scRNAseq analysis

Abinet Joan

2024-10-18 11:10:20 +0200

Loading packages

```
suppressMessages(library(dplyr))
suppressMessages(library(Seurat))
suppressMessages(library(patchwork))
suppressMessages(library(ggplot2))
suppressMessages(library(SingleR))
suppressMessages(library(formatR))
```

Each sample correspond to one experimental conditions of lung myeloid cells, defined as living singlet CD45+, F4/80+, and/or CD11b+ cells from lung single-cell suspensions pooled from five C57BL/6 male WT mice, either mock-infected (Myeloid-Ctrl) or PR8-infected at day 10 after IAV (2 biological replicates: Myeloid-PR81 and Myeloid-PR82).

The Mouse1, Mouse2 and Mouse3 from GEO (GSE244727) correspond respectively to Myeloid-Ctrl, Myeloid-PR81 and Myeloid-PR82

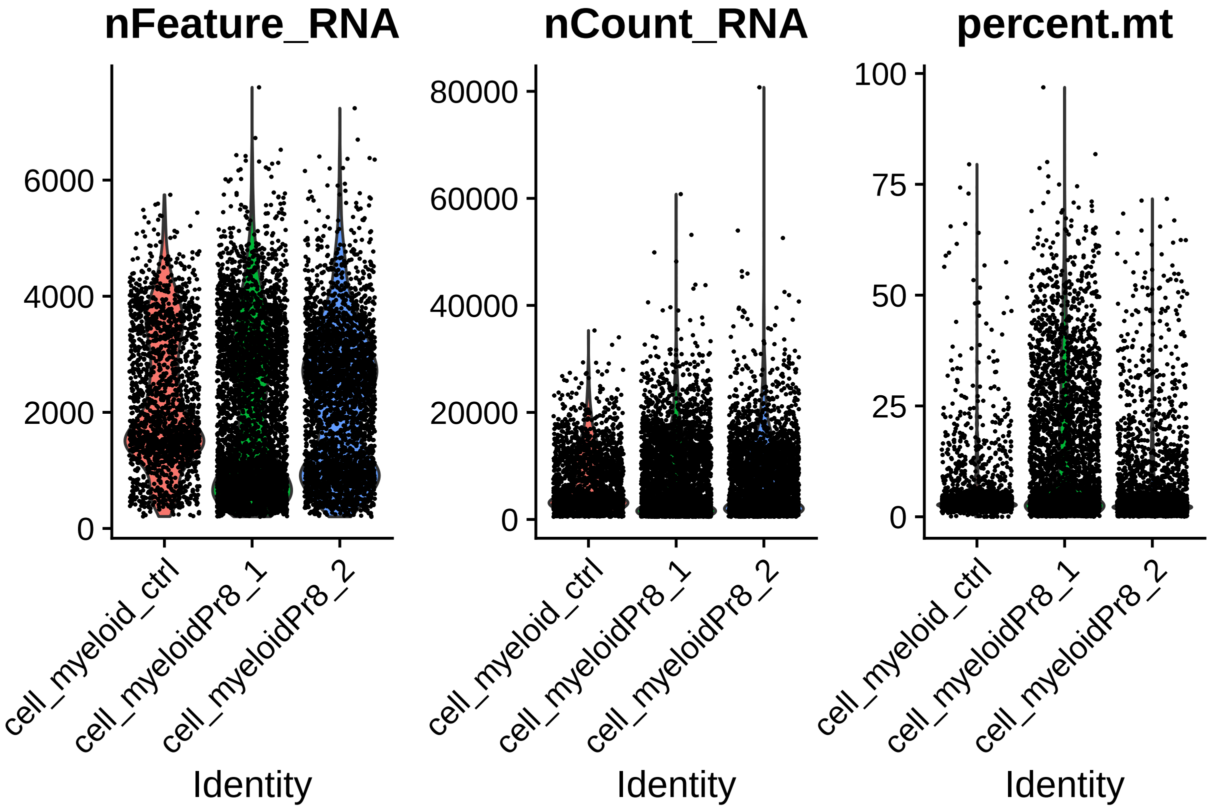
It is required to create 3 folders, one per mouse, and to rename the files in each folder matrix.mtx, features.tsv and barcodes.tsv.

