

Cell Type Anotation

Cell type annotation for the merged Seurat object

Ximing Ran

2025-04-09

Contents

1. Read the merged Data	2
-------------------------	---

2. Culstering	3
---------------	---

```
library(tibble)
library(tidyr)
library(dplyr)
library(rtracklayer)
library(dplyr)
library(Seurat)
library(Signac)
library(EnsDb.Hsapiens.v86)
library(ggplot2)
library(cowplot)
library(simspec)
library(cowplot)
library(AnnotationHub)
library(Seurat)
library(tidyverse)
library(openxlsx)
library(dplyr)
library(conflicted)
library(harmony)
library(data.table)
library(pheatmap)
library(ggplot2)
library(patchwork)

conflict_prefer("filter", "dplyr") # Prefer dplyr's filter()
conflict_prefer("lag", "dplyr")    # Prefer dplyr's lag()

# load function from local files

# load function from local files
load(here::here("data","reference","annotations.rdata")) # load the annotations
```

1. Read the merged Data

```
seurat_obj <- readRDS(here::here("data", "processed-data", "seurat-merged_2025-03-15.rds"))

##### Dimension reduction #####
DefaultAssay(seurat_obj) <- "RNA"
seurat_obj <- NormalizeData(seurat_obj)
seurat_obj <- FindVariableFeatures(seurat_obj, nfeatures = 3000)
seurat_obj <- ScaleData(seurat_obj)
seurat_obj <- RunPCA(seurat_obj, npcs = 50)
seurat_obj <- RunUMAP(seurat_obj, dims = 1:20, reduction.name = "umap_rna", reduction.key = "UMAPRNA_")

seurat_obj <- RunHarmony(seurat_obj, group.by.vars = "orig.ident")
seurat_obj <- RunUMAP(seurat_obj, reduction = "harmony", dims = 1:20,
                      reduction.name = "umap_harmony", reduction.key = "UMAPH_")
```

