3.3 Tutorial-3: scATAC Longitudinal data (n=4 and 6 weeks follow-up)

This tutorial allows users to explore single cell ATACseq genscore data measured from 4 healthy donors over 6 timepoints (week 2-7). Single cell ATAC data available at GSE190992. (1) pbmc_scatac_archr_genescore_longitudinal_data (2) AIFI-Metadata.Rda (clinical metadata). Longitudinal dataset have 4 donors and 18 samples. To infer the variations at single cell ATAC please follow following steps.

3.3.1 Load Library

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

3.3.2 Load data and assign paramaters (Time <30sec)

```
#scATAC object
```

Load genescorematrix from archR or relevant tools (Aggregate data at celltypes (pseudo-bulk))

```
load("data/AIFI-scATAC-PBMC-FinalData.Rda")
datamatrix <- log2(scatac_gm+1)

#Load annotation data
load("data/AIFI-Metadata.Rda")
```

3.3.3 Create PALMO object (Time <30sec)

```
#Create PALMO object
palmo_obj <- createPALMOobject(anndata=ann, data=datamatrix)

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

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

#Check for replicates
palmo_obj <- checkReplicates(data_object=palmo_obj, mergeReplicates = T)
```
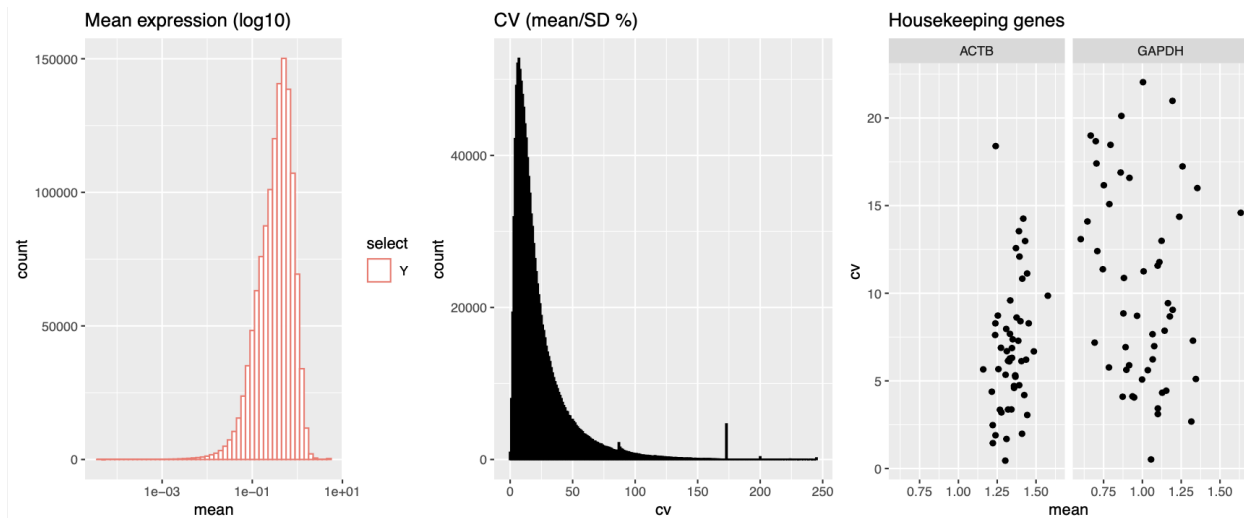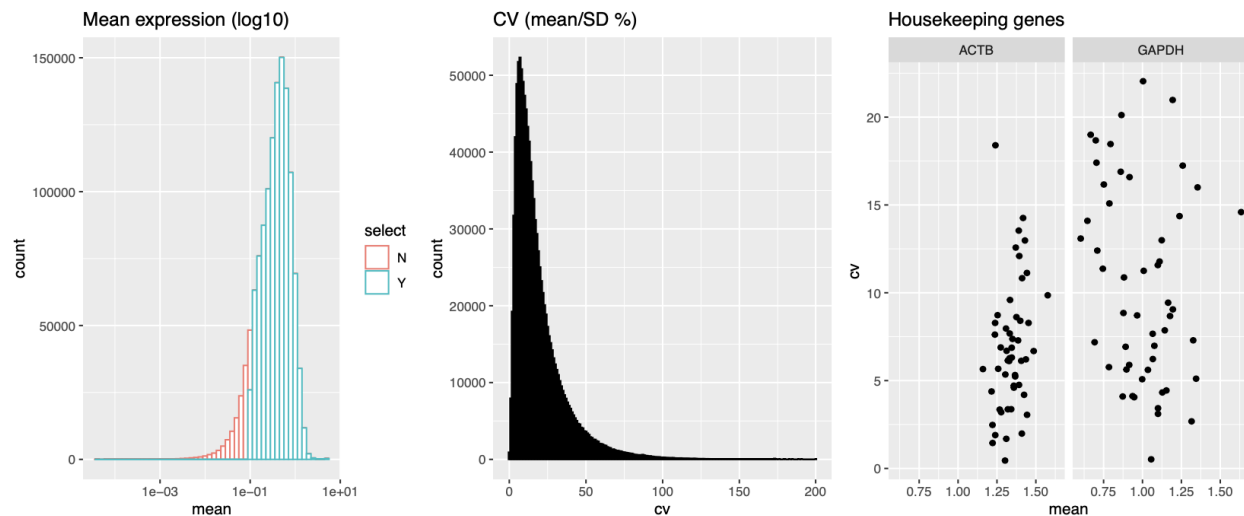
3.3.4 CV profile (Time ~2min)

```
palmo_obj <- cvCalcSCProfile(data_object=palmo_obj,
                            housekeeping_genes=c("GAPDH", "ACTB"),
                            fileName="scatac")
```

```
palmo_obj <- cvCalcSCProfile(data_object=palmo_obj,
                             housekeeping_genes=c("GAPDH", "ACTB"),
                             meanThreshold = 0.1,
                             fileName="scatac")
```



```
#Sample Celltype Mean-CV plot
cvSCsampleprofile(data_object=palmo_obj, meanThreshold = 0.1,
                  cvThreshold = 10, fileName="scatac")
#plots saved in output directory
```

