

Enio Gjerga - R Notebook for the analysis of Hong etal 2020 data.

This is an R Markdown Notebook for the Hong etal 2020 manuscript (submitted in International Journal of Cancer). When you execute code within the notebook, the results appear beneath the code.

R code for the analysis

```
## Loading the packages required
library(readr)
library(ggplot2)
library(reshape)
library(pheatmap)
library(gridExtra)
library(grid)
library(cowplot)

##
## *****

## Note: As of version 1.0.0, cowplot does not change the
## default ggplot2 theme anymore. To recover the previous
## behavior, execute:
## theme_set(theme_cowplot())

## *****

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:reshape':
##
## stamp
library(ggrepel)

## Loading the data and creating sample groups
load(file = "src/Yinfeng_processed_normalized_gene_counts_wt.RData")
groups <- c("GM-CSF", "M-CSF", "Hox", "GM-CSF", "M-CSF", "Hox", "GM-CSF", "M-CSF", "Hox")
names(groups) <- colnames(data)

## PCA analysis of samples
data.pca <- t(data)
data.pca <- cbind(data.pca, as.matrix(groups))
colnames(data.pca)[ncol(data.pca)] <- "Group"
data.pca <- as.data.frame(data.pca)
data.pca[, 1:(ncol(data.pca)-1)] <- lapply(data.pca[, 1:(ncol(data.pca)-1)],
                                           function(x) as.numeric(as.character(x)))

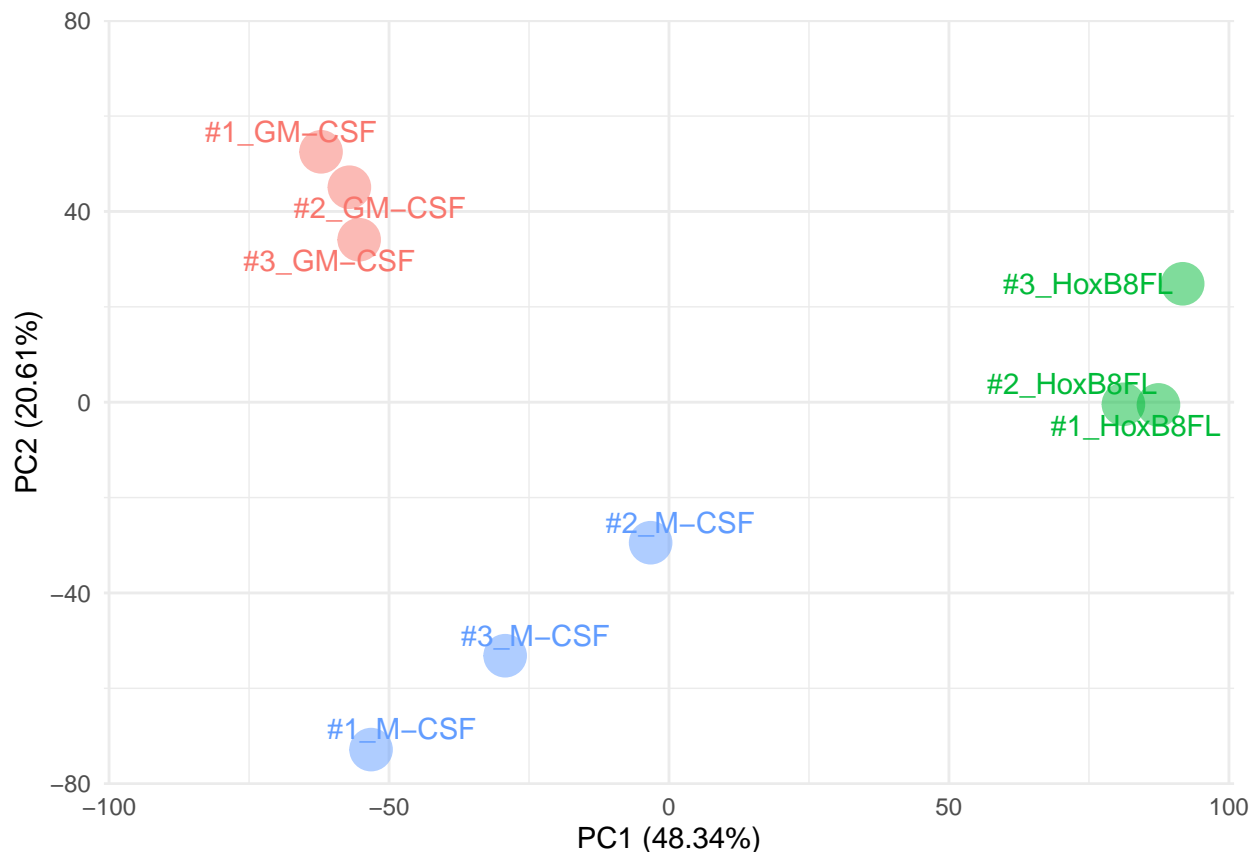
res.pca <- prcomp(data.pca[, -ncol(data.pca)], scale. = TRUE)
res.plot <- as.data.frame(cbind(res.pca$x[, 1], res.pca$x[, 2],
                                as.character(data.pca$Group), rownames(data.pca)))
```

```
res.plot[, 1:2] <- lapply(res.plot[, 1:2], function(x) as.numeric(as.character(x)))
res.plot[, 3:4] <- lapply(res.plot[, 3:4], function(x) as.character(x))
colnames(res.plot) <- c("pc1", "pc2", "Group", "sample")
percentages <- ((res.pca$sdev)^2 / sum(res.pca$sdev^2)*100)[1:2]
```

```
pp <- ggplot(res.plot, aes(x=pc1, y=pc2, color=Group)) +
  geom_point(size=7, alpha = 0.5) +
  scale_alpha_discrete(range=c(0.3, 1.0)) +
  #geom_path(arrow=arrow()) +
  theme_minimal() +
  xlab(paste0("PC1 (", round(x = percentages[1], digits = 2), "%)")) +
  ylab(paste0("PC2 (", round(x = percentages[2], digits = 2), "%)")) +
  xlim(c(-max(abs(res.pca$x[, 1])), max(abs(res.pca$x[, 1])))) +
  ylim(c(-max(abs(res.pca$x[, 2])), max(abs(res.pca$x[, 2])))) +
  theme(legend.position = "none") +
  geom_text_repel(data = res.plot, aes(label=sample))
```

