# Differential Expression Analysis for bulk RNA-seq data $$\operatorname{CTRL}$$ Condition: Vehicle 100 vs 10

# Ximing Ran

## 2025-03-24

# Contents

1. Read the count data	2
2. Differential expression analysis	2
3. Visualization for reuslt         (1) Sample information	<b>4</b> 4 6
4. GSVA analysis	15
5. Pathway Enrichment Analysis	18
Session information	30
<pre>library(tibble) library(tidyr) library(dplyr) library(rtracklayer)</pre>	
# load function from local files	
<pre># load function from local files source(here::here("source", "DEG_functions.R"))</pre>	

#### 1. Read the count data

In this section, we will read the clean count data from the synaptosomes\_bulkRNA folder. We will read the data and merge them into a single table. The final table will be stored in ../dataresults/bulkRNA\_counts\_clean.csv.

```
input_count <- read.csv(here::here("data", "bulkRNA",</pre>
                                        "bulkRNA_counts_cleaned.csv"))
counts <- as.data.frame(input_count) %>%
  column_to_rownames(var = "gene")
colnames(counts) <- gsub("_", "-", colnames(counts))</pre>
# raw sample list
sample_list_raw <- read.csv(here::here("data", "bulkRNA",</pre>
                                        "sample_info_AD.csv")) %>%
                     mutate(condition = paste0(Diagosis, "_", Treatment),
                             sample = gsub("_", "-", Sample.name))
# Ensure the column names of counts exist in Sample.name
new_colnames <- sample_list_raw$Label[match(colnames(counts), sample_list_raw$sample )]</pre>
# Assign new column names
colnames(counts) <- new_colnames</pre>
# sort the columns by the colname
condition_list <- data.frame(</pre>
  group =sample list raw$condition
row.names(condition_list) <- sample_list_raw$Label</pre>
counts<- counts[, rownames(condition list)]</pre>
gene_name_mapping<- readRDS(here::here("data","ref" ,"gene_name_mapping.rds"))</pre>
```

## 2. Differential expression analysis

In this section, we will perform differential expression analysis using DESeq2. We will compare the 22q vs Control in the vehicle condition. The results will be stored in results/02-DEG-V\_10/DESeq2\_results.csv.

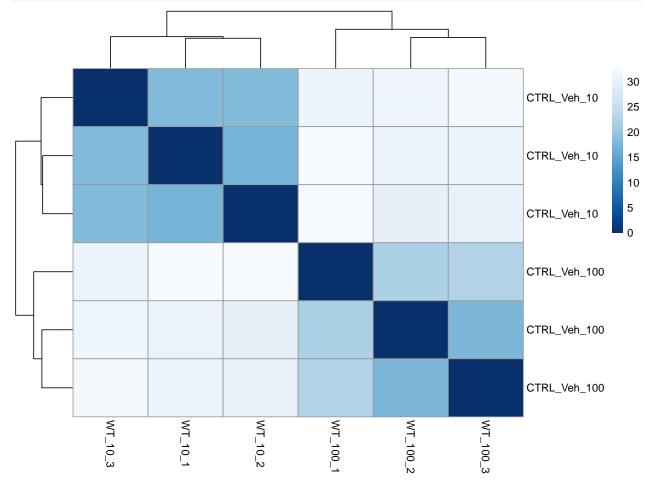
```
# Init the result folder structure for the result
result_folder_all = './results/02-DEG-CTRL_10_vs_100'
result_folder = result_folder_all
Result_folder_structure(result_folder)

# load the comparison group information
reference_group <- "CTRL_Veh_10"
compare_group <- "CTRL_Veh_100"
reference_group_short <- "CTRL_10"
compare_group_short <- "Veh_100"</pre>
```

```
filter_sample_info <- condition_list %>%
  filter(group %in% c(reference_group, compare_group))
filter_counts <- counts[, rownames(filter_sample_info)]</pre>
# Run the DESeq2 analysis
dds_obj <- DEAnalysis(counts =filter_counts,</pre>
                       reference_group = reference_group,
                       compare group = compare group,
                        condition_list = filter_sample_info,
                       target_gene = target_gene,
                       result_folder = result_folder)
res <- results(dds_obj)</pre>
resOrdered <- res[order(res$padj), ]
# omit the NA values
resOrdered <- resOrdered[!is.na(resOrdered$padj),]</pre>
dds_obj <- dds_obj[rownames(resOrdered),]</pre>
write.csv(resOrdered, file.path(result_folder, "02-DEG", "01_all_gene_results.csv"))
# DEG with log2fc > 1 and padj < 0.05
deg_1 <- resOrdered %>% as.data.frame() %>% rownames_to_column(var = "gene") %>%
 filter(padj < 0.05 & abs(log2FoldChange) > 1) %>% arrange(padj)
deg 1 <- deg 1[!is.na(deg 1$padj),]
write.csv(deg_1, file.path(result_folder, "02-DEG", "02_DEG_log2fc_1.csv"), row.names = FALSE)
# DEG with log2fc > 1.5 and padj < 0.05
deg_1.5 <- resOrdered %>% as.data.frame() %>% rownames_to_column(var = "gene") %>%
 filter(padj < 0.05 & abs(log2FoldChange) > 1.5) %>% arrange(padj)
deg_1.5 <- deg_1.5 [!is.na(deg_1.5 $padj),]
write.csv(deg_1.5 , file.path(result_folder, "02-DEG", "03_DEG_log2fc_1_5.csv"), row.names = FALSE)
print("DEG analysis is done")
## [1] "DEG analysis is done"
# Save the normalized counts
normalized_counts <- counts(dds_obj, normalized = TRUE)</pre>
write.csv(normalized_counts, file.path(result_folder, "02-DEG", "DESeq2_normalized_counts.csv"))
```