Loading data

```
myeloid_ctrl.data <- Read10X(data.dir = "Counts/mouse1/")  # it is require to specify 10x
directory
cell_MyeloidCtrl <- CreateSeuratObject(counts = myeloid_ctrl.data$`Gene Expression`,
    project = "cell_myeloid_ctrl", min.cells = 3, min.features = 200)

myeloid_PR81.data <- Read10X(data.dir = "Counts/mouse2/")
cell_myeloidPr8_1 <- CreateSeuratObject(counts = myeloid_PR81.data$`Gene Expression`,
    project = "cell_myeloidPr8_1", min.cells = 3, min.features = 200)

myeloid_PR82.data <- Read10X(data.dir = "Counts/mouse3/")
cell_myeloidPr8_2 <- CreateSeuratObject(counts = myeloid_PR82.data$`Gene Expression`,
    project = "cell_myeloidPr8_2", min.cells = 3, min.features = 200)

# Merge the 3 seurat object
myeloid_cell <- merge(cell_MyeloidCtrl, y = c(cell_myeloidPr8_1, cell_myeloidPr8_2),
    add.cell.ids = c("1", "2", "3"), project = "Myeloid_cells")
```

Quality control

```
myeloid_cell[["percent.mt"]] <- PercentageFeatureSet(myeloid_cell, pattern = "^mt-")
VlnPlot(myeloid_cell, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"), ncol = 3,
    pt.size = 0.1)
```



