

### 3.5 Tutorial-5: COVID19 longitudinal dataset (CNP0001102)

This tutorial allows users to explore single cell RNAseq data variability across COVID and FLU donors. PBMC from the patients were collected longitudinally. Single cell data from Zhu et al. 2020 downloaded from here. Metadata is downloaded from table and can be found in the data. To infer variability (inter- and Intra-) and identify stable genes, please follow following steps.

#### 3.5.1 Load Library

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

#### 3.5.2 Load data and assign paramaters (Time < 1min)

```
#Load scRNA data
pbmc <- readRDS("data/CNP0001102_Final_nCoV_0716_upload.RDS")
#Add column Sample
pbmc@meta.data$Sample <- pbmc@meta.data$batch
#check celltypes
sort(unique(pbmc@meta.data$cell_type))
#[1] Cytotoxic CD8 T cells Naive T cells           NKs
#[4] MAIT          Activated CD4 T cells Naive B cells
#[7] Plasma         Memory B cells               XCL+ NKs
#[10] Cycling T cells   Monocytes            DCs
#[13] Cycling Plasma    Stem cells           Megakaryocytes

#Clinical annotations Table S1. Clinical data of the enrolled subjects
metadata <- read.csv("data/CNP0001102-annotation.csv", stringsAsFactors = F)

#Exploring only COVID samples
metadata <- metadata[metadata$Participant %in% c("COV-1", "COV-2", "COV-3", "COV-4", "COV-5"),]
#Exploring only FLU samples
#metadata <- metadata[metadata$Participant %in% c("IAV-1", "IAV-2"),]
```

#### 3.5.3 Create PALMO object (Time < 1min)

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

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

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

#Aggregate data (Psuedo-bulk)
```

```

palmo_obj <- avgExpCalc(data_object=palmo_obj, assay="RNA",
                           group_column="cell_type")
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)

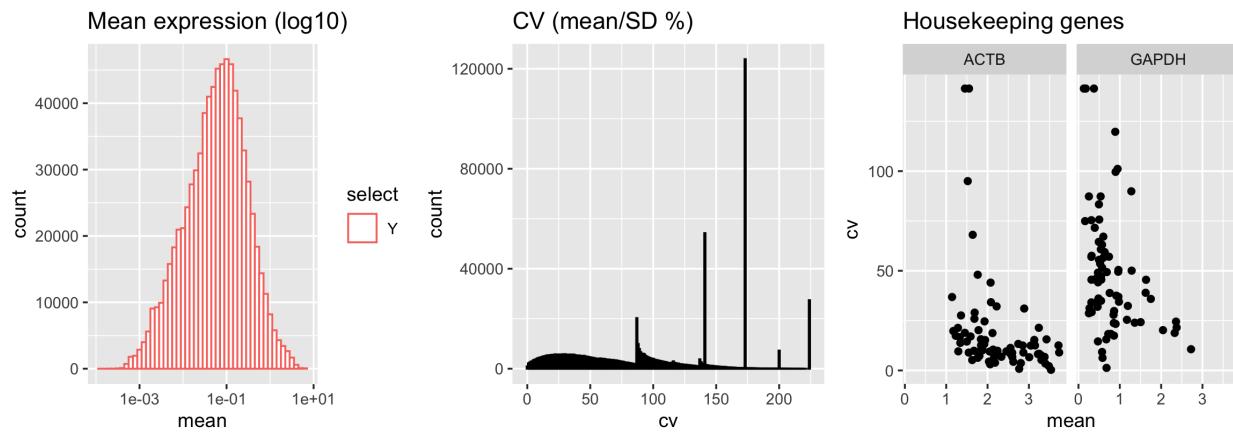
```

### 3.5.4 CV profile (Time ~ 1min)

```

palmo_obj <- cvCalcSCPProfile(data_object=palmo_obj,
                                housekeeping_genes=c("GAPDH", "ACTB"),
                                fileName="CNP0001102")