3.3.5   Features contributing towards donor variations (Time ~ 8min)

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

#Variance decomposition
featureSet <- c("PTID","Time","group")
palmo_obj <- lmeVariance(data_object=palmo_obj,
```

```
                    featureSet=featureSet,
                    meanThreshold=0.1, cl=4,
                    fileName="scatac")
var_decomp <- palmo_obj@result$variance_decomposition
head(var_decomp[,featureSet])
#                     PTID       Time       group
#FIRRE          88.88239 0.2606612 2.8339971
#XIST           87.05064 0.3986387 0.5178851
#RHD            81.77230 0.3641998 1.4574140
#GSTM1          80.29766 0.1786628 6.2277645
#ZNF705D        70.69163 0.4342325 0.4752960
#LOC105376805   66.62977 1.2741294 9.8918653
```
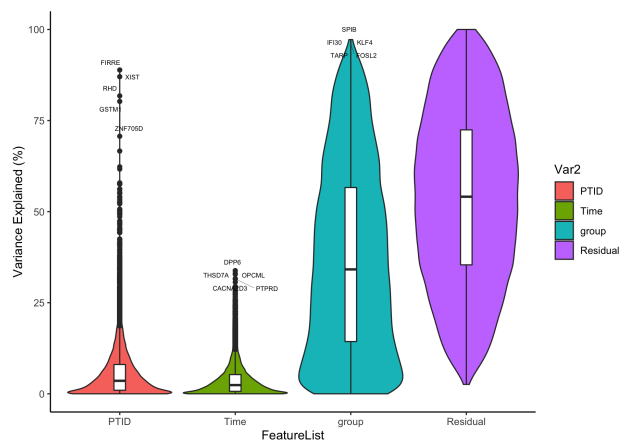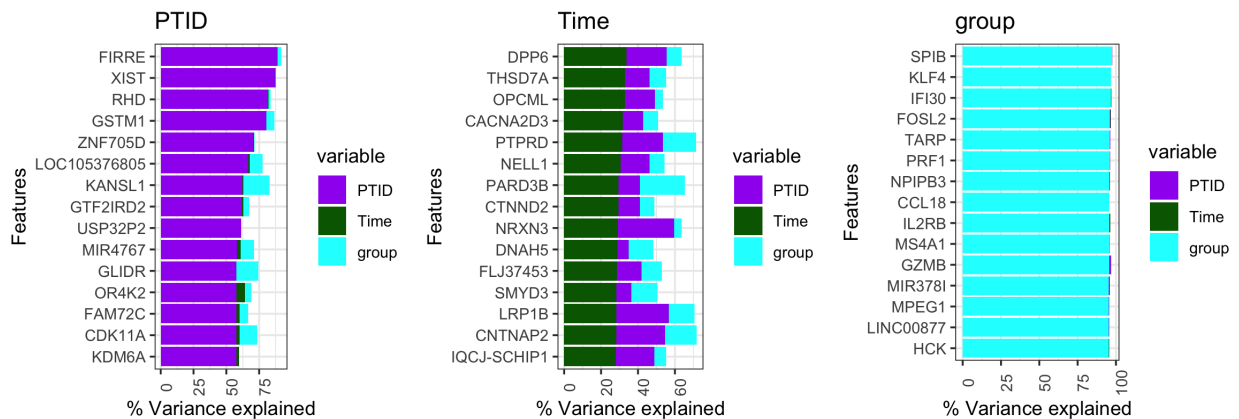


