### pca-and-heatmap.Rmd

Medhanie Mulaw 5/19/2020

## Principal Component Analysis (PCA) and heatmap based on gene expression data

The first part of the script generates interactive PCA based 3D plot using the function proomp from the stats package (R Core Team, 2019) and rgl (Adler et al., 2020). The second part of the script generates heatmap from gene expression data using the Heatplus package (Ploner, 2020).

In the first step, we load all the required libraries and functions. The corrdist and holust.avl are distance and clustering options as described in the Heatplus package.

#### Load expression Data, grouping factors, and coloring schemes

Normalized expression dataset (preprocessing steps to generate and normalize count data are not described here), the grouping parameters (treatments or experimental arms), and colors of choice are loaded and defined. We then run the pca as described below:

#### Perform PCA and generate the 3D plot

Additionally, one can note above that 3d plot perspective parameters have been predfined (zoom, userMatrix, and windowRect). These can be loaded from text files with predifined values or one can interactively experiment with the 3d plot to obtain the desired viewing angle and save the values. These settings are very helpful in reproducing the 3d plot in a given perspective. Following the pca and parameter settings, one can the generate the plot as follows:

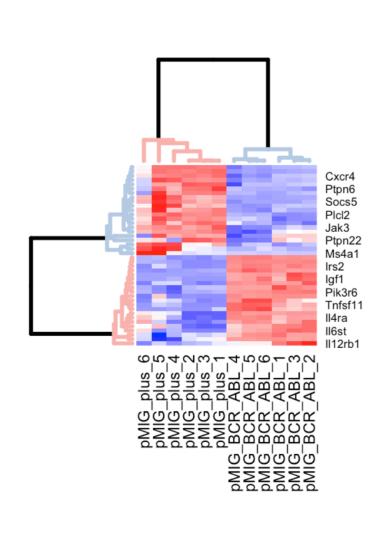
The optional step of saving into a portable document format (pdf) has also been indicated and labels and fonts can further be modified using pdf editors.

# Generating heatmap using a predifined subset of gene expression data

A subet of gene expression dataset (for instance from a differential expression analysis) can be visualized using various tools and packages. We describe here how one can easily generate such plot using the R package Heatplus.

Similar to the steps under PCA, we first load the desired dataset. Optionally, one can also load grouping parameter to later add group annotation

The plot is then generated as follows:



-2 -1 0 1 2

```
sessionInfo()
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
## attached base packages:
## [1] stats graphics grDevices utils
                                               datasets methods base
## other attached packages:
## [1] gplots_3.0.3 rgl_0.100.54 Heatplus_2.32.1
## loaded via a namespace (and not attached):
## [1] Rcpp 1.0.4.6
                                knitr 1.28
                                                         magrittr_1.5
## [4] xtable_1.8-4 R6_2.4.1

## [7] fastmap_1.0.1 stringr_1.4.0

## [10] tools_3.6.1 webshot_0.5.2
                                                         rlang_0.4.5
                                                         caTools_1.18.0
## [10] tools 3.6.1
                                webshot 0.5.2
                                                         xfun_0.13
## [13] KernSmooth 2.23-16
                                miniUI_0.1.1.1
                                                         gtools 3.8.2
## [16] htmltools 0.4.0
                                crosstalk 1.1.0.1
                                                         yaml_2.2.1
## [19] digest_0.6.25
                                shiny_1.4.0.2
                                                         RColorBrewer 1.1-2
                                bitops 1.0-6
## [22] later_1.0.0
                                                         htmlwidgets_1.5.1
## [25] promises_1.1.0
                                manipulateWidget_0.10.1 evaluate_0.14
## [28] mime_0.9
                                rmarkdown 2.1
                                                         gdata_2.18.0
## [31] stringi_1.4.6
                                compiler 3.6.1
                                                         jsonlite 1.6.1
## [34] httpuv_1.5.2
```

References: R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Daniel Adler, Duncan Murdoch and others (2020). rgl: 3D Visualization Using OpenGL. https://CRAN.R-project.org/package=rgl
Alexander Ploner (2020). Heatplus: Heatmaps with row and/or column covariates and colored clusters. https://github.com/alexploner/Heatplus