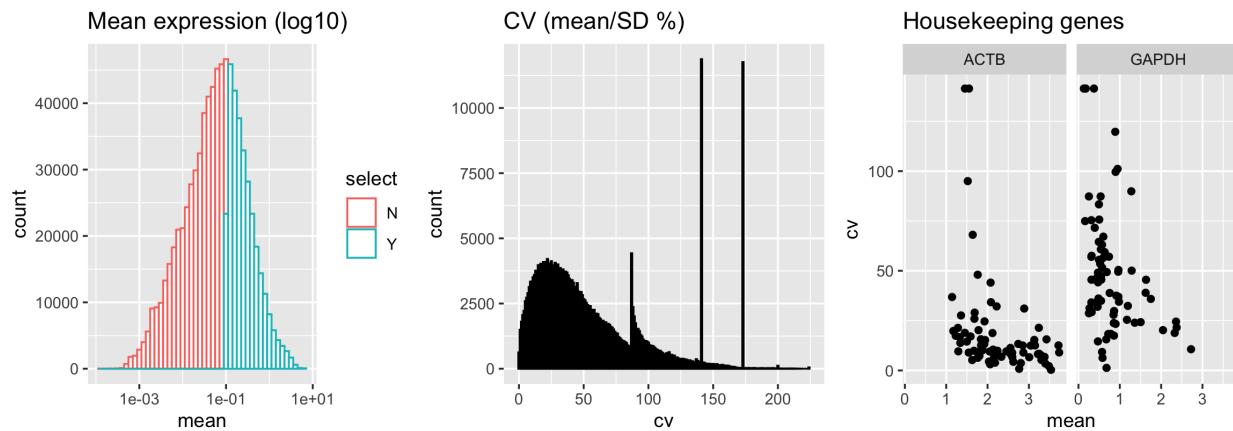
```



```

#Sample Celltype Mean-CV plot (output directory)
palmo_obj <- cvCalcSCPProfile(data_object=palmo_obj,
                                housekeeping_genes=c("GAPDH", "ACTB"),
                                meanThreshold = 0.1,
                                fileName="CNP0001102")

```



```

cvSCsampleprofile(data_object=palmo_obj,
                   meanThreshold = 0.1, plot_log10=T,
                   cvThreshold = 25)