```
#Variance contributing features
plots <- variancefeaturePlot(vardata=var_decomp, featureSet=featureSet,
                             cols=c("purple", "darkgreen", "cyan"))
plot_grid(plotlist = plots, ncol=3)
```



```
#Top genes
plots <- gene_featureplot(data_object=palmo_obj,
                          featureList=c("FIRRE", "XIST",
                                        "DPP6", "THSD7A",
                                        "SPIB", "KLF4"),
                          x_group_by="PTID", var_oi="Time",
```
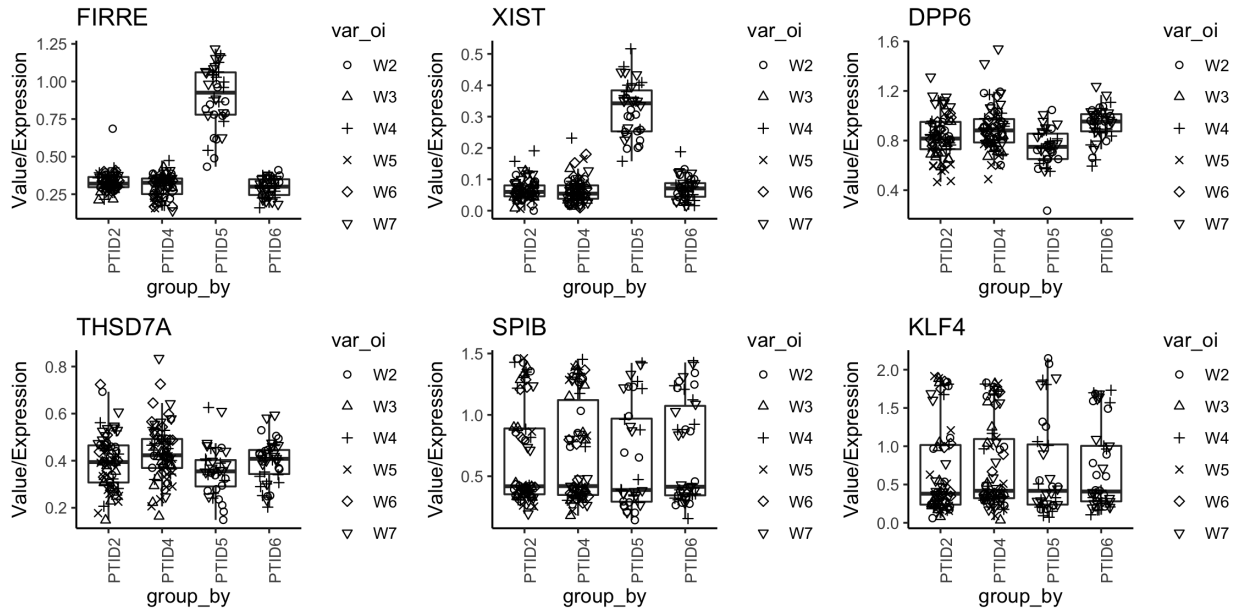
23