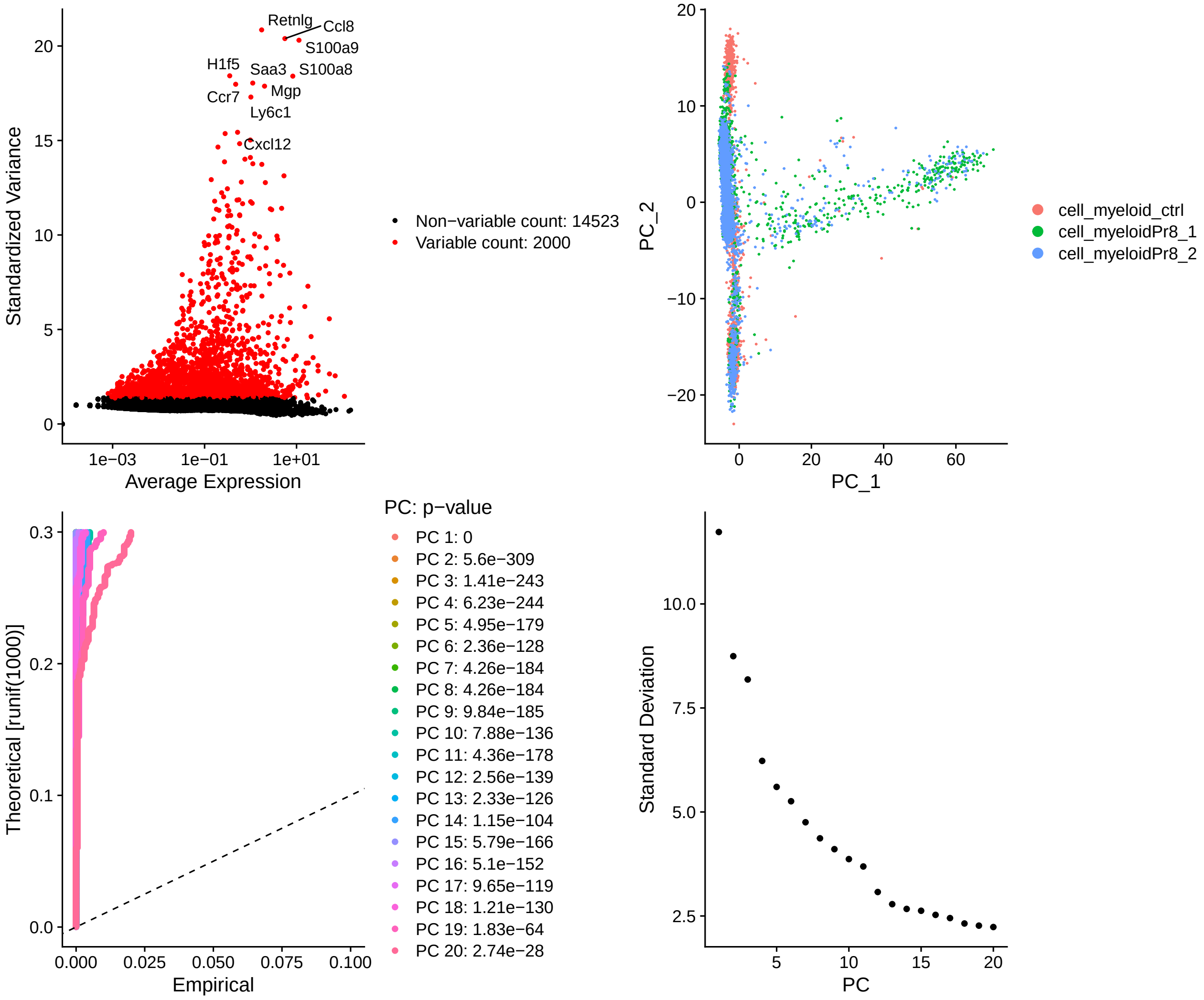
Pre-processing workflow

```
# removing low quality cells
myeloid_cell <- subset(myeloid_cell, subset = nFeature_RNA > 200 & nFeature_RNA <
    5000 & nCount_RNA < 20000 & percent.mt < 10)

# Normalizing the data
myeloid_cell <- NormalizeData(myeloid_cell, normalization.method = "LogNormalize",
    scale.factor = 10000)

# Feature selection
myeloid_cell <- FindVariableFeatures(myeloid_cell, selection.method = "vst", nfeatures =
2000)

# Identify the 10 most highly variable genes
top10 <- head(VariableFeatures(myeloid_cell), 10)

# plot variable features without labels
plot1 <- VariableFeaturePlot(myeloid_cell)
plot1 <- LabelPoints(plot = plot1, points = top10, repel = TRUE)
```

```r
# Scaling the data
all.genes <- rownames(myeloid_cell)
myeloid_cell <- ScaleData(myeloid_cell, features = all.genes)

# Linear dimensional reduction
myeloid_cell <- RunPCA(myeloid_cell, features = VariableFeatures(object = myeloid_cell))
plot2 <- DimPlot(myeloid_cell, reduction = "pca")

# Determine the ‘dimensionality’ of the dataset
myeloid_cell <- JackStraw(myeloid_cell, num.replicate = 100)
myeloid_cell <- ScoreJackStraw(myeloid_cell, dims = 1:20)
plot3 <- JackStrawPlot(myeloid_cell, dims = 1:20)
plot4 <- ElbowPlot(myeloid_cell, ndims = 20)

# Cluster cells in umap
myeloid_cell <- FindNeighbors(myeloid_cell, dims = 1:15)
myeloid_cell <- FindClusters(myeloid_cell, resolution = 0.25)
myeloid_cell <- RunUMAP(myeloid_cell, dims = 1:15)
```