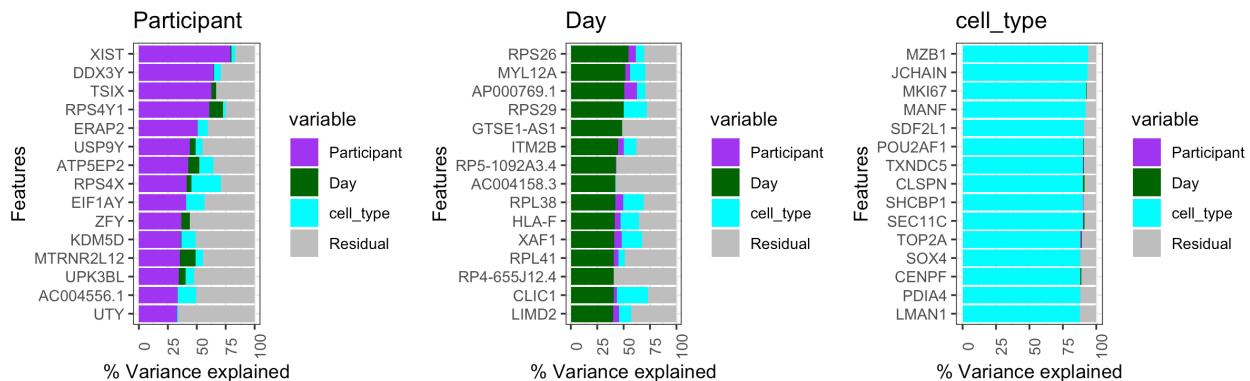
```

### 3.5.5 Features contributing towards donor variations (Time ~ 5min)

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

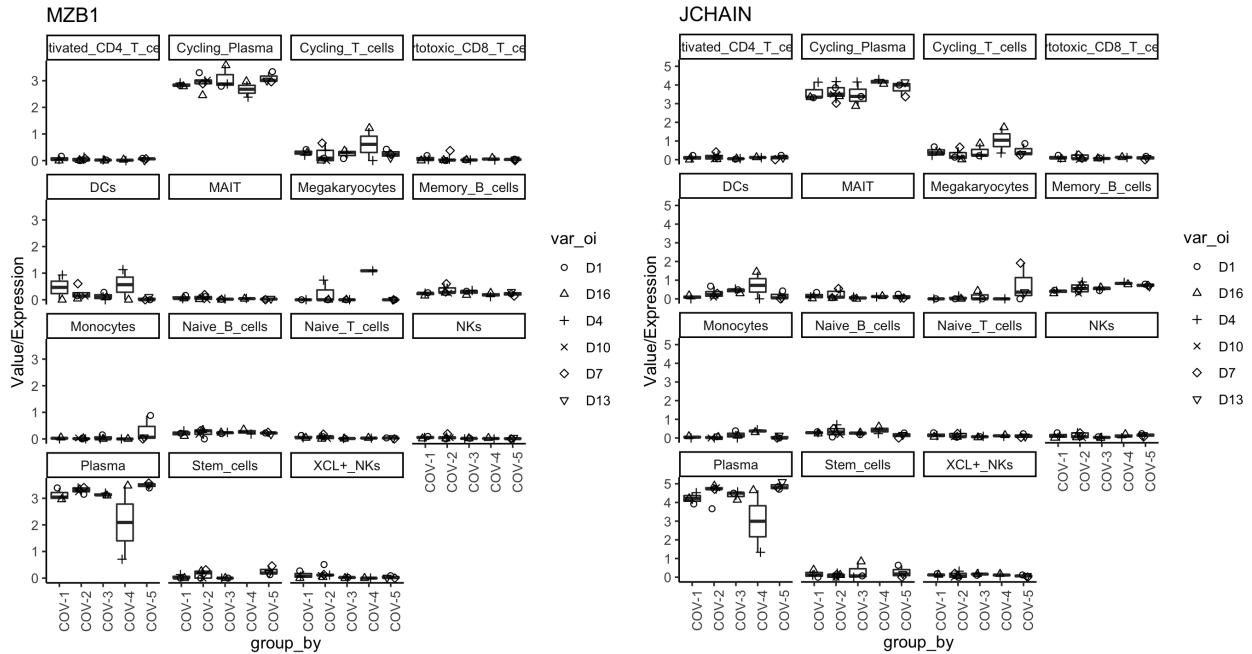
#Variance decomposition
featureSet <- c("Participant", "Day", "cell_type")
palmo_obj <- lmeVariance(data_object=palmo_obj,
                           featureSet=featureSet,
                           meanThreshold=0.1, cl=4,
                           fileName="CNP0001102")