```
                        x_text_angle=90)
plot_grid(plotlist=plots, ncol= 3, align="hv")
```
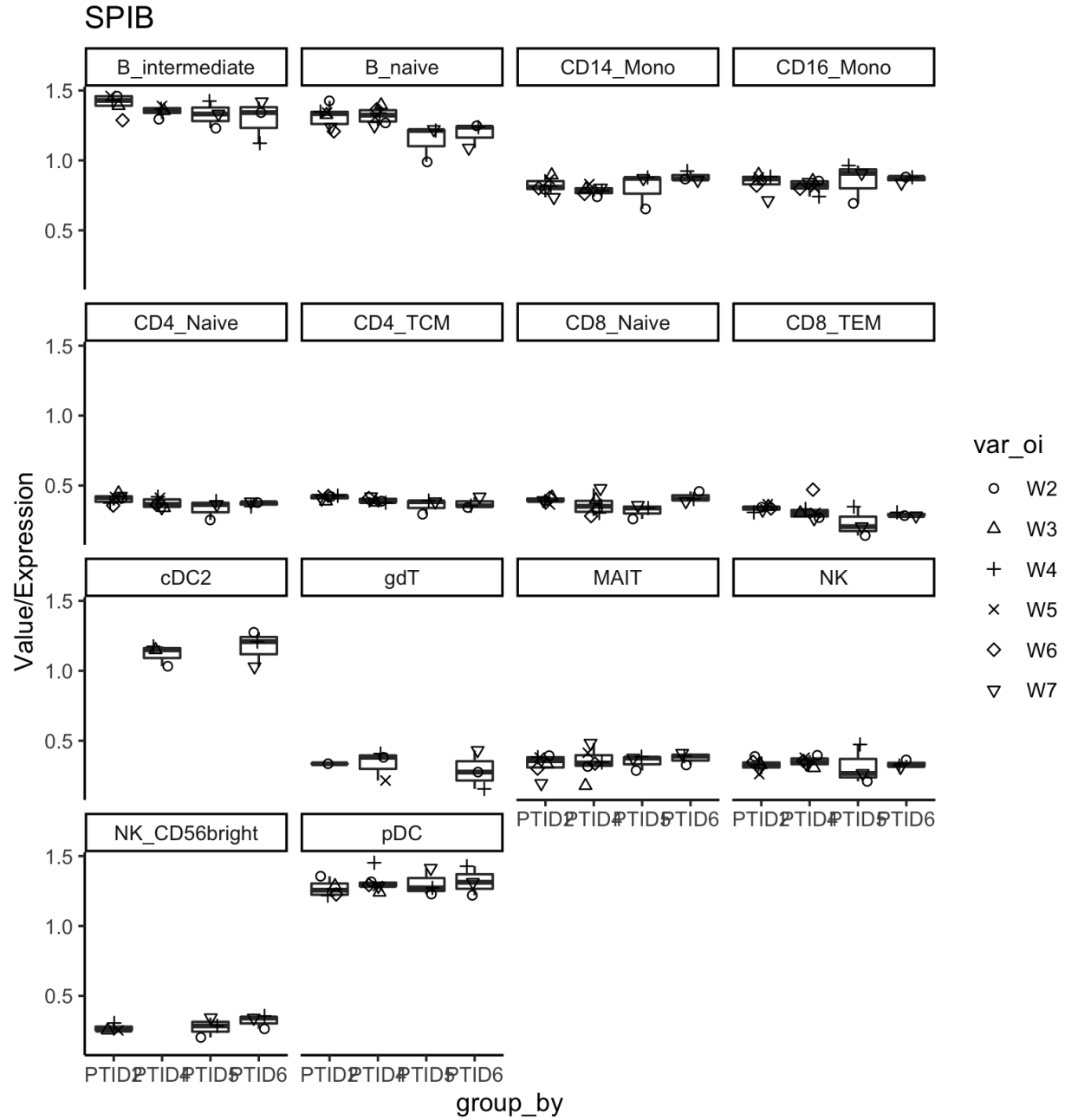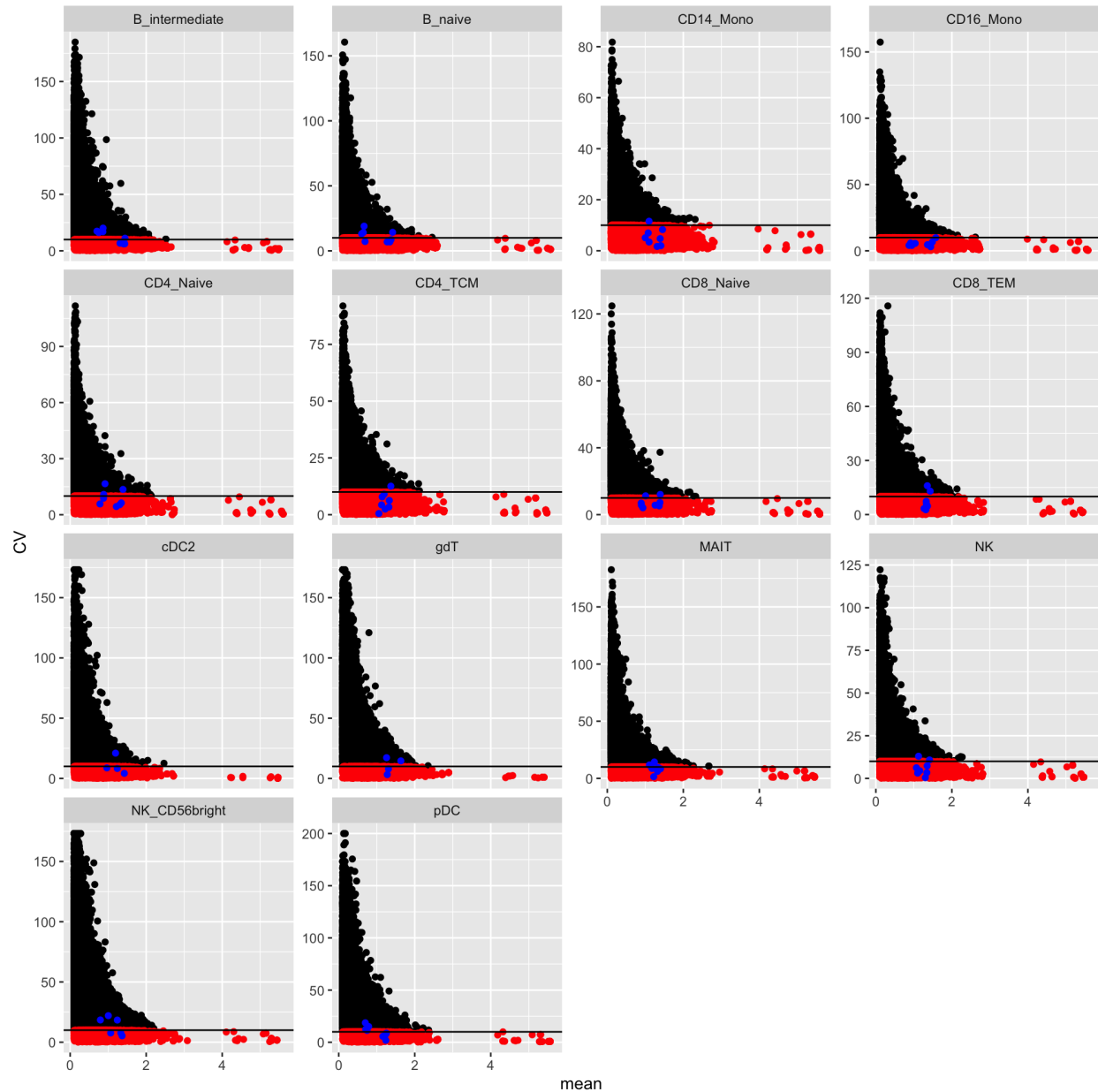