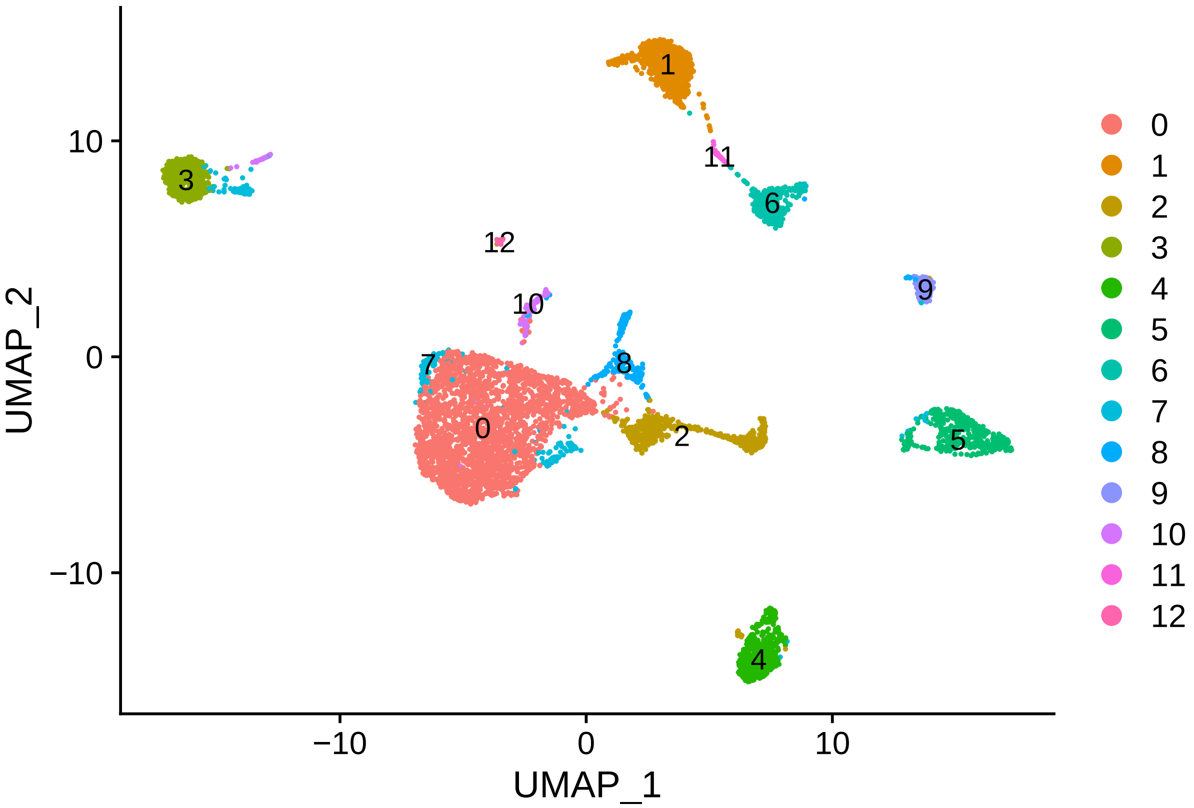
```r
plot1 + plot2 + plot3 + plot4
```

```
## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Removed 28000 rows containing missing values (`geom_point()`).
```



# Visualizing all clusters

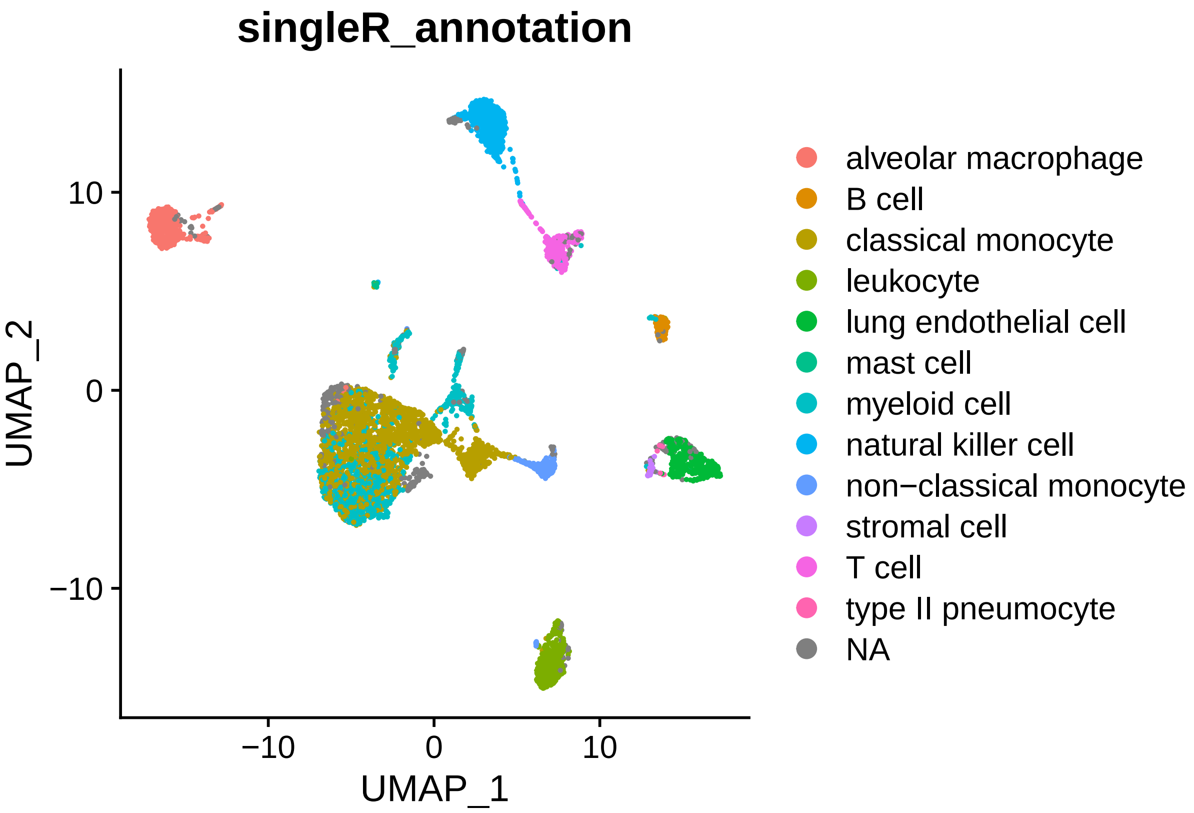
```
DimPlot(myeloid_cell, reduction = "umap", label = T)
```