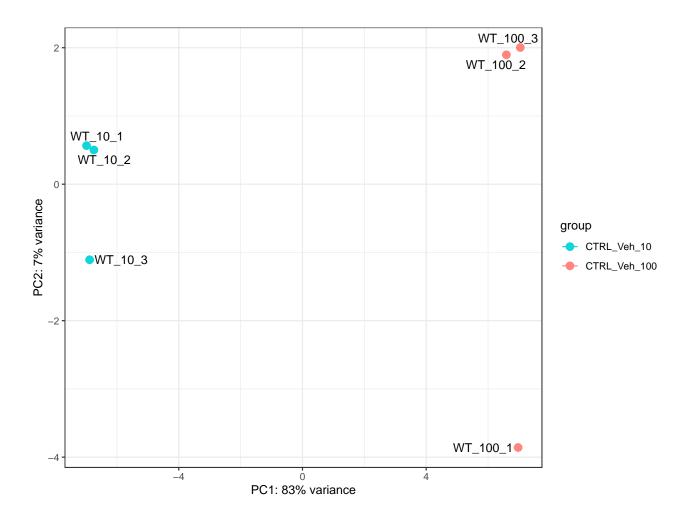
## 3. Visualization for reuslt

### (1) Sample information



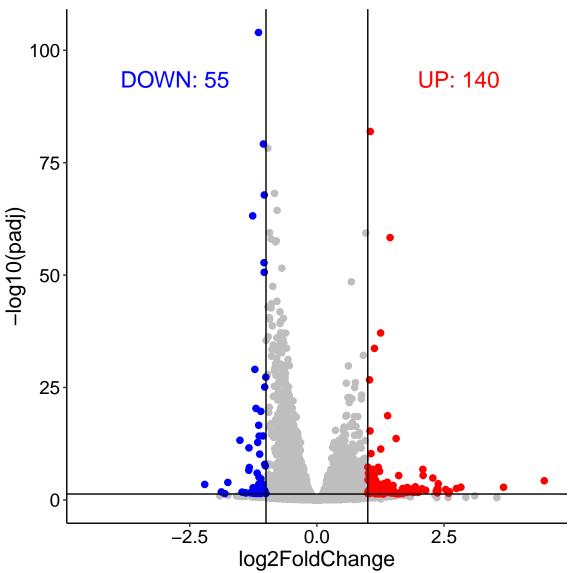
## [1] "Sample distance heatmap is done"

## [1] "PCA plot is done"



#### (2) DEG visualization - Volcano plot and Heatmap

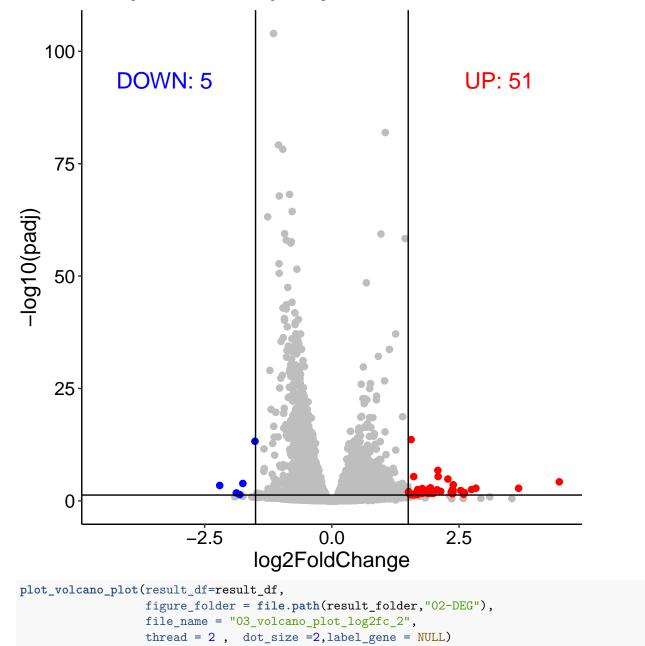
## [1] "Volcano plot for 02\_volcano\_plot\_log2fc\_1"



