

3.7 Tutorial-7: Mouse brain dataset (GSE129788)

This tutorial allows users to explore single cell RNAseq data from Mouse brain to show the application of PALMO on tissue samples. Ximerakis et al (2019) study was used out to explore the transcriptomic difference in aging brain. In this case study we used PALMO to identify stable features associated with brain celltypes across aging brain samples. The dataset includes total of 16 mice brains samples (8 young and 8 old) with 37,089 single cells GSE129788. To infer variability (inter- and Intra-) and identify stable genes, please follow following steps.

3.7.1 Load Library

```
#Load Library
library("PALMO")
library("Hmisc")
library("ggpubr")
library("cowplot")
```

3.7.2 Load data and assign paramaters (Time ~ 30sec)

```
#Load scRNA data
mbrain <- readRDS("data/GSE129788_seurat.RDS")
metaDF <- mbrain@meta.data
#check celltypes
sort(unique(mbrain@meta.data$cluster))
#[1] "ABC"      "ARP"       "ASC"       "CPC"       "DC"        "EC"        "EPC"
#[8] "Hb_VC"    "HypEPC"   "ImmN"     "MAC"       "MG"        "MNC"      "mNEUR"
#[15] "NendC"    "NEUT"     "NRP"      "NSC"       "OEG"      "OLG"      "OPC"
#[22] "PC"        "TNC"      "VLMC"     "VSMC"

#Clinical annotations Table S1. Clinical data of the enrolled subjects
metadata <- read.csv("data/GSE129788-annotation.csv", stringsAsFactors = F)
```

3.7.3 Create PALMO object (Time ~ 1min)

```
#Create PALMO object
palmo_obj <- createPALMOobject(anndata=metadata, data=mbrain)

#Assign Sample, PTID and Time parameters
palmo_obj <- annotateMetadata(data_object=palmo_obj,
                                sample_column= "Sample",
                                donor_column= "Age_group",
                                time_column= "Subject_id")

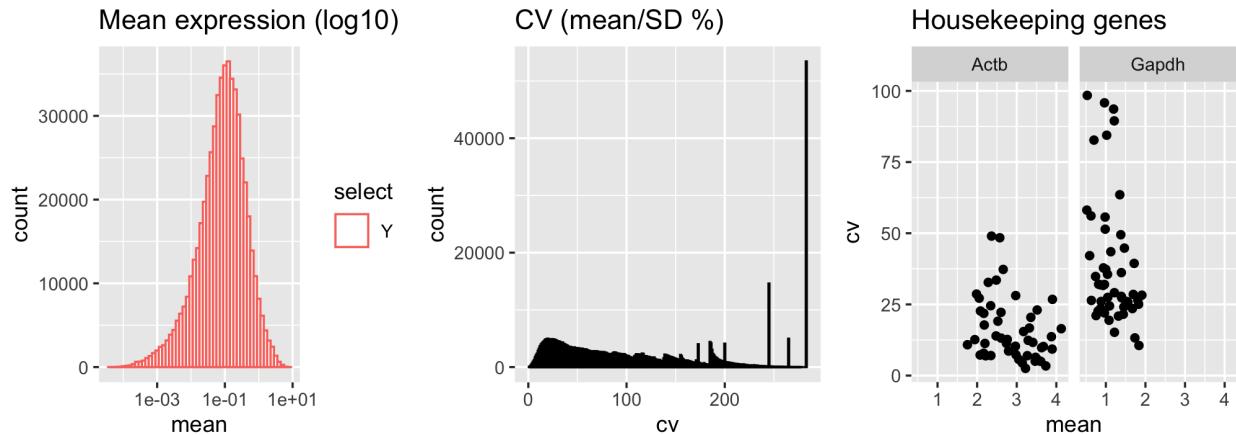
#Sample overlap and final matrix
palmo_obj <- mergePALMOdata(data_object=palmo_obj, datatype="singlecell")

#Aggregate data (Psuedo-bulk)
avgGroup <- "cluster"
palmo_obj <- avgExpCalc(data_object=palmo_obj, assay="RNA",
                         group_column="cluster")
head(palmo_obj@curated[["anndata"]]) #merged annotation data
head(palmo_obj@curated[["data"]]) #scRNA average expression data
```