Cell annotation

```
library(SingleR)
library(ExperimentHub)
library(scuttle)

eh <- ExperimentHub()
query(eh, "TabulaMurisData")

ref <- eh[["EH1617"]]
myeloid_ref <- ref[, !is.na(ref$cell_ontology_class)]
myeloid_ref <- myeloid_ref[, myeloid_ref$tissue == "Lung"]

myeloid_ref <- logNormCounts(myeloid_ref)

tested_data <- as.SingleCellExperiment(myeloid_cell)
tested_data <- logNormCounts(tested_data)

results <- SingleR(test = tested_data, ref = myeloid_ref, labels =
myeloid_ref$cell_ontology_class)
cell_annotations <- results

myeloid_cell[["singleR_annotation"]] <- cell_annotations[, c(4)]
```

```
DimPlot(myeloid_cell, reduction = "umap", group.by = "singleR_annotation")
```



Removing the variable used for annotation

```
rm(tested_data)
rm(myeloid_ref)
```

The next step is to subset the clusters of myeloid cells (Macropages, neutrophils, DCs). The others clusters will be removed.

Removing contamination

```
myeloid_cells <- subset(myeloid_cell, seurat_clusters %in% c(0, 2, 3, 4, 7, 8, 10))
```

Saving file

```
saveRDS(myeloid_cells, "Myeloid_cells_Part1.rds")
```

```
sessionInfo()
```

```
## R version 4.3.3 (2024-02-29)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.4 LTS
##
```

```
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;  LAPACK version 3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=fr_BE.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=fr_BE.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=fr_BE.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=fr_BE.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Brussels
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] TabulaMurisData_1.18.0      scuttle_1.10.1
##  [3] SingleCellExperiment_1.22.0 ExperimentHub_2.8.0
##  [5] AnnotationHub_3.8.0         BiocFileCache_2.8.0
##  [7] dbplyr_2.3.2                formatR_1.14
##  [9] SingleR_2.2.0               SummarizedExperiment_1.30.2
## [11] Biobase_2.60.0              GenomicRanges_1.52.0
## [13] GenomeInfoDb_1.36.0         IRanges_2.34.0
## [15] S4Vectors_0.38.1            BiocGenerics_0.46.0
## [17] MatrixGenerics_1.12.2       matrixStats_1.0.0
## [19] ggplot2_3.4.2               patchwork_1.1.2
## [21] SeuratObject_4.1.3          Seurat_4.3.0
## [23] dplyr_1.1.2
##
## loaded via a namespace (and not attached):
##   [1] RcppAnnoy_0.0.21              splines_4.3.3
##   [3] later_1.3.1                   filelock_1.0.2
##   [5] bitops_1.0-7                  tibble_3.2.1
##   [7] R.oo_1.25.0                   polyclip_1.10-4
##   [9] lifecycle_1.0.3               globals_0.16.2
##  [11] lattice_0.22-5                MASS_7.3-60.0.1
##  [13] magrittr_2.0.3                plotly_4.10.2
##  [15] rmarkdown_2.23                yaml_2.3.7
##  [17] httpuv_1.6.11                 sctransform_0.3.5
##  [19] spam_2.9-1                    sp_2.0-0
##  [21] spatstat.sparse_3.0-2         reticulate_1.30
##  [23] cowplot_1.1.1                 pbapply_1.7-2
##  [25] DBI_1.1.3                     RColorBrewer_1.1-3
##  [27] abind_1.4-5                   zlibbioc_1.46.0
##  [29] Rtsne_0.16                    purrr_1.0.1
##  [31] R.utils_2.12.2                RCurl_1.98-1.12
##  [33] rappdirs_0.3.3                GenomeInfoDbData_1.2.10
##  [35] ggrepel_0.9.3                 irlba_2.3.5.1
##  [37] listenv_0.9.0                 spatstat.utils_3.0-3
##  [39] goftest_1.2-3                 spatstat.random_3.1-5
##  [41] fitdistrplus_1.1-11           parallelly_1.36.0
##  [43] DelayedMatrixStats_1.22.1     leiden_0.4.3
##  [45] codetools_0.2-19              DelayedArray_0.26.3
##  [47] tidyselect_1.2.0              farver_2.1.1
##  [49] ScaledMatrix_1.8.1            spatstat.explore_3.2-1
##  [51] jsonlite_1.8.7                ellipsis_0.3.2
##  [53] progressr_0.13.0              ggridges_0.5.4
##  [55] survival_3.5-8                tools_4.3.3
##  [57] ica_1.0-3                     Rcpp_1.0.11
##  [59] glue_1.6.2                    gridExtra_2.3
##  [61] xfun_0.39                     withr_2.5.0
##  [63] BiocManager_1.30.21           fastmap_1.1.1
##  [65] fansi_1.0.4                   digest_0.6.33
##  [67] rsvd_1.0.5                    R6_2.5.1
##  [69] mime_0.12                     colorspace_2.1-0
##  [71] scattermore_1.2               tensor_1.5
##  [73] RSQLite_2.3.1                 spatstat.data_3.0-1
##  [75] R.methodsS3_1.8.2             utf8_1.2.3
##  [77] tidyr_1.3.0                   generics_0.1.3
##  [79] data.table_1.14.8             httr_1.4.6
##  [81] htmlwidgets_1.6.2             S4Arrays_1.2.1
##  [83] uwot_0.1.16                   pkgconfig_2.0.3
##  [85] gtable_0.3.3                  blob_1.2.4
##  [87] lmtest_0.9-40                 XVector_0.40.0
##  [89] htmltools_0.5.5               dotCall64_1.0-2
##  [91] scales_1.2.1                  png_0.1-8
##  [93] knitr_1.43                    rstudioapi_0.14
##  [95] reshape2_1.4.4                curl_5.0.1
##  [97] nlme_3.1-164                  cachem_1.0.8
##  [99] zoo_1.8-12                    stringr_1.5.0
## [101] BiocVersion_3.17.1            KernSmooth_2.23-22
## [103] parallel_4.3.3                miniUI_0.1.1.1
## [105] vipor_0.4.5                   AnnotationDbi_1.62.1
## [107] ggrastr_1.0.2                 pillar_1.9.0
## [109] grid_4.3.3                    vctrs_0.6.3
## [111] RANN_2.6.1                    promises_1.2.0.1
## [113] BiocSingular_1.16.0           beachmat_2.16.0
## [115] xtable_1.8-4                  cluster_2.1.6
## [117] beeswarm_0.4.0                evaluate_0.21
## [119] cli_3.6.1                     compiler_4.3.3
## [121] rlang_1.1.1                   crayon_1.5.2
## [123] future.apply_1.11.0           labeling_0.4.2
## [125] plyr_1.8.8                    ggbeeswarm_0.7.2
## [127] stringi_1.7.12                viridisLite_0.4.2
## [129] deldir_1.0-9                  BiocParallel_1.34.2
## [131] Biostrings_2.68.1             munsell_0.5.0
## [133] lazyeval_0.2.2                spatstat.geom_3.2-4
## [135] Matrix_1.6-1                  bit64_4.0.5
## [137] sparseMatrixStats_1.12.0      future_1.33.0
## [139] KEGGREST_1.40.0               shiny_1.7.4.1
## [141] interactiveDisplayBase_1.38.0 highr_0.10
## [143] ROCR_1.0-11                   memoise_2.0.1
## [145] igraph_1.5.0.1                bit_4.0.5
```