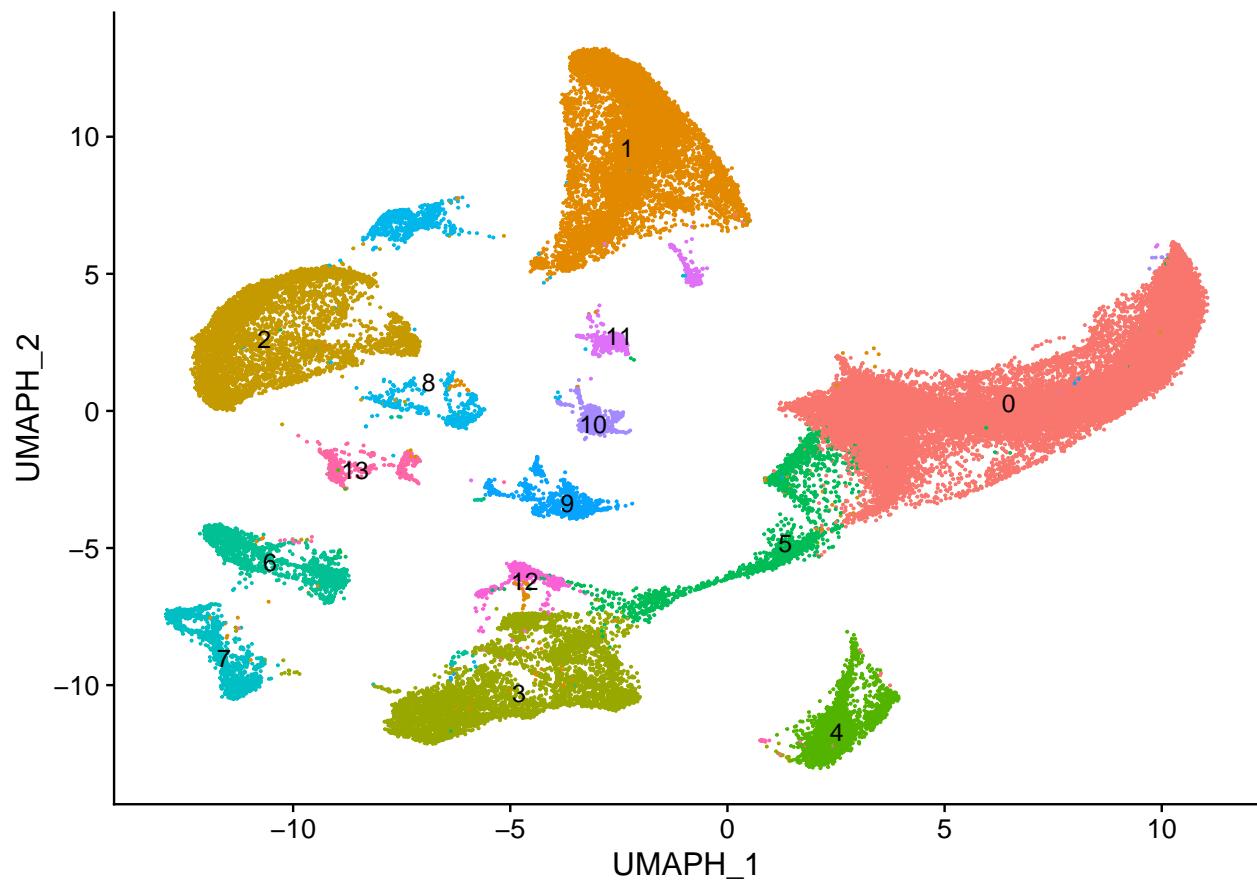
2. Culstering

```
#### Run clustering res from 0.1 to 0.5 ####
for (res in seq(0.1, 0.5, by = 0.1)) {
  seurat_obj <- FindNeighbors(seurat_obj, reduction = "harmony", dims = 1:20)
  seurat_obj <- FindClusters(seurat_obj, resolution = res)

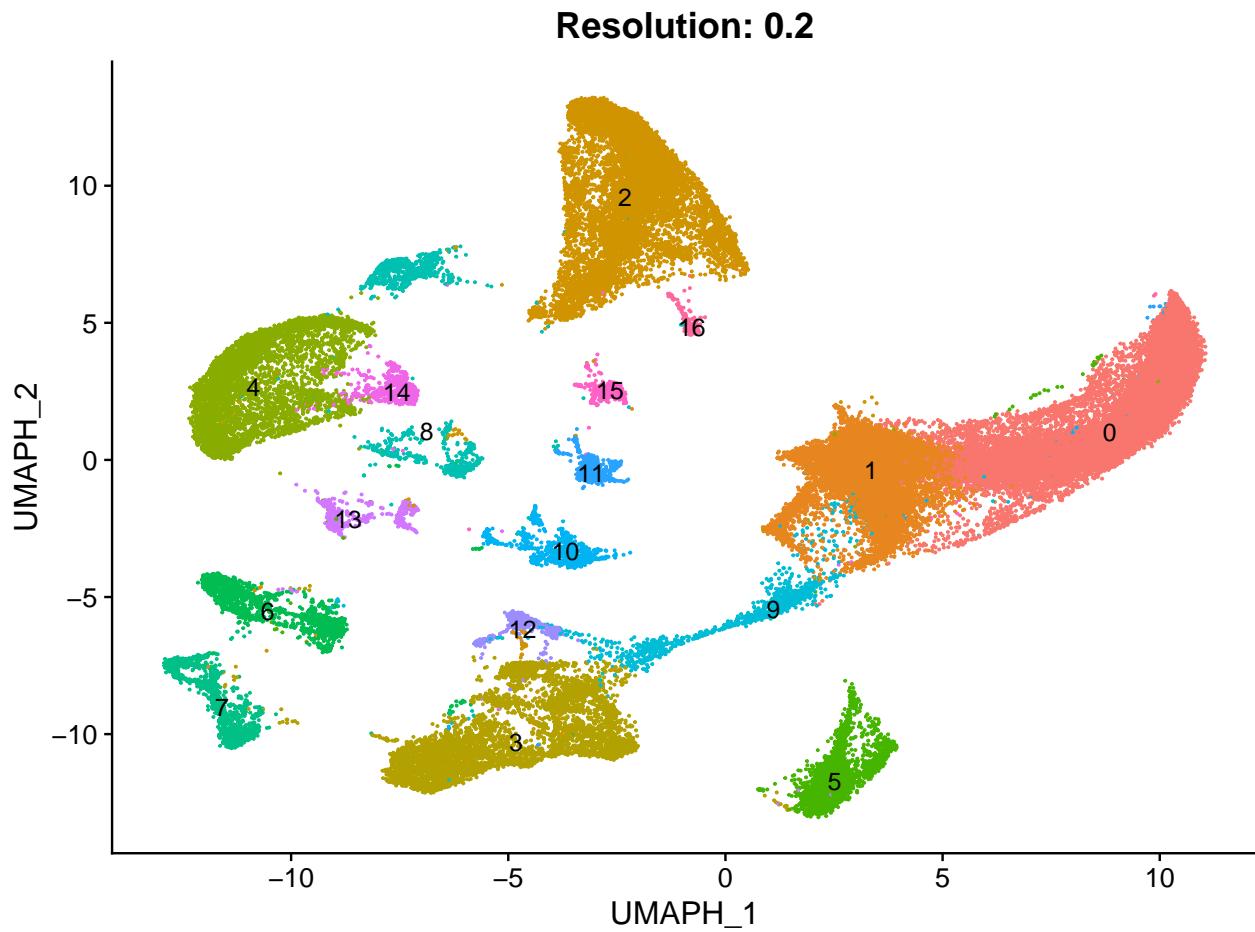
  # plot the UMAP with the cluster labels
  p <- DimPlot(seurat_obj, reduction = "umap_harmony", group.by = "seurat_clusters", label = TRUE) +
    ggtitle(paste("Resolution:", res)) +
    theme(legend.position = "none")
  print(p)
  ggsave(paste0("result/01-Clustering/umap_plot/umap_harmony_res_", res, ".pdf"), plot = p, width = 8, height = 6)
  ggsave(paste0("result/01-Clustering/umap_plot/umap_harmony_res_", res, ".png"), plot = p, width = 8, height = 6)
}

## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 58119
## Number of edges: 1880350
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.9797
## Number of communities: 14
## Elapsed time: 7 seconds
```

