% c("opls_res.vipVn")),]

df_VIP_1.5 <- cbind(df2$cell3,df_VIP_1.5)
names(df_VIP_1.5)[1] <- "cell"
Heatmap(t(scale(df_VIP_1.5[,-1])),
        column_title = "VIP score > 2",
        col = colorRamp2(c(-1.5, 0, 1.5), c("blue", "white", "red")),
        name = "scaled_expr",
        show_column_names = FALSE, width = unit(8, "cm"),show_row_names = FALSE,
        heatmap_legend_param = list(title = "Scaled expr"),
        cluster_columns = TRUE,cluster_rows = TRUE,
        clustering_method_rows = "ward.D2", clustering_distance_rows = "euclidean",
        top_annotation = ha)
res.pca <- prcomp(df_VIP_1.5[,-1], scale = TRUE)
fviz_pca_ind(res.pca, title = "PCA",
             col.ind = cell2,
             addEllipses = TRUE, # Concentration ellipses
             ellipse.level=0.95,ellipse.type = "confidence",
             legend.title = "Groups",
             repel = TRUE,
             invisible = "quali", geom.ind="point",pointsize=3, palette = c("hotpink","red",

"skyblue", "blue"))  + ggtitle("VIP score > 1.5")

data <- data.frame(opls_res@coefficientMN[names(df_VIP_1.5[-1]),],opls_res@vipVn[names(df_VIP_1.5[-1])])
data$color <- ifelse(data$opls_res.coefficientMN.names.df_VIP_1.5..1.....>0, "Tph", "Tfh")
```

```
names(data) <- c("coefficient","VIP_score","color")
data_Tph <- subset(data, color=="Tph")
data_Tfh <- subset(data, color=="Tfh")
Heatmap(t(scale(df_VIP_1.5[,rownames(data_Tph)])),
        column_title = "Tph-signature genes¥nwith VIP score > 1.5",
        col = colorRamp2(c(-1.5, 0, 1.5), c("blue", "white", "red")),
        name = "scaled_expr",
        show_column_names = FALSE, width = unit(8, "cm"),show_row_names = FALSE,
        heatmap_legend_param = list(title = "Scaled expr"),
        cluster_columns = TRUE,cluster_rows = TRUE,
        clustering_method_rows = "ward.D2", clustering_distance_rows = "euclidean",
        top_annotation = ha)
Heatmap(t(scale(df_VIP_1.5[,rownames(data_Tfh)])),
        column_title = "Tfh-signature genes¥nwith VIP score > 1.5",
        col = colorRamp2(c(-1.5, 0, 1.5), c("blue", "white", "red")),
        name = "scaled_expr",
        show_column_names = FALSE, width = unit(8, "cm"),show_row_names = FALSE,
        heatmap_legend_param = list(title = "Scaled expr"),
        cluster_columns = TRUE,cluster_rows = TRUE,
        clustering_method_rows = "ward.D2", clustering_distance_rows = "euclidean",
        top_annotation = ha)

data_Tph <- merge(res,data_Tph,by.x="ensembl_gene_id", by.y="row.names")
head(data_Tph)
dim(data_Tph)
write.csv(data_Tph,"Tph_signature_genes_VIP1.5.csv")
data_Tfh <- merge(res,data_Tfh,by.x="ensembl_gene_id", by.y="row.names")
head(data_Tfh)
dim(data_Tfh)
write.csv(data_Tfh,"Tfh_signature_genes_VIP1.5.csv")

```
```
```