```
gene_featureplot(data_object=palmo_obj, featureList="SPIB", facet_by="group")
```

3.3.6 Intra-donor variations over time (Time ~ 5min)

```
palmo_obj <- cvCalcSC(data_object=palmo_obj,
                  meanThreshold=0.1, cvThreshold=10,
                  housekeeping_genes=c("GAPDH", "ACTB"),
                  fileName="scatac")
```

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

```
#Find stable and variable features in longitudinal data
donorThreshold <- 4
groupThreshold <- 28 #number of donors * number of celltypes/2 (4x14/2)
topFeatures <- 25
palmo_obj <- StableFeatures(data_object=palmo_obj,
                             cvThreshold=10,
                             donorThreshold=4, groupThreshold=28,
                             topFeatures=25,
                             fileName="scatac")
stable_gene <- palmo_obj@result$stable_genes
```

Super stable 25

Stable genes: 228
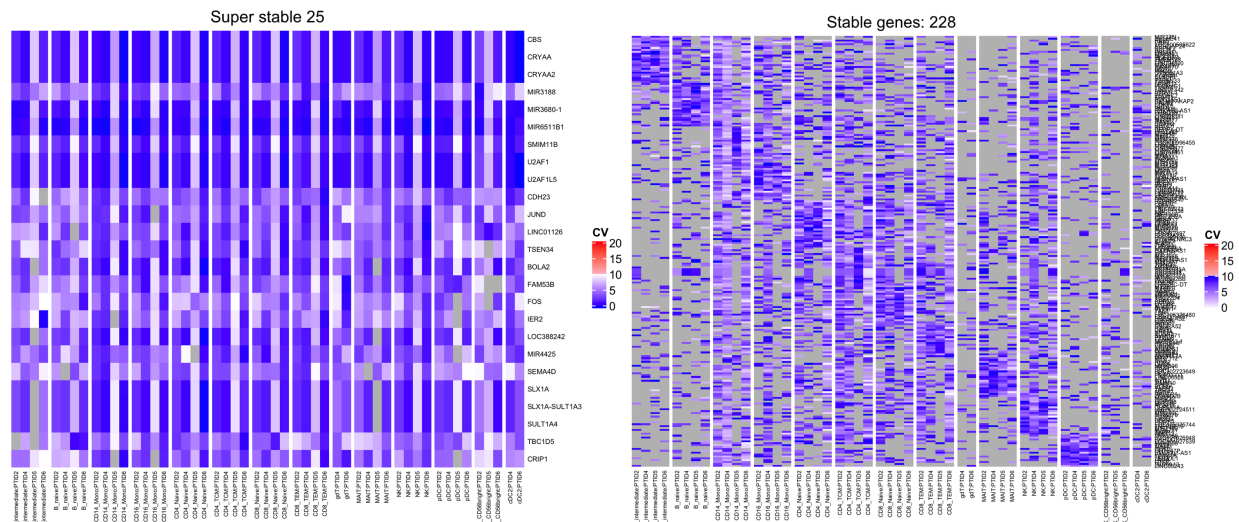