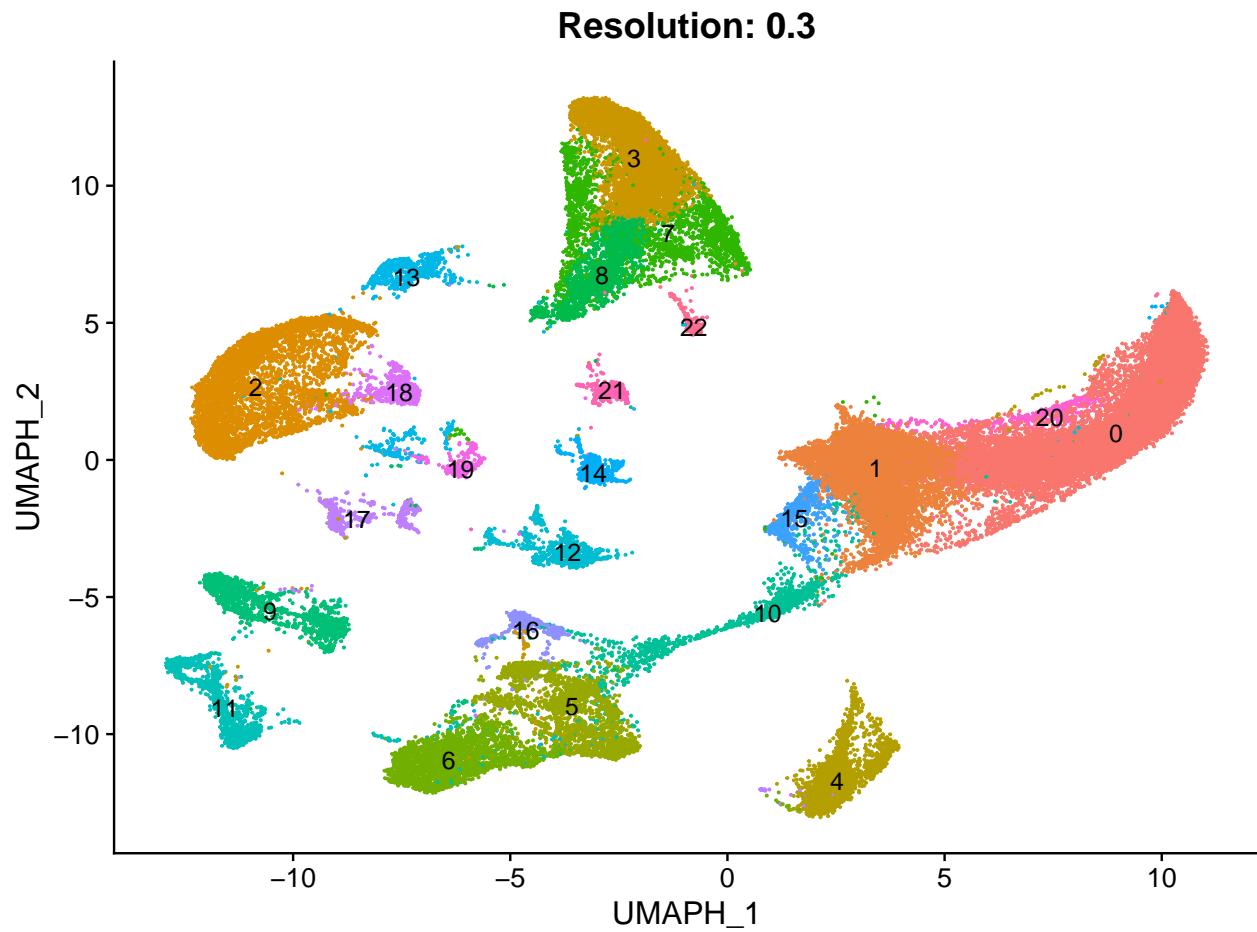
Resolution: 0.1



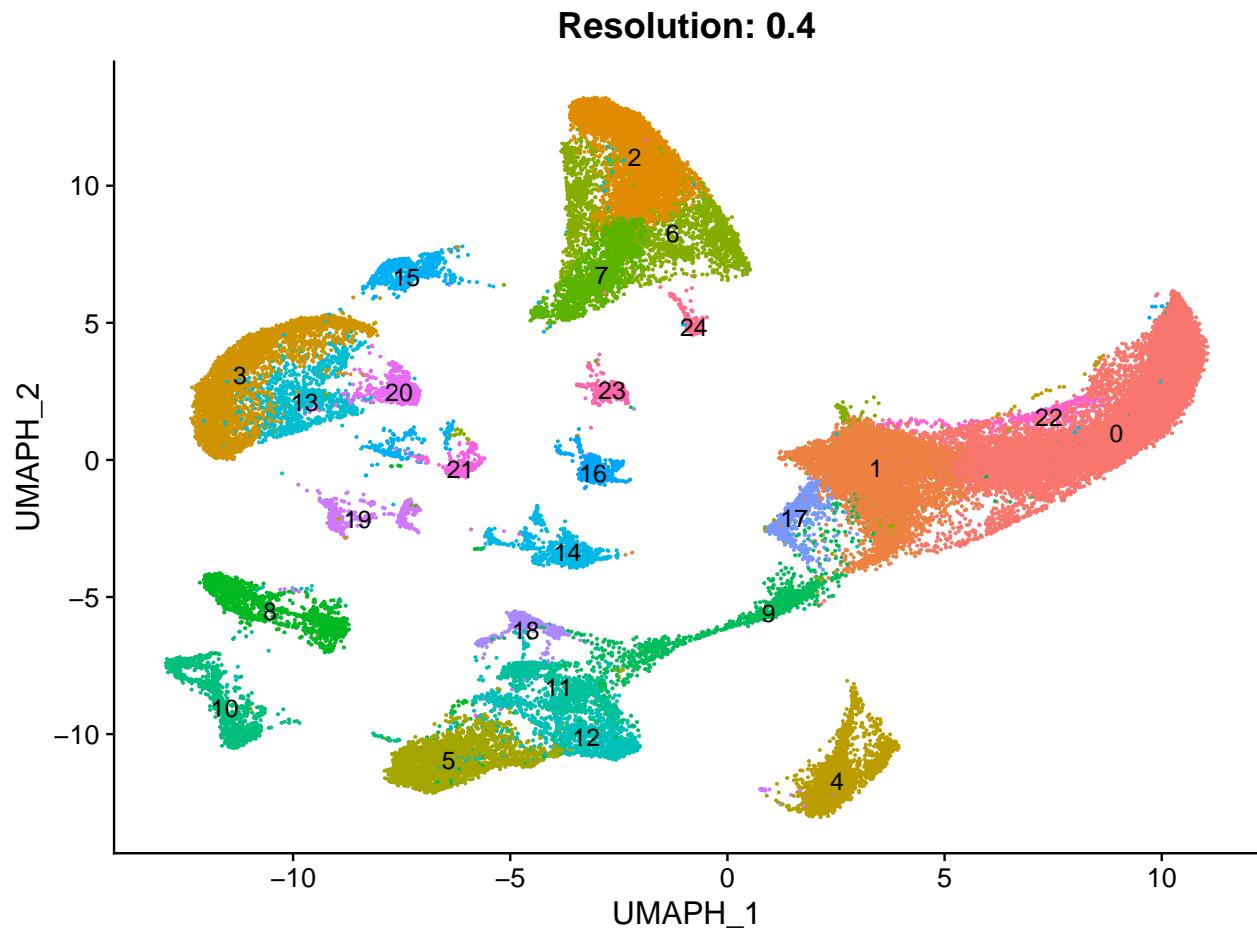
```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 58119
## Number of edges: 1880350
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.9669
## Number of communities: 17
## Elapsed time: 8 seconds
```



```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 58119
## Number of edges: 1880350
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.9583
## Number of communities: 23
## Elapsed time: 8 seconds
```