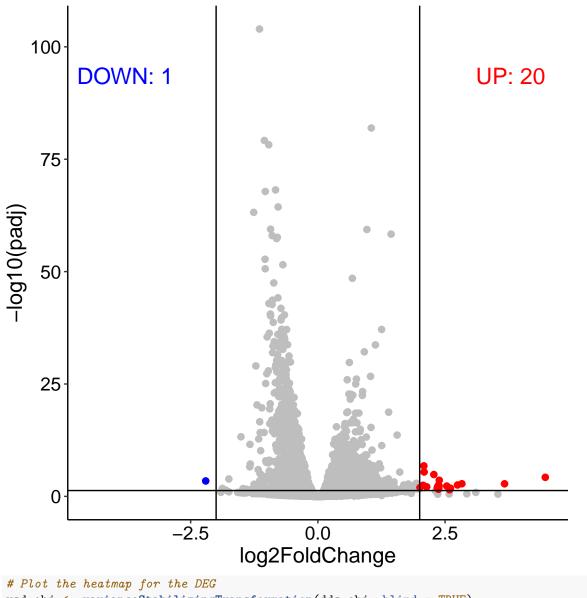
plot\_volcano\_plot(result\_df=result\_df,

```
figure_folder = file.path(result_folder,"02-DEG"),
file_name = "03_volcano_plot_log2fc_1.5",
thread = 1.5 , dot_size =2,label_gene = NULL)
```

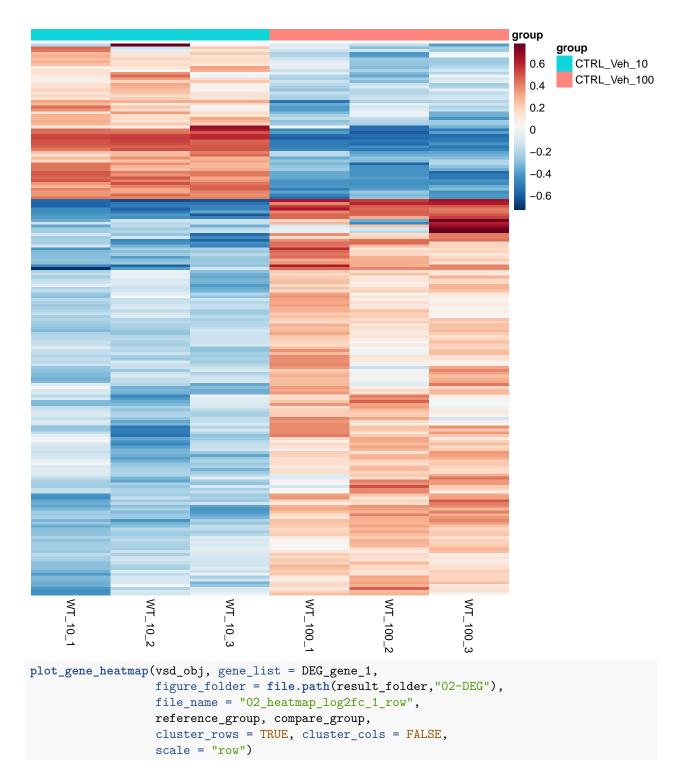
## [1] "Volcano plot for 03\_volcano\_plot\_log2fc\_1.5"



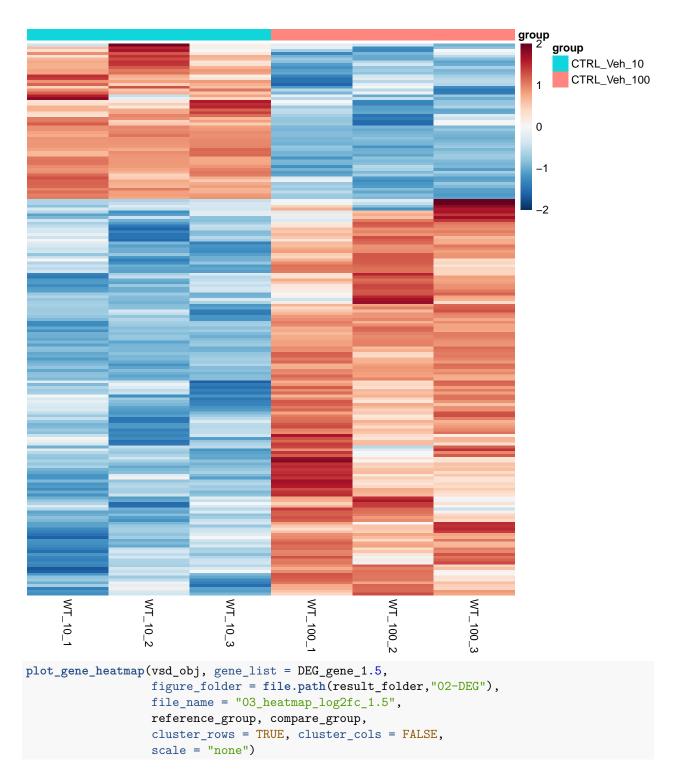
## [1] "Volcano plot for 03\_volcano\_plot\_log2fc\_2"