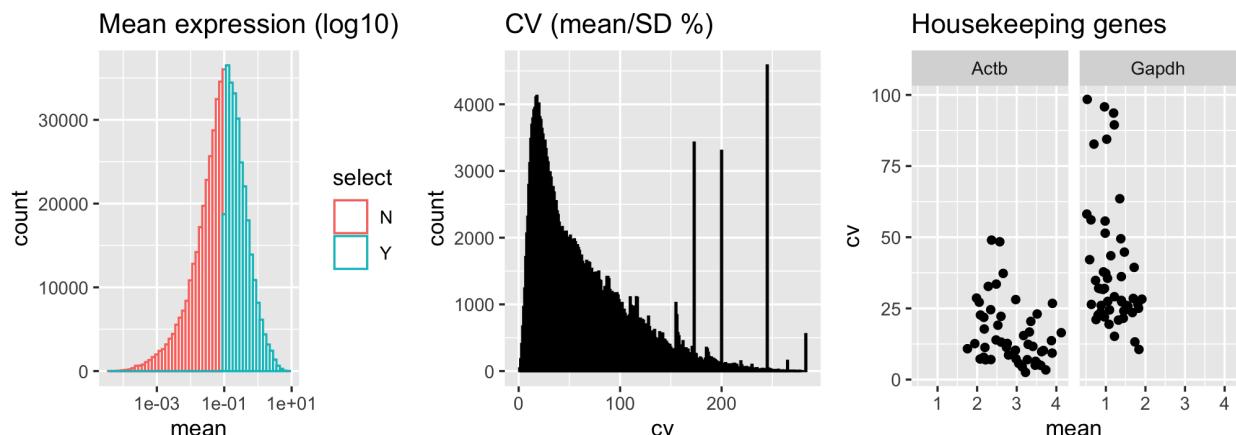
```
#Check for replicates
palmo_obj <- checkReplicates(data_object=palmo_obj, mergeReplicates = T)
```

3.7.4 CV profile (Time ~ 2min)

```
palmo_obj <- cvCalcSCPProfile(data_object=palmo_obj,
                                housekeeping_genes=c("Gapdh", "Actb"),
                                fileName="GSE129788")
```



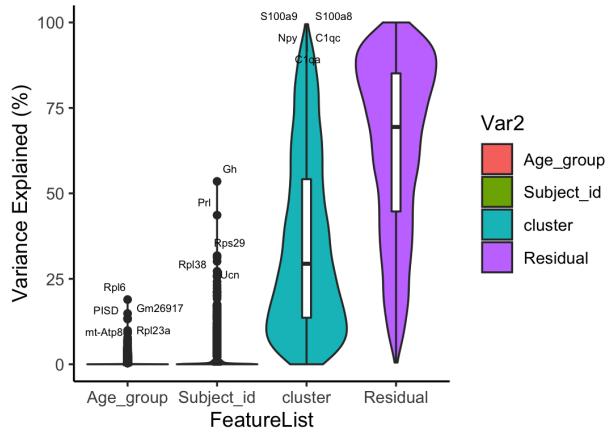
```
#Sample Celltype Mean-CV plot (output directory)
palmo_obj <- cvCalcSCPProfile(data_object=palmo_obj,
                                housekeeping_genes=c("Gapdh", "Actb"),
                                meanThreshold = 0.1,
                                fileName="GSE129788")
```