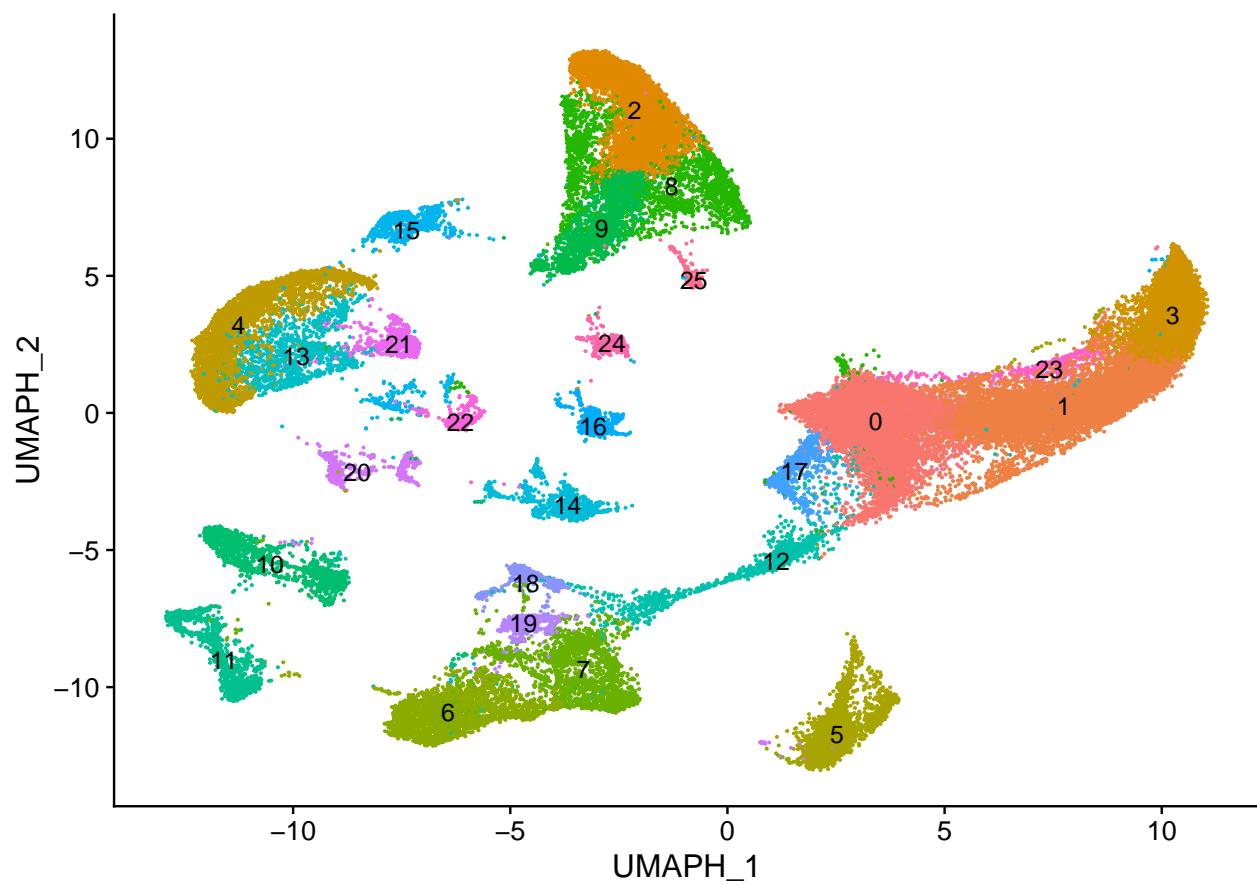


```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 58119
## Number of edges: 1880350
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.9512
## Number of communities: 25
## Elapsed time: 8 seconds
```



```
## Modularity Optimizer version 1.3.0 by Ludo Waltman and Nees Jan van Eck
##
## Number of nodes: 58119
## Number of edges: 1880350
##
## Running Louvain algorithm...
## Maximum modularity in 10 random starts: 0.9444
## Number of communities: 26
## Elapsed time: 8 seconds
```