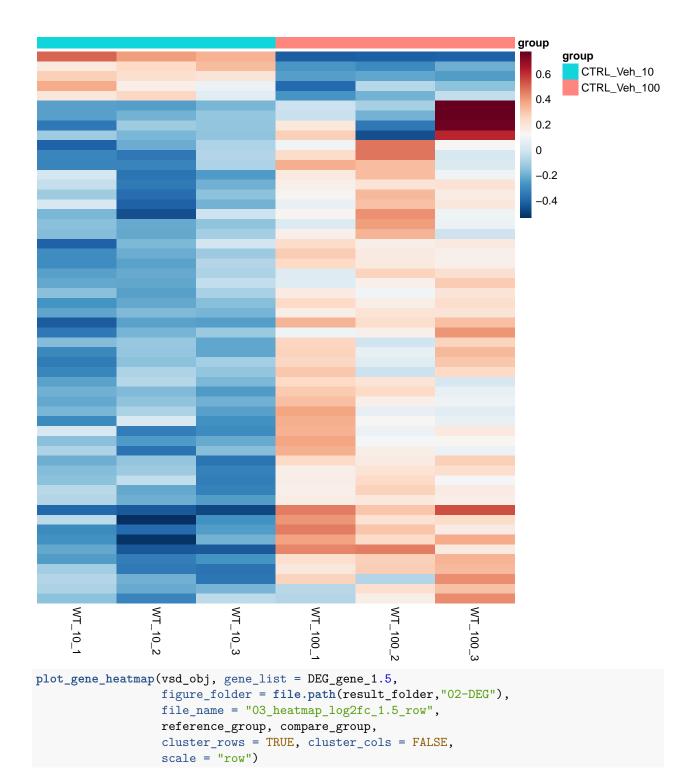
## [1] "Heatmap for 02\_heatmap\_log2fc\_1 "



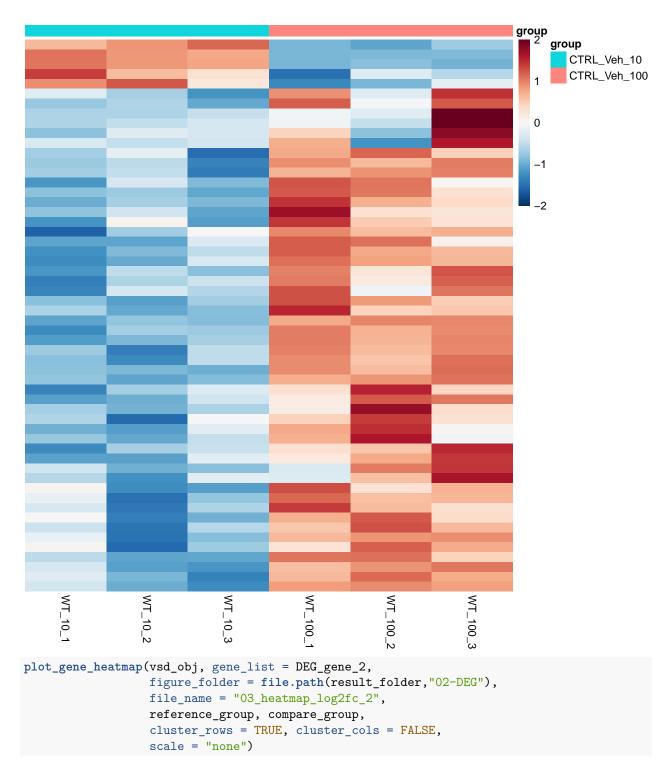
## [1] "Heatmap for 02\_heatmap\_log2fc\_1\_row "



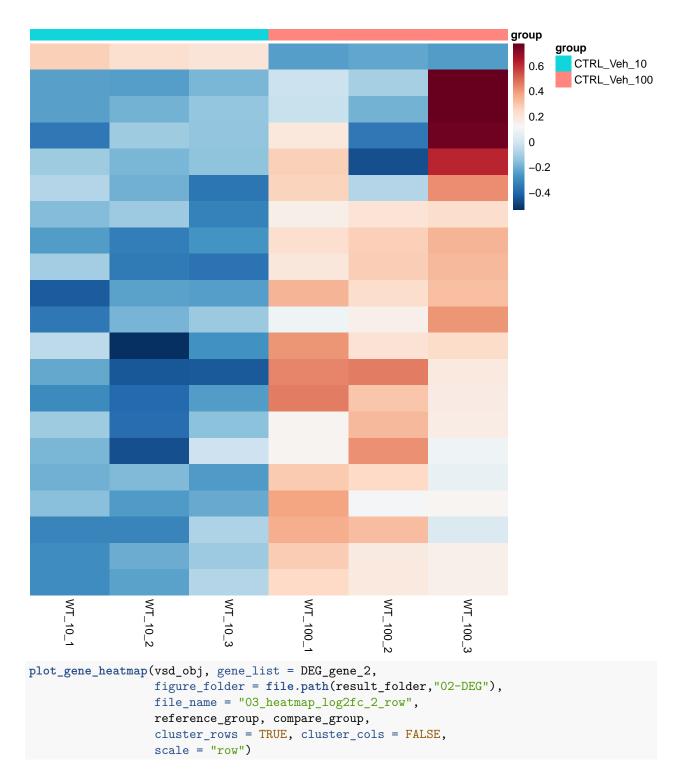
## [1] "Heatmap for 03\_heatmap\_log2fc\_1.5 "