var_decomp <- palmo_obj@result$variance_decomposition
head(var_decomp[,featureSet])
  Participant      Day cell_type
#XIST      78.82243  0.936695  3.510383
#DDX3Y     64.44897  0.637655  5.824030
#TSIX      62.57711  3.997646  0.339527
#RPS4Y1    60.73146  11.768212 2.346928
#ERAP2      50.92503  0.000000  8.745452
#USP9Y     44.21867  4.657177  5.997529

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



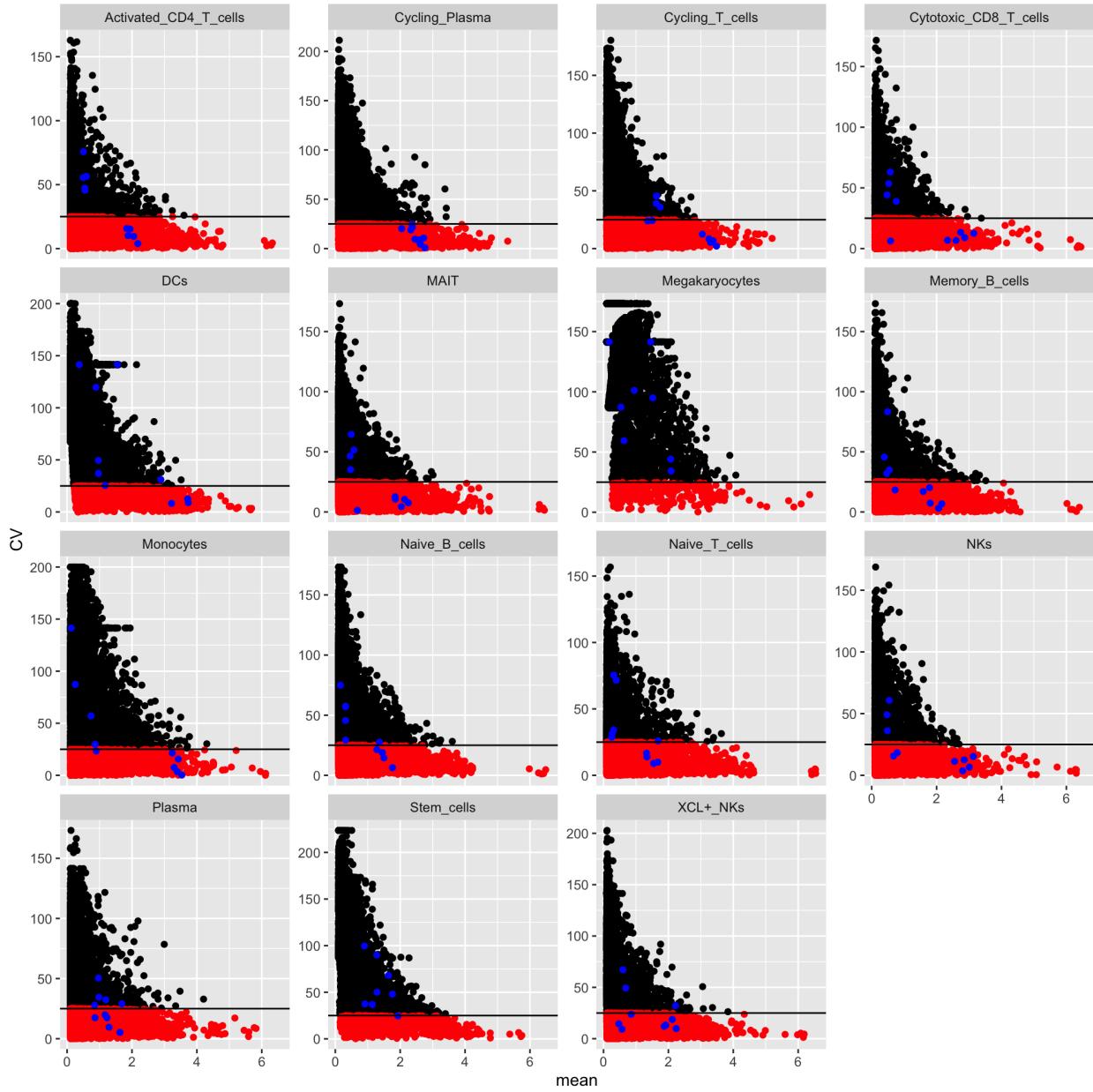
### 3.5.6 Plot the variables (Time ~ 10sec)