Resolution: 0.5



```
# save the seurat object  
saveRDS(seurat_obj, file = here::here("data", "processed-data", "seurat-cluster_2025-04-09.rds"))
```

```

sessionInfo()

## R version 4.4.0 (2024-04-24)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sonoma 14.3.1
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;  LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats4      stats       graphics   grDevices  utils      datasets   methods
## [8] base
##
## other attached packages:
##  [1] patchwork_1.3.0          pheatmap_1.0.12
##  [3] data.table_1.17.0        harmony_1.2.3
##  [5] Rcpp_1.0.14              conflicted_1.2.0
##  [7] openxlsx_4.2.8           AnnotationHub_3.12.0
##  [9] BiocFileCache_2.12.0     dbplyr_2.5.0
## [11] simspec_0.0.0.9000       cowplot_1.1.3
## [13] EnsDb.Hsapiens.v86_2.99.0 ensemblDb_2.28.1
## [15] AnnotationFilter_1.28.0  GenomicFeatures_1.56.0
## [17] AnnotationDbi_1.66.0     Biobase_2.64.0
## [19] Signac_1.14.0            Seurat_5.2.1
## [21] SeuratObject_5.0.2       sp_2.2-0
## [23] rtracklayer_1.64.0       GenomicRanges_1.56.2
## [25] GenomeInfoDb_1.40.1      IRanges_2.38.1
## [27] S4Vectors_0.42.1         BiocGenerics_0.50.0
## [29] knitr_1.50                lubridate_1.9.4
## [31]forcats_1.0.0             stringr_1.5.1
## [33] dplyr_1.1.4              purrr_1.0.4
## [35] readr_2.1.5              tidyverse_1.3.1
## [37] tibble_3.2.1             ggplot2_3.5.1
## [39] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] RcppAnnoy_0.0.22          splines_4.4.0
##  [3] later_1.4.2               BiocIO_1.14.0
##  [5] filelock_1.0.3            bitops_1.0-9
##  [7] polyclip_1.10-7           XML_3.99-0.18
##  [9] fastDummies_1.7.5          lifecycle_1.0.4
## [11] rprojroot_2.0.4            globals_0.16.3
## [13] lattice_0.22-7            MASS_7.3-65
## [15] magrittr_2.0.3             plotly_4.10.4
## [17] rmarkdown_2.29              yaml_2.3.10
## [19] httpuv_1.6.15              sctransform_0.4.1
## [21] zip_2.3.2                 spam_2.11-1