3.7.5 Features contributing towards donor variations (Time ~ 5min)

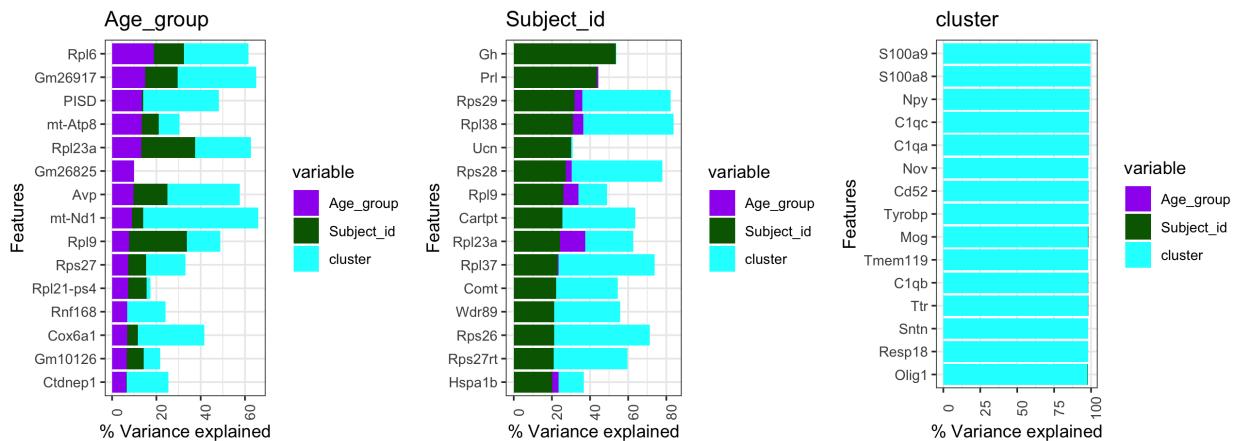
```
#Check the group of interest
head(palmo_obj@curated$anndata)
```

```
#Variance decomposition
featureSet <- c("Age_group", "Subject_id", "cluster")
palmo_obj <- lmeVariance(data_object=palmo_obj,
                           featureSet=featureSet,
                           meanThreshold=0.1, cl=4,
                           fileName="GSE129788")
```



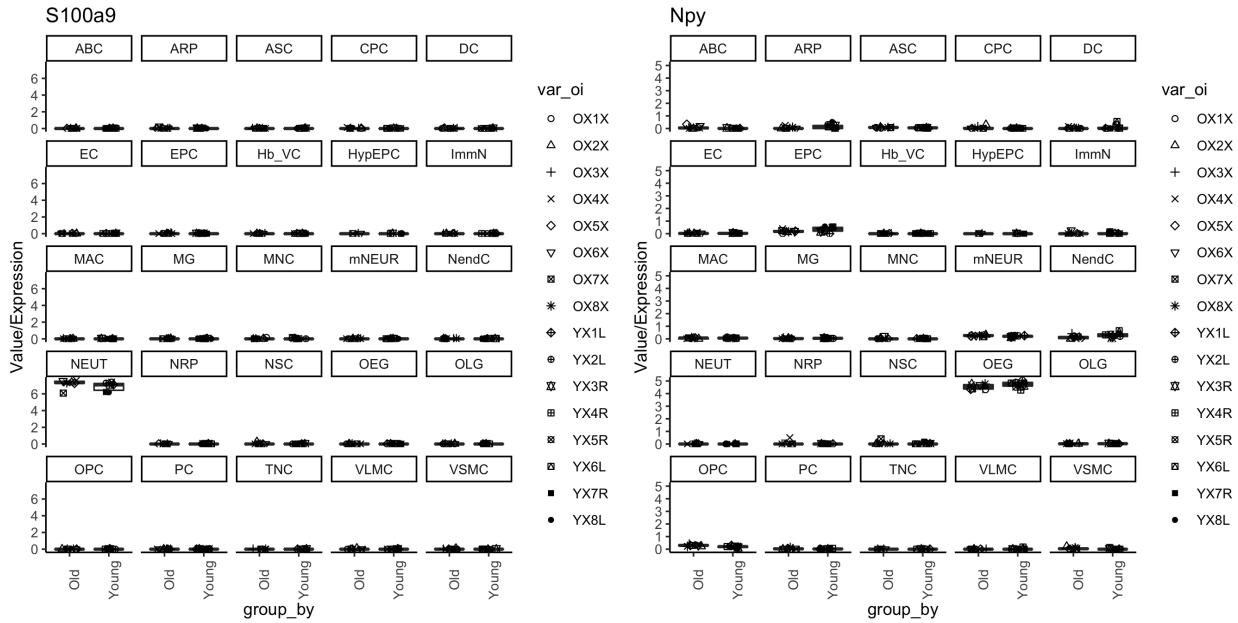
```
var_decomp <- palmo_obj@result$variance_decomposition
head(var_decomp[,featureSet])
#      Age_group Subject_id   cluster
#Rpl6     18.927915 13.6210861 29.057360
#Gm26917 14.853344 14.8230225 35.393884
#PISD    13.413587  0.5785149 34.217585
#mt-Atp8 13.361711  7.6618346 9.437131
#Rpl23a  13.238554 24.1634462 25.245072
#Gm26825 9.885817  0.0000000 0.000000
```

```
#Variance explained (Donor, Time, and celltype)
plots <- variancefeaturePlot(vardata=var_decomp, featureSet=featureSet,
                               cols=c("purple", "darkgreen", "cyan"))
plot_grid(plotlist = plots, ncol=3)
```