```
plots <- gene_featureplot(data_object=palmo_obj,
                           featureList=c("MZB1", "JCHAIN"),
                           facet_by="cell_type", x_text_angle=90)
```



### 3.5.7 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="CNP0001102")
```

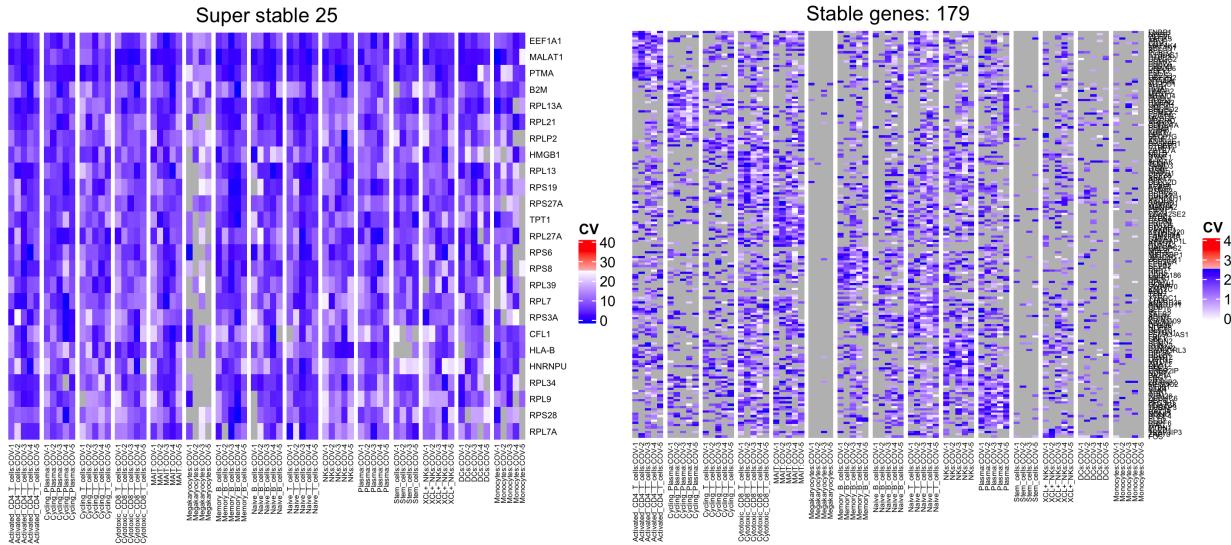


### 3.5.8 Find stable and variable features in longitudinal data (Time ~ 10sec)

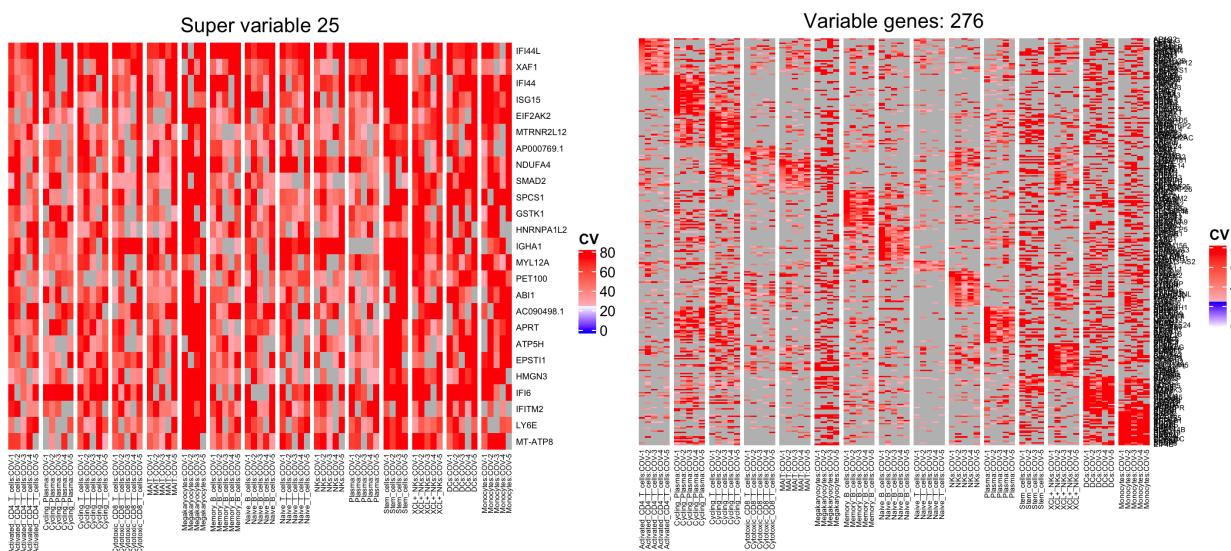
```

donorThreshold <- 5 #number of donors
groupThreshold <- 38 #number of donors * number of celltypes/2 (5x15/2)
topFeatures <- 25
palmo_obj <- StableFeatures(data_object=palmo_obj,
                               cvThreshold=25,
                               donorThreshold=5, groupThreshold=38,
                               topFeatures=25,
                               fileName="CNP0001102")
stable_genes <- palmo_obj@result$stable_genes

```

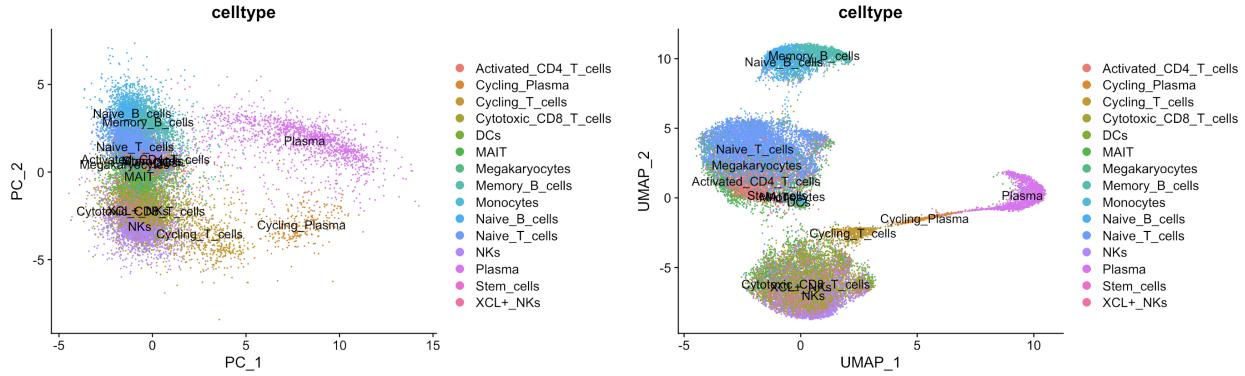


```
palmo_obj <- VarFeatures(data_object=palmo_obj,
                           cvThreshold=25,
                           donorThreshold=5, groupThreshold=38,
                           topFeatures=25,
                           fileName="CNP0001102")
var_genes <- palmo_obj@result$var_genes
```