```

```

## [23] spatstat.sparse_3.1-0      reticulate_1.42.0
## [25] pbapply_1.7-2             DBI_1.2.3
## [27] RColorBrewer_1.1-3       abind_1.4-8
## [29] zlibbioc_1.50.0          Rtsne_0.17
## [31] RCurl_1.98-1.17          rappdirs_0.3.3
## [33] GenomeInfoDbData_1.2.12  ggrepel_0.9.6
## [35] irlba_2.3.5.1           listenv_0.9.1
## [37] spatstat.utils_3.1-3      goftest_1.2-3
## [39] RSpectra_0.16-2          spatstat.random_3.3-3
## [41] fitdistrplus_1.2-2        parallelly_1.43.0
## [43] codetools_0.2-20          DelayedArray_0.30.1
## [45] RcppRoll_0.3.1           tidyselect_1.2.1
## [47] UCSC.utils_1.0.0          farver_2.1.2
## [49] matrixStats_1.5.0         spatstat.explore_3.4-2
## [51] GenomicAlignments_1.40.0  jsonlite_2.0.0
## [53] progressr_0.15.1          ggridges_0.5.6
## [55] survival_3.8-3           systemfonts_1.2.2
## [57] tools_4.4.0               ragg_1.3.3
## [59] ica_1.0-3                 glue_1.8.0
## [61] gridExtra_2.3              SparseArray_1.4.8
## [63] here_1.0.1                xfun_0.52
## [65] MatrixGenerics_1.16.0     withr_3.0.2
## [67] BiocManager_1.30.25       fastmap_1.2.0
## [69] digest_0.6.37             timechange_0.3.0
## [71] R6_2.6.1                  mime_0.13
## [73] textshaping_1.0.0         colorspace_2.1-1
## [75] scattermore_1.2            tensor_1.5
## [77] spatstat.data_3.1-6       RSQLite_2.3.9
## [79] RhpcBLASctl_0.23-42       generics_0.1.3
## [81] httr_1.4.7                htmlwidgets_1.6.4
## [83] S4Arrays_1.4.1            uwot_0.2.3
## [85] pkgconfig_2.0.3            gtable_0.3.6
## [87] blob_1.2.4                lmtest_0.9-40
## [89] XVector_0.44.0            htmltools_0.5.8.1
## [91] dotCall64_1.2              ProtGenerics_1.36.0
## [93] scales_1.3.0               png_0.1-8
## [95] spatstat.univar_3.1-2     rstudioapi_0.17.1
## [97] tzdb_0.5.0                reshape2_1.4.4
## [99] rjson_0.2.23              nlme_3.1-168
## [101] curl_6.2.2                cachem_1.1.0
## [103] zoo_1.8-13               BiocVersion_3.19.1
## [105] KernSmooth_2.23-26       parallel_4.4.0
## [107] miniUI_0.1.1.1           restfulr_0.0.15
## [109] pillar_1.10.2             grid_4.4.0
## [111] vctrs_0.6.5               RANN_2.6.2
## [113] promises_1.3.2            xtable_1.8-4
## [115] cluster_2.1.8.1           evaluate_1.0.3
## [117] tinytex_0.56              cli_3.6.4
## [119] compiler_4.4.0            Rsamtools_2.20.0
## [121] rlang_1.1.5               crayon_1.5.3
## [123] future.apply_1.11.3       labeling_0.4.3
## [125] plyr_1.8.9                stringi_1.8.7
## [127] viridisLite_0.4.2          deldir_2.0-4
## [129] BiocParallel_1.38.0        munsell_0.5.1

```

```
## [131] Biostrings_2.72.1           lazyeval_0.2.2
## [133] spatstat.geom_3.3-6         Matrix_1.7-3
## [135] RcppHNSW_0.6.0              hms_1.1.3
## [137] bit64_4.6.0-1               future_1.34.0
## [139] KEGGREST_1.44.1             shiny_1.10.0
## [141] SummarizedExperiment_1.34.0 ROCR_1.0-11
## [143] memoise_2.0.1               igraph_2.1.4
## [145] fastmatch_1.1-6             bit_4.6.0
```