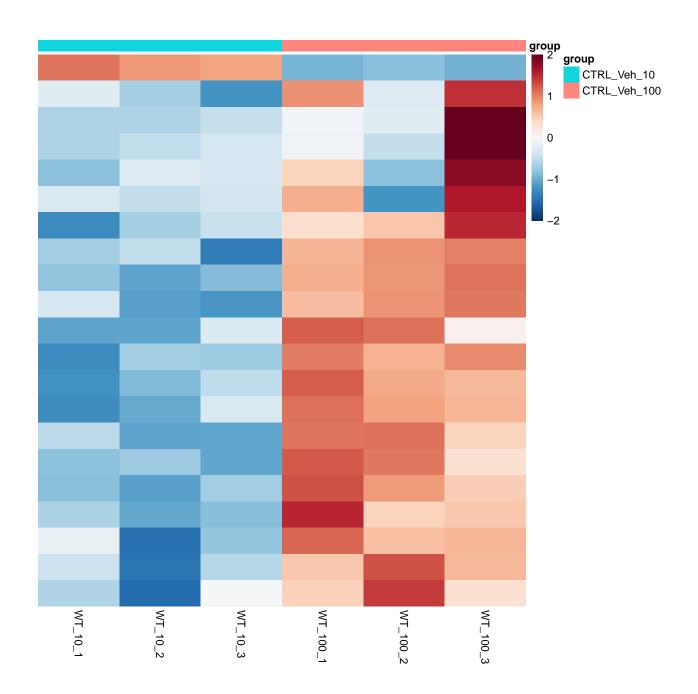
## [1] "Heatmap for O3\_heatmap\_log2fc\_1.5\_row "



## [1] "Heatmap for 03\_heatmap\_log2fc\_2 "



## [1] "Heatmap for 03\_heatmap\_log2fc\_2\_row "

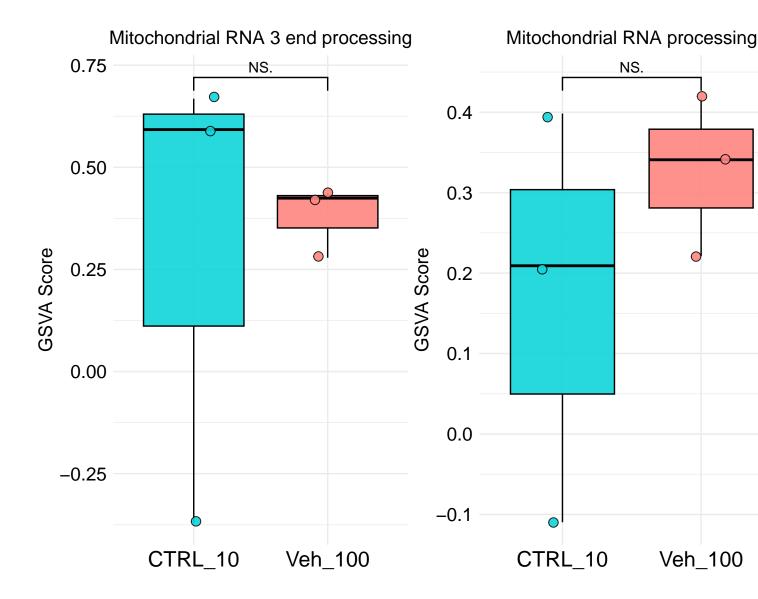


#### 4. GSVA analysis

```
# The following code is used to generate the GSVA matrix , only need to run once
gmxFile <- here::here("data", "ref", "c5.go.v2023.1.Hs.symbols.gmt")</pre>
go_list <- getGmt(gmxFile)</pre>
geneset <- go_list</pre>
dat <- as.matrix(counts)</pre>
gsvapar <- gsvaParam(dat, geneset, maxDiff=TRUE)</pre>
gsva_es <- gsva(gsvapar)</pre>
## No annotation package name available in the input data object.
## Attempting to directly match identifiers in data to gene sets.
## Estimating GSVA scores for 10531 gene sets.
## Estimating ECDFs with Gaussian kernels
##
gsva_matrix <- as.data.frame(gsva_es)</pre>
# save the result
write.csv(gsva_matrix, file.path(result_folder,"04-GSVA", "01_GSVA_matrix.csv"))
gsva_matrix <- read.csv(file.path(result_folder, "04-GSVA", "01_GSVA_matrix.csv"),</pre>
                         row.names = 1)
colnames(gsva_matrix) <- sub("^X", "", colnames(gsva_matrix))</pre>
condition_list_label <- condition_list %>%
 filter(group %in% c(reference_group, compare_group)) %>%
 mutate(group = case when(
    group == reference_group ~ reference_group_short,
    group == compare_group ~ compare_group_short,
    TRUE ~ group # fallback just in case
  ))
# plot the heatmap for the GSVA result
pathway_list <- read.csv(here::here("data", "ref", "focus-pathway_2024_10_03.csv"))</pre>
# # plot for all pathway
# for (i in 1:nrow(pathway list)){
   if (i %% 10 == 0) print(i)
   pathway name <- pathway list$pathway[i]</pre>
#
    plot_gsva_boxplot(gsva_matrix,
                       condition_list_label =condition_list_label,
#
                       pathway_name = pathway_name,
#
                       figure_folder = file.path(result_folder, "04-GSVA", "Boxplot"),
#
                       file_name = pasteO("GSVA_", pathway_name),
#
                       fig.height = 6, fig.width = 4,
#
                       reference_group =reference_group_short ,
                       compare_group = compare_group_short)
```