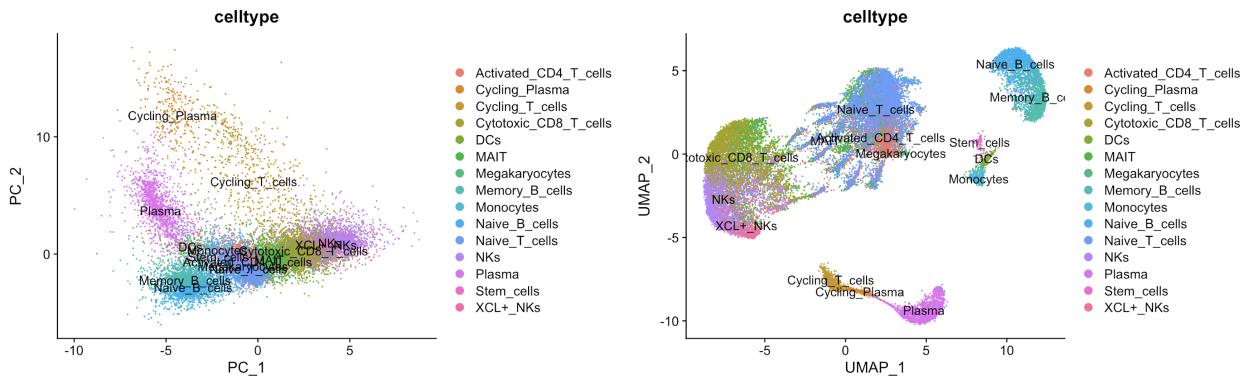


### 3.5.9 UMAP Plot (Time ~ 2min)

```
#Stable genes UMAP
dimUMAPPlot(data_object=palmo_obj, nPC=15,
             gene_oi=unique(stable_genes$gene),
             group_column="cell_type", plotname="stable",
             fileName="CNP0001102")
```



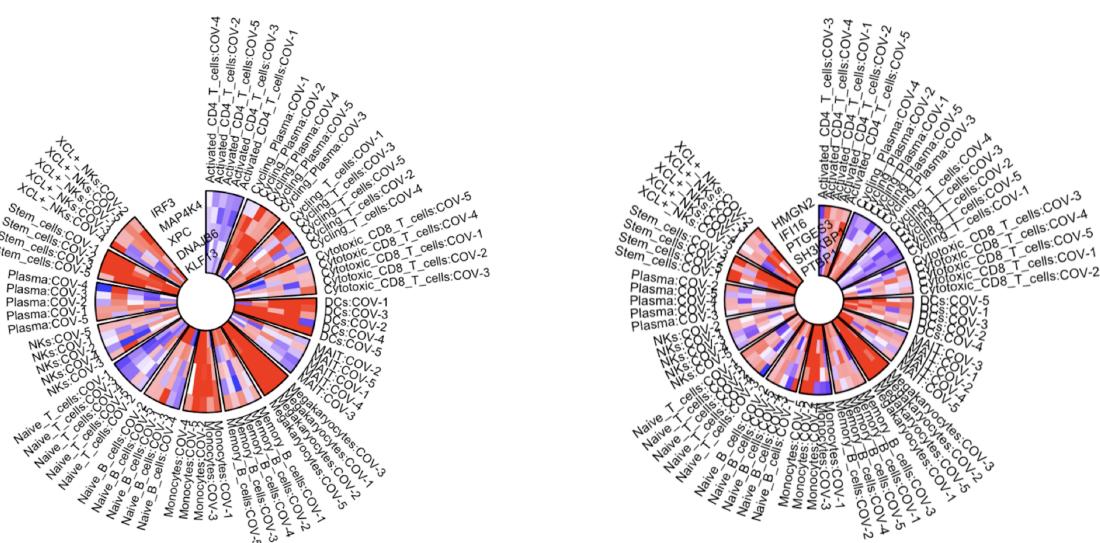
```
#Variable genes UMAP
dimUMAPPPlot(data_object=palmo_obj, nPC=15,
              gene_oi=unique(var_genes$gene),
              group_column="cell_type", plotname="variable",
              fileName="CNP0001102")
```



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

```
#Activated CD4 T-cells
geneList <- c("IRF3", "MAP4K4", "XPC", "DNAJB6", "KLF13")
plotres <- genecircosPlot(data_object=palmo_obj,
                           geneList=toList(geneList), colorThreshold=25)

#Cycling T-cells
geneList <- c("HMGN2", "IFI16", "PTGES3", "SH3KBP1", "PTBP1")
plotres <- genecircosPlot(data_object=palmo_obj,
                           geneList=toList(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

```