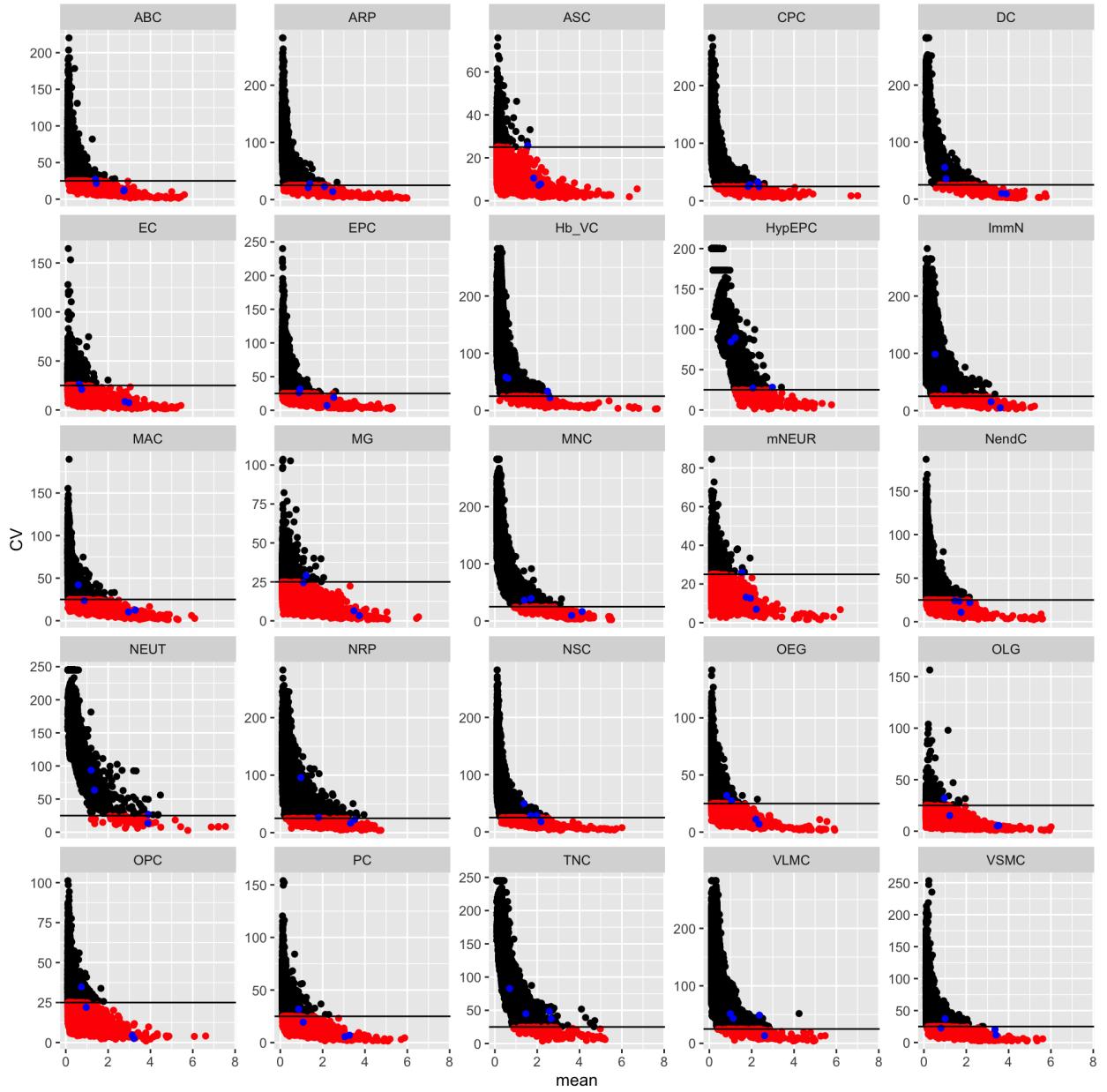
3.7.6 Plot the variables (Time ~ 10sec)

```
plots <- gene_featureplot(data_object=palmo_obj,
                            featureList=c("S100a9", "Npy"),
                            x_group_by="Age_group", var_oi="Subject_id",
                            facet_by = "cluster",
                            x_text_angle=90)
plot_grid(plotlist=plots, ncol= 2, align="hv")
```