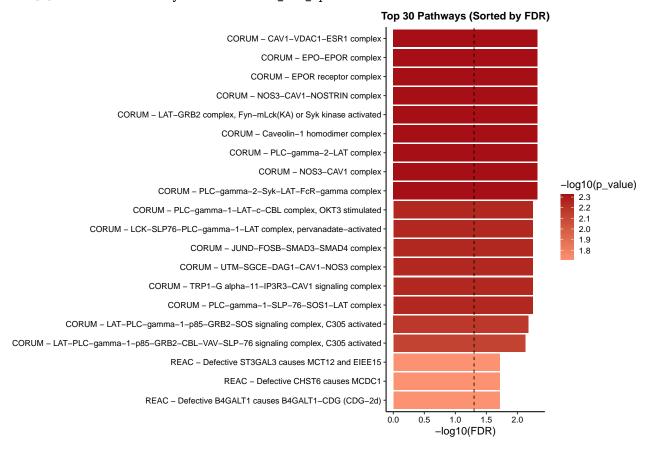
```
# }
box_plot_folder<- file.path(result_folder,"04-GSVA","Boxplot")</pre>
# create the folder
dir.create(box_plot_folder, showWarnings = FALSE)
# plot for the focus pathway
for (i in 1:2){
  pathway_name <- pathway_list$pathway[i]</pre>
  print(pathway_name)
  p<-plot_gsva_boxplot(gsva_matrix,</pre>
                     condition_list_label =condition_list_label,
                     pathway_name = pathway_name,
                     figure_folder = file.path(result_folder, "04-GSVA", "Boxplot"),
                     file_name = paste0("GSVA_", pathway_name),
                     fig.height = 6, fig.width = 4,
                     reference_group =reference_group_short ,
                     compare_group = compare_group_short)
  print(p)
```

- ## [1] "GOBP\_MITOCHONDRIAL\_RNA\_3\_END\_PROCESSING"
- ## [1] "GOBP\_MITOCHONDRIAL\_RNA\_PROCESSING"

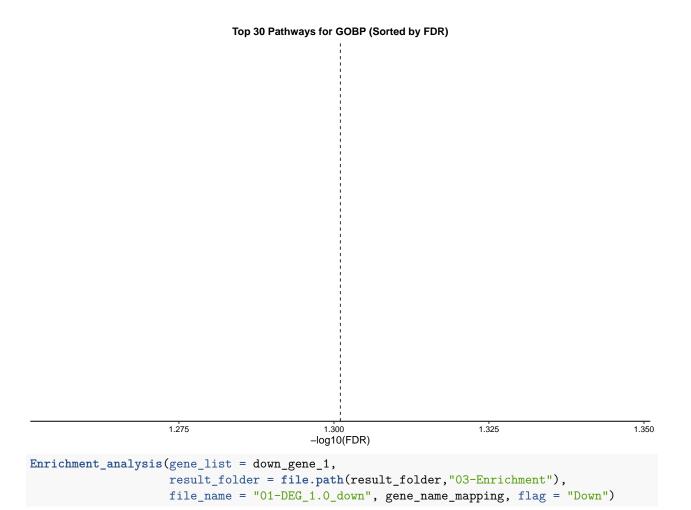


## 5. Pathway Enrichment Analysis

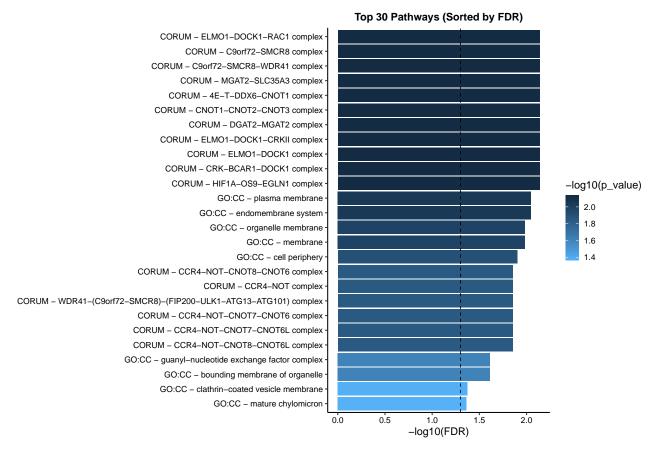
## [1] "Enrichment analysis for 01-DEG\_1.0\_up "



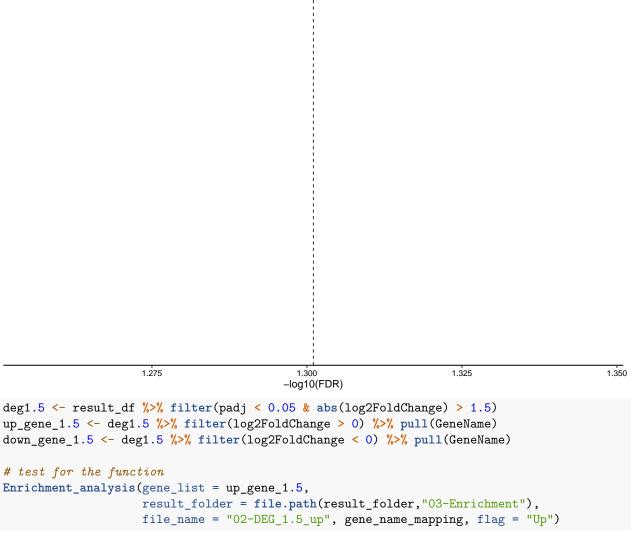
## [1] "Enrichment analysis for GOBP 01-DEG\_1.0\_up "