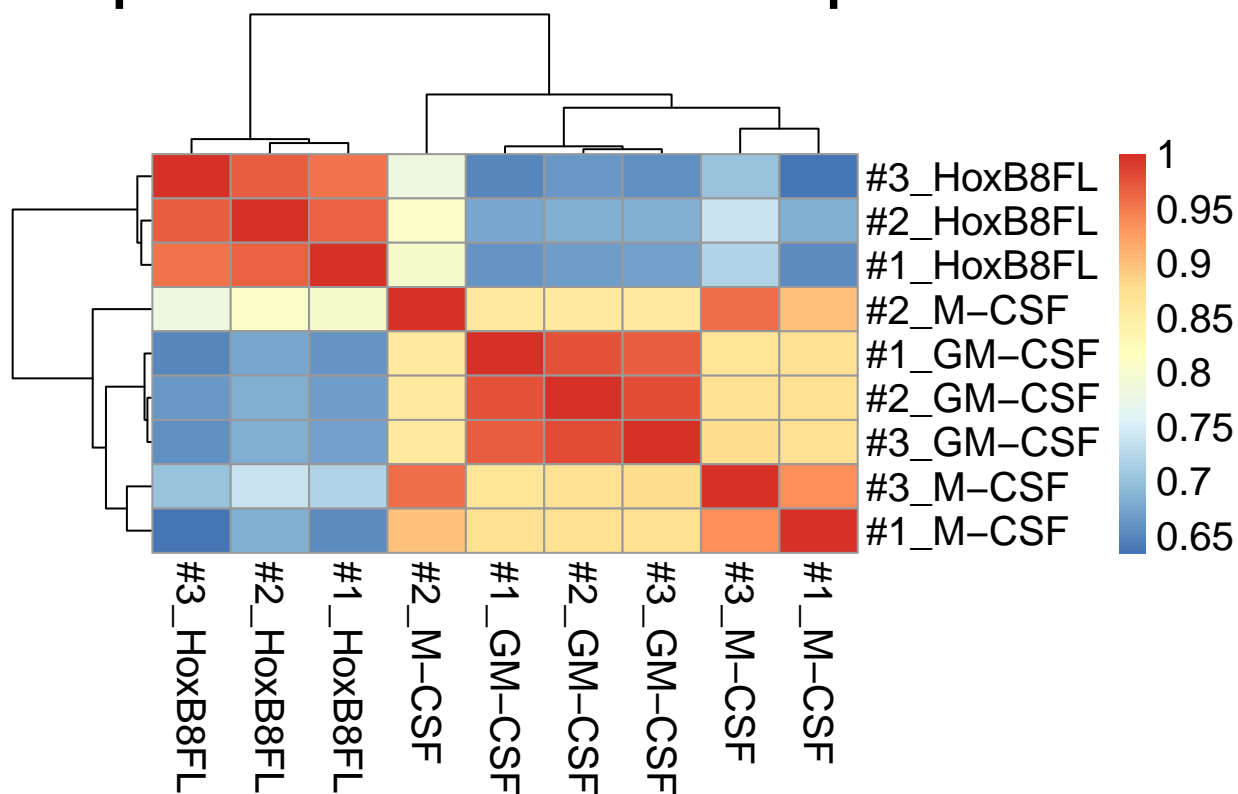
```
## Warning: Using alpha for a discrete variable is not advised.
```

```
plot(pp)
```



```
## Correlation heatmap of samples
pheatmap(mat = cor(x = data, method = "spearman"),
  main = "Spearman Correlation of Samples",
  fontsize_number = 5, fontsize = 15)
```

Spearman Correlation of Samples



Enio Gjerga 2020