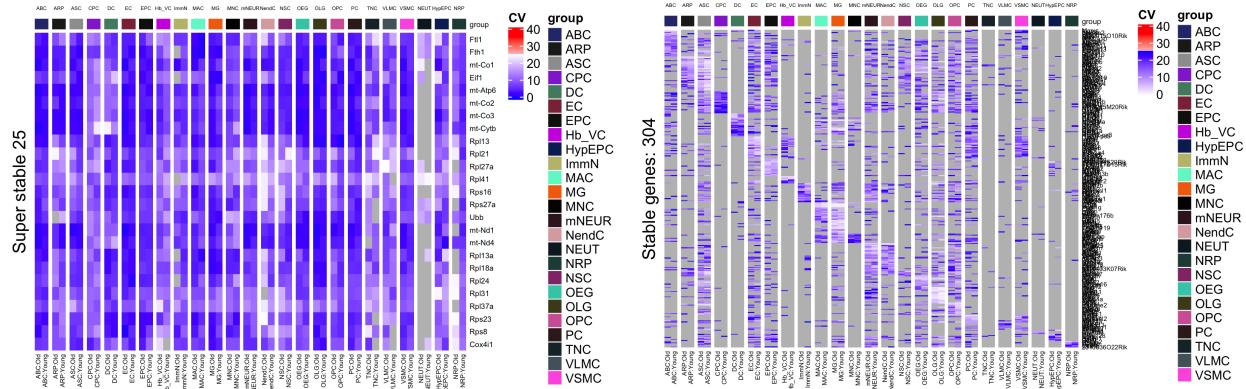
3.7.7 Intra-donor variations over time (Time ~ 4min)

```
#Calculate CV
palmo_obj <- cvCalcSC(data_object=palmo_obj,
                        meanThreshold=0.1, cvThreshold=25,
                        housekeeping_genes=c("Gapdh", "Actb"),
                        fileName="GSE129788")
```



3.7.8 Find stable and variable features in longitudinal data (Time ~ 30sec)

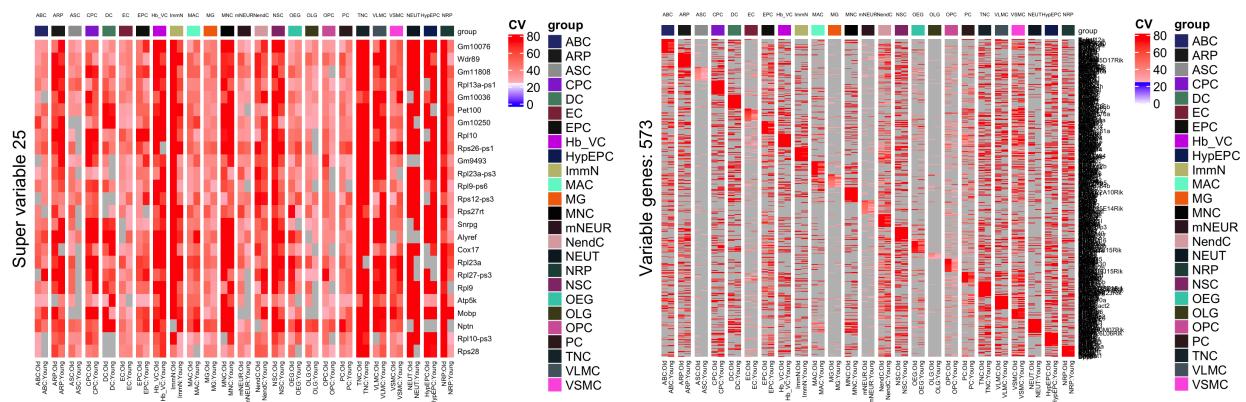
```
palmo_obj <- StableFeatures(data_object=palmo_obj,
                               cvThreshold=25,
                               topFeatures=25,
                               fileName="GSE129788")
stable_genes <- palmo_obj@result$stable_genes
```



```

palmo_obj <- VarFeatures(data_object=palmo_obj,
                           cvThreshold=25,
                           topFeatures=25,
                           fileName="GSE129788")
var_genes <- palmo_obj@result$var_genes