```
palmo_obj <- VarFeatures(data_object=palmo_obj,
                         cvThreshold=10,
                         donorThreshold=4, groupThreshold=28,
                         topFeatures=25,
                         fileName="scatac")
var_gene <- palmo_obj@result$var_genes
```


Super variable 25

Variable genes: 232
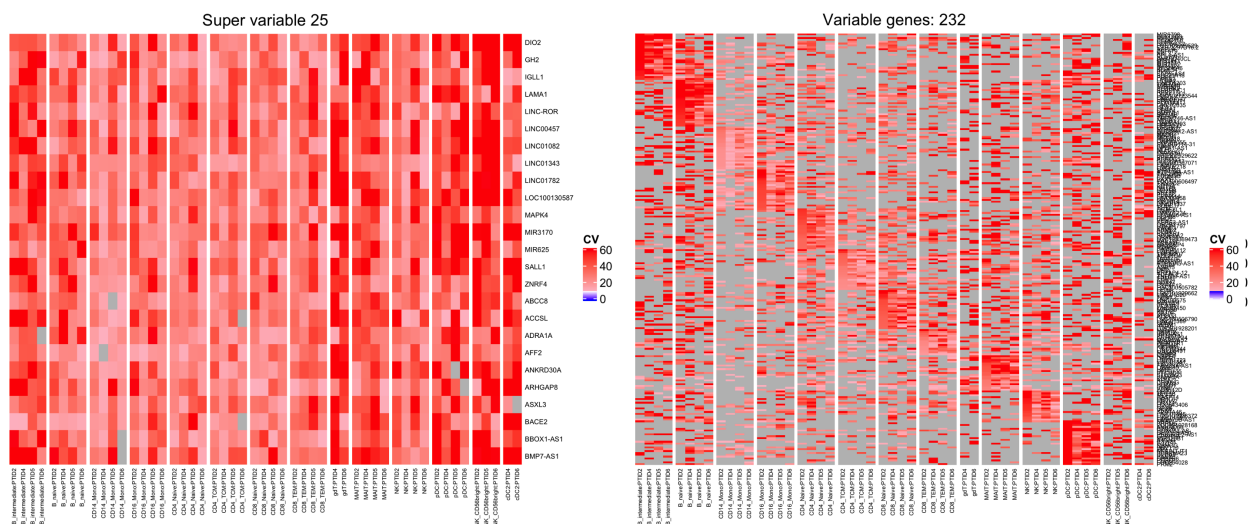
3.3.8   Circos CV plot (Time ~ 10sec)

```
geneList <- c("HLA-A","HLA-B","HLA-C","HLA-DRA","HLA-DPA1","HLA-DRB1",
              "ACTB","GAPDH")
plotmatrix <- genecircosPlot(data_object=palmo_obj, geneList=geneList,
                             colorThreshold=15)

#order by user-defined group order
celltype_oi <- c("CD4_Naive","CD4_TEM","CD4_TCM","CD4_CTL",
                 "CD8_Naive","CD8_TEM","CD8_TCM","Treg","MAIT","gdT",
```

27

```
            "NK", "NK_CD56bright",
            "B_naive","B_memory", "B_intermediate",
            "CD14_Mono","CD16_Mono",
            "cDC2","pDC")
plotmatrix <- genecircosPlot(data_object=palmo_obj, geneList=geneList,
                             group_oi=celltype_oi, colorThreshold=15)
```
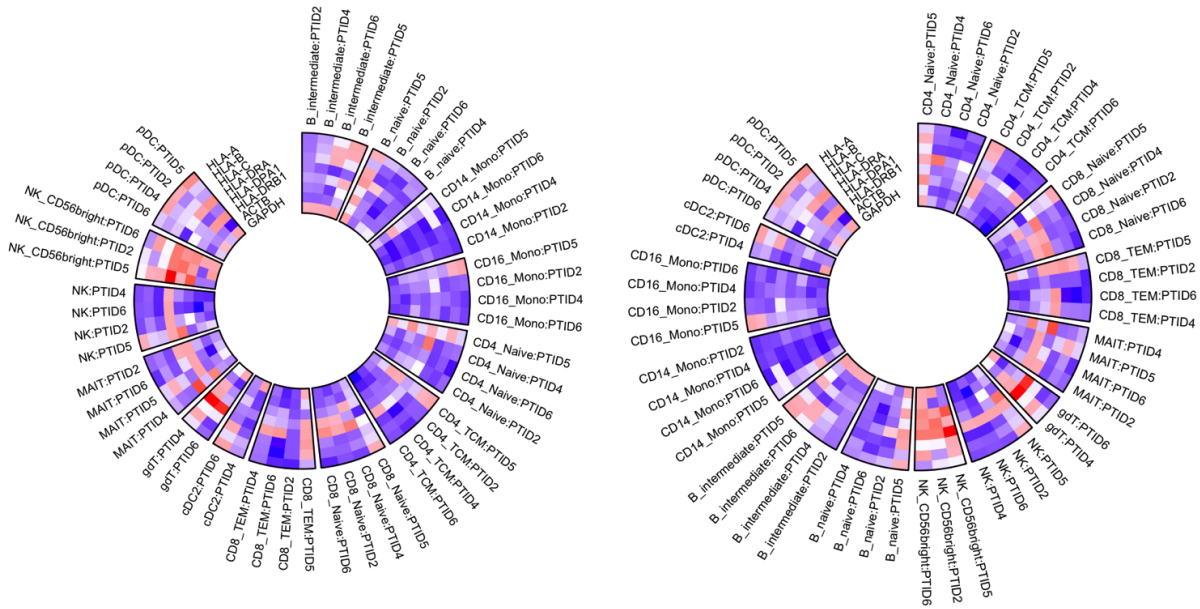
6 Session info