## [1] "Enrichment analysis for 01-DEG\_1.0\_down "

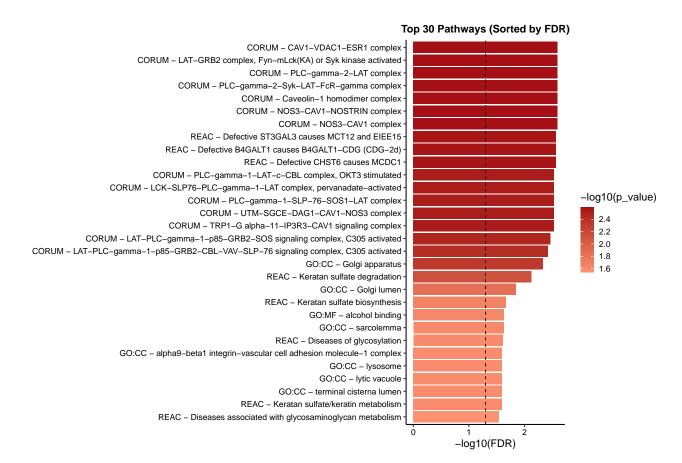


## [1] "Enrichment analysis for GOBP 01-DEG\_1.0\_down "

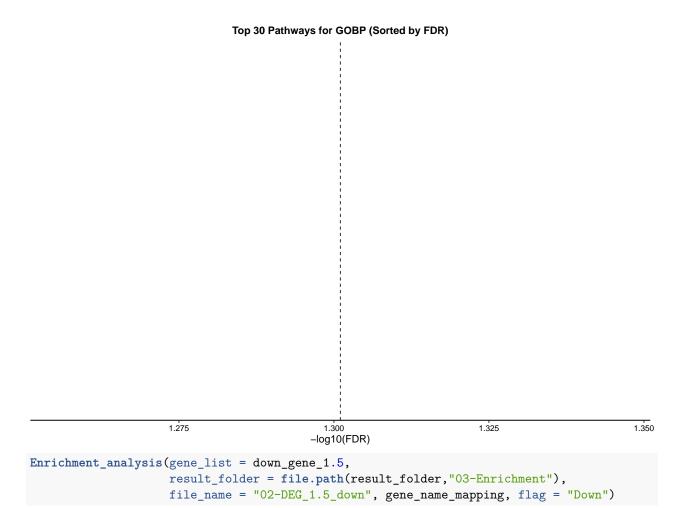


Top 30 Pathways for GOBP (Sorted by FDR)

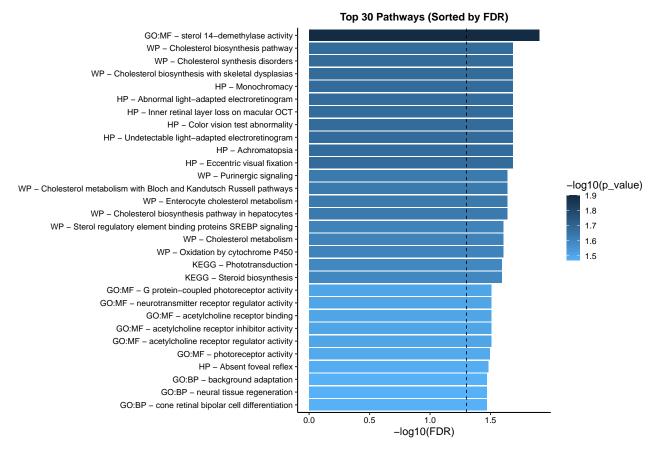
## [1] "Enrichment analysis for 02-DEG\_1.5\_up "



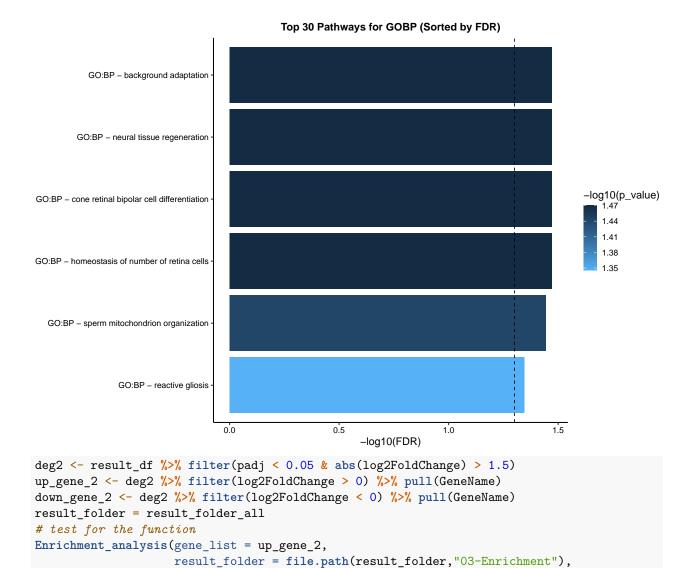
## [1] "Enrichment analysis for GOBP 02-DEG\_1.5\_up "  $\,$ 