```

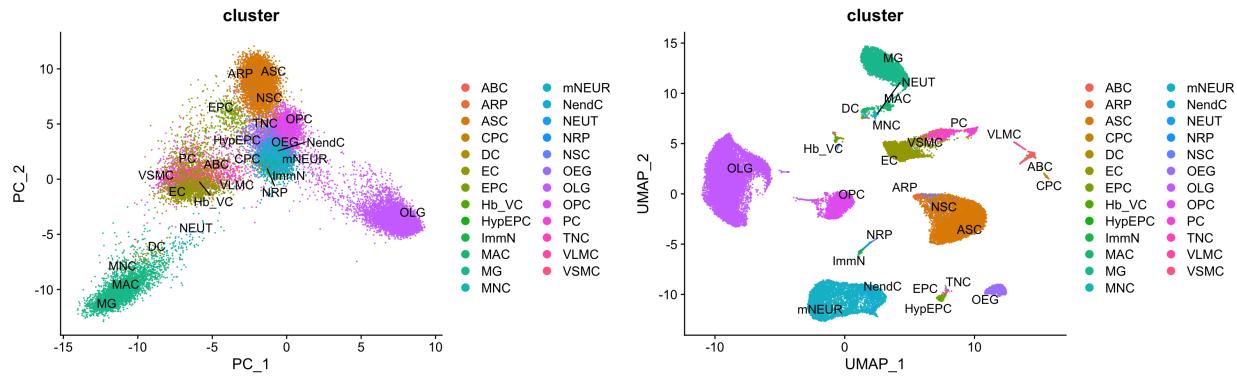


3.7.9 UMAP Plot (Time ~ 2min)

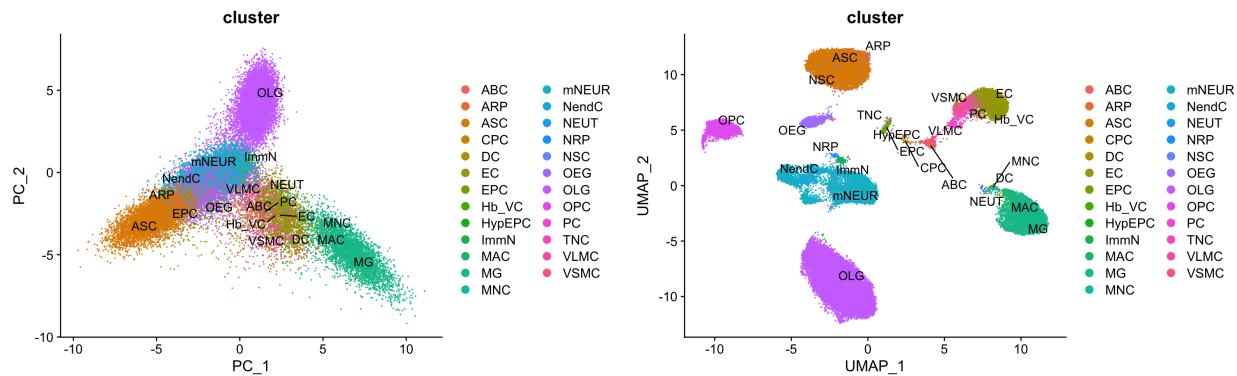
```

#Stable genes UMAP
dimUMAPPlot(data_object=palmo_obj, nPC=15,
             gene_oi=unique(stable_genes$gene),
             group_column="cluster", plotname="stable",
             fileName="GSE129788")