```r
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             listenv_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                  RColorBrewer_1.1-2
#>  [65] ellipsis_0.3.2              Seurat_4.0.0
#>  [67] factoextra_1.0.7.999        ica_1.0-2
#>  [69] farver_2.1.0                pkgconfig_2.0.3
#>  [71] uwot_0.1.10                 dbplyr_1.4.4
#>  [73] deldir_1.0-6                utf8_1.2.2
#>  [75] tidyselect_1.1.1            rlang_0.4.12
#>  [77] reshape2_1.4.4              later_1.3.0
#>  [79] munsell_0.5.0               cellranger_1.1.0
#>  [81] tools_4.0.3                 generics_0.1.1
#>  [83] broom_0.7.2                 ggridges_0.5.3
#>  [85] evaluate_0.14               stringr_1.4.0
#>  [87] fastmap_1.1.0               yaml_2.2.1
#>  [89] goftest_1.2-3               knitr_1.30
#>  [91] fs_1.5.0                    fitdistrplus_1.1-6
#>  [93] purrr_0.3.4                 RANN_2.6.1
#>  [95] pbapply_1.5-0               future_1.23.0
#>  [97] nlme_3.1-149                mime_0.12
#>  [99] xml2_1.3.2                  compiler_4.0.3
#> [101] plotly_4.10.0               png_0.1-7
#> [103] spatstat.utils_2.2-0        reprex_0.3.0
#> [105] tweenr_1.0.1                tibble_3.1.6
#> [107] statmod_1.4.35              stringi_1.7.5
#> [109] forcats_0.5.0               lattice_0.20-41
#> [111] Matrix_1.3-4                nloptr_1.2.2.2
#> [113] vctrs_0.3.8                 pillar_1.6.4
#> [115] lifecycle_1.0.1             lmtest_0.9-39
#> [117] GlobalOptions_0.1.2         RcppAnnoy_0.0.19
#> [119] bitops_1.0-7                data.table_1.14.2
#> [121] cowplot_1.1.1               irlba_2.3.3
#> [123] GenomicRanges_1.40.0        httpuv_1.6.3
#> [125] patchwork_1.1.1             R6_2.5.1
#> [127] promises_1.2.0.1            KernSmooth_2.23-17
#> [129] gridExtra_2.3               IRanges_2.22.2
#> [131] parallelly_1.28.1           codetools_0.2-16
#> [133] boot_1.3-25                 MASS_7.3-53
#> [135] assertthat_0.2.1            SummarizedExperiment_1.18.2
#> [137] MAST_1.14.0                 rjson_0.2.20
#> [139] SeuratObject_4.0.2          sctransform_0.3.2
#> [141] GenomeInfoDbData_1.2.3      S4Vectors_0.26.1
#> [143] mgcv_1.8-33                 parallel_4.0.3
#> [145] hms_0.5.3                   rpart_4.1-15
#> [147] tidyverse_1.3.0             tidyr_1.1.4
#> [149] minqa_1.2.4                 rmarkdown_2.5
#> [151] Rtsne_0.15                  ggforce_0.3.2
#> [153] Biobase_2.48.0              shiny_1.7.1
#> [155] lubridate_1.7.9
```