```



```
#Variable genes UMAP
dimUMAPPlot(data_object=palmo_obj, nPC=15,
             gene_oi=unique(var_genes$gene),
             group_column="cluster", plotname="variable",
             fileName="GSE129788")
```

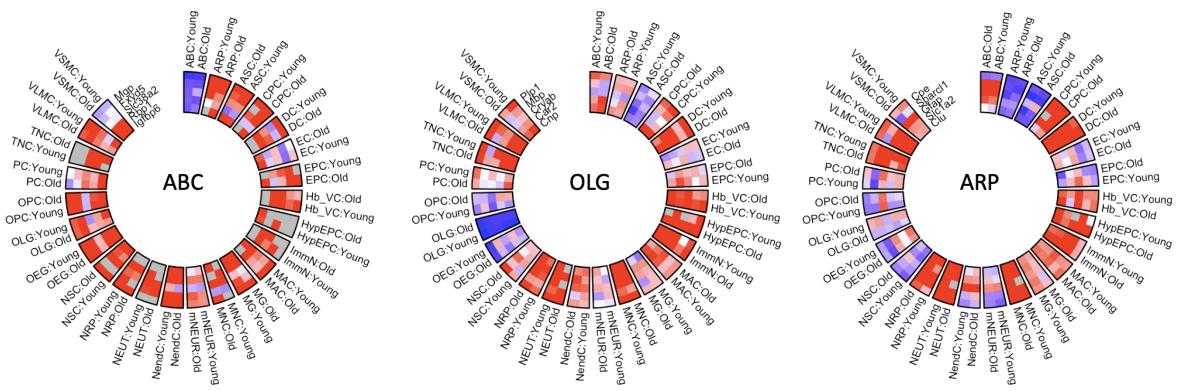


3.7.10 Celltype-specific Circos CV Plot (Time ~ 30sec)

```
#ABC: Arachnoid barrier cells
geneList <- c("Mgp", "Fxyd5", "Slc38a2", "Rbp1", "Igfbp6")
plotres <- genecircosPlot(data_object=palmo_obj,
                           geneList=geneList, colorThreshold=25)

#Olg: Oligodendrocytes
geneList <- c("Plp1", "Mbp", "Cryab", "Car2", "Cnp")
plotres <- genecircosPlot(data_object=palmo_obj,
                           geneList=geneList, colorThreshold=25)

#ARP: Astrocyte-restricted precursors
geneList <- c("Cpe", "Sparcl1", "Gfap", "Slc1a2", "Clu")
plotres <- genecircosPlot(data_object=palmo_obj,
                           geneList=geneList, colorThreshold=25)
```



6 Session info

```
sessionInfo()
#> R version 4.0.3 (2020-10-10)
#> Platform: x86_64-apple-darwin17.0 (64-bit)
#> Running under: macOS Catalina 10.15.7
#>
#> Matrix products: default
#> BLAS:    /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
#> LAPACK:  /Library/Frameworks/R.framework/Versions/4.0/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] grid      stats     graphics  grDevices  utils     datasets  methods
#> [8] base
#>
#> other attached packages:
#> [1] PALMO_0.1.0
#>
#> loaded via a namespace (and not attached):
#> [1] readxl_1.3.1           backports_1.2.0
#> [3] circlize_0.4.11        plyr_1.8.6
#> [5] igraph_1.2.8           lazyeval_0.2.2
#> [7] splines_4.0.3          listenr_0.8.0
#> [9] scattermore_0.7         GenomeInfoDb_1.24.2
#> [11] ggplot2_3.3.5          digest_0.6.28
#> [13] htmltools_0.5.2        fansi_0.5.0
#> [15] magrittr_2.0.1          tensor_1.5
#> [17] cluster_2.1.0          ROCR_1.0-11
#> [19] ComplexHeatmap_2.4.3   globals_0.14.0
#> [21] readr_1.4.0             modelr_0.1.8
#> [23] matrixStats_0.61.0     colorspace_2.0-2
#> [25] rvest_0.3.6            blob_1.2.1
#> [27] ggrepel_0.9.1          haven_2.3.1
#> [29] xfun_0.25              dplyr_1.0.7
#> [31] RCurl_1.98-1.2          crayon_1.4.2
#> [33] jsonlite_1.7.2          lme4_1.1-25
#> [35] spatstat_1.64-1         spatstat.data_2.1-0
#> [37] survival_3.2-7          zoo_1.8-9
#> [39] glue_1.5.0              polyclip_1.10-0
#> [41] gtable_0.3.0             zlibbioc_1.34.0
#> [43] XVector_0.28.0          leiden_0.3.9
#> [45] DelayedArray_0.14.1     GetoptLong_1.0.4
#> [47] SingleCellExperiment_1.10.1 future.apply_1.8.1
#> [49] shape_1.4.5              BiocGenerics_0.34.0
#> [51] abind_1.4-5              scales_1.1.1
#> [53] pheatmap_1.0.12          DBI_1.1.0
#> [55] miniUI_0.1.1.1          Rcpp_1.0.7
#> [57] viridisLite_0.4.0         xtable_1.8-4
#> [59] clue_0.3-57              reticulate_1.22
#> [61] stats4_4.0.3              htmlwidgets_1.5.4
```

```

#> [63] httr_1.4.2
#> [65] ellipsis_0.3.2
#> [67] factoextra_1.0.7.999
#> [69] farver_2.1.0
#> [71] uwot_0.1.10
#> [73] deldir_1.0-6
#> [75] tidyselect_1.1.1
#> [77] reshape2_1.4.4
#> [79] munsell_0.5.0
#> [81] tools_4.0.3
#> [83] broom_0.7.2
#> [85] evaluate_0.14
#> [87] fastmap_1.1.0
#> [89] goftest_1.2-3
#> [91] fs_1.5.0
#> [93] purrrr_0.3.4
#> [95] pbapply_1.5-0
#> [97] nlme_3.1-149
#> [99] xml2_1.3.2
#> [101] plotly_4.10.0
#> [103] spatstat.utils_2.2-0
#> [105] tweenr_1.0.1
#> [107] statmod_1.4.35
#> [109]forcats_0.5.0
#> [111] Matrix_1.3-4
#> [113] vctrs_0.3.8
#> [115] lifecycle_1.0.1
#> [117] GlobalOptions_0.1.2
#> [119] bitops_1.0-7
#> [121] cowplot_1.1.1
#> [123] GenomicRanges_1.40.0
#> [125] patchwork_1.1.1
#> [127] promises_1.2.0.1
#> [129] gridExtra_2.3
#> [131] parallelly_1.28.1
#> [133] boot_1.3-25
#> [135] assertthat_0.2.1
#> [137] MAST_1.14.0
#> [139] SeuratObject_4.0.2
#> [141] GenomeInfoDbData_1.2.3
#> [143] mgcv_1.8-33
#> [145] hms_0.5.3
#> [147] tidyverse_1.3.0
#> [149] minqa_1.2.4
#> [151] Rtsne_0.15
#> [153] Biobase_2.48.0
#> [155] lubridate_1.7.9

RColorBrewer_1.1-2
Seurat_4.0.0
ica_1.0-2
pkgconfig_2.0.3
dbplyr_1.4.4
utf8_1.2.2
rlang_0.4.12
later_1.3.0
cellranger_1.1.0
generics_0.1.1
ggridges_0.5.3
stringr_1.4.0
yaml_2.2.1
knitr_1.30
fitdistrplus_1.1-6
RANN_2.6.1
future_1.23.0
mime_0.12
compiler_4.0.3
png_0.1-7
reprex_0.3.0
tibble_3.1.6
stringi_1.7.5
lattice_0.20-41
nloptr_1.2.2.2
pillar_1.6.4
lmtest_0.9-39
RcppAnnoy_0.0.19
data.table_1.14.2
irlba_2.3.3
httpuv_1.6.3
R6_2.5.1
KernSmooth_2.23-17
IRanges_2.22.2
codetools_0.2-16
MASS_7.3-53
SummarizedExperiment_1.18.2
rjson_0.2.20
sctransform_0.3.2
S4Vectors_0.26.1
parallel_4.0.3
rpart_4.1-15
tidyR_1.1.4
rmarkdown_2.5
ggforce_0.3.2
shiny_1.7.1

```