## [1] "Enrichment analysis for 02-DEG\_1.5\_down "

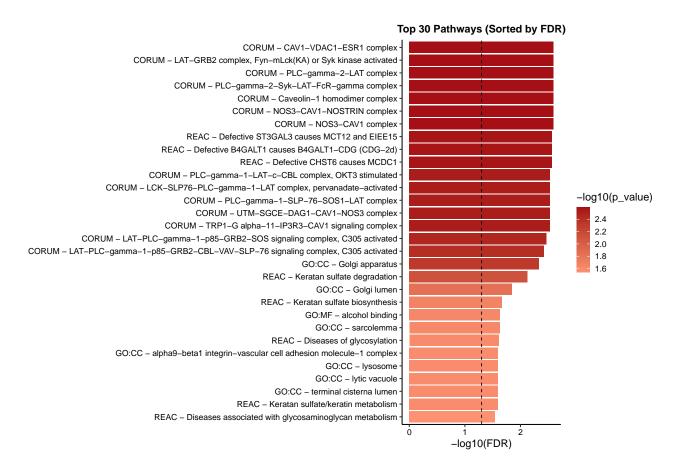


## [1] "Enrichment analysis for GOBP 02-DEG\_1.5\_down "

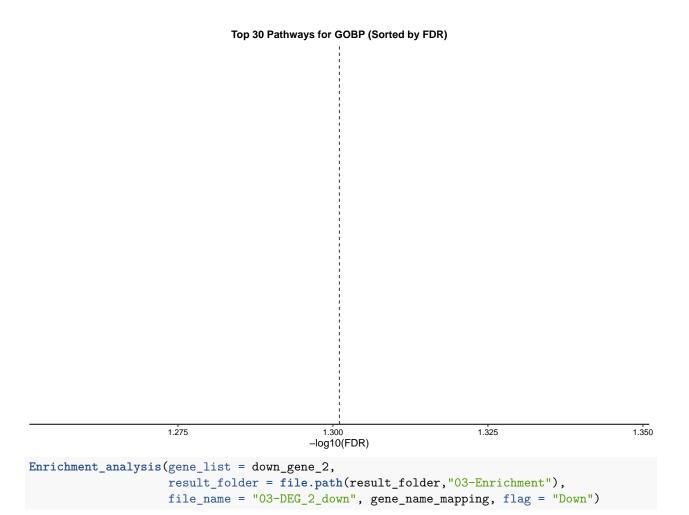


file\_name = "03-DEG\_2\_up", gene\_name\_mapping, flag = "Up")

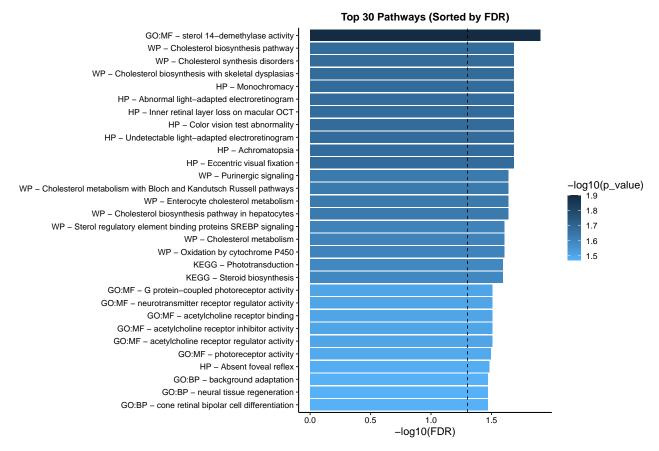
## [1] "Enrichment analysis for 03-DEG\_2\_up "



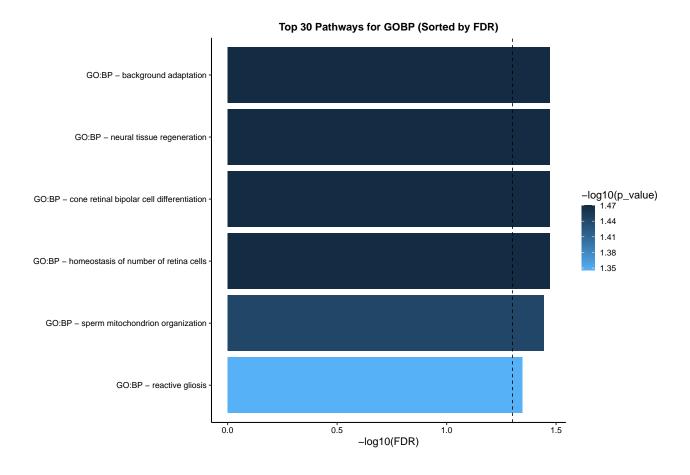
## [1] "Enrichment analysis for GOBP 03-DEG\_2\_up "



## [1] "Enrichment analysis for 03-DEG\_2\_down "



## [1] "Enrichment analysis for GOBP 03-DEG\_2\_down "



#### Session information

```
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;
##
## 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] parallel stats4
                                     graphics grDevices utils
                                                                    datasets
## [8] methods
                 base
##
## other attached packages:
## [1] GSEABase_1.66.0
                                    graph_1.82.0
## [3] annotate_1.82.0
                                    XML_3.99-0.18
## [5] extrafont_0.19
                                    ggsignif_0.6.4
## [7] patchwork_1.3.0
                                    decoupleR_2.10.0
## [9] GSVA_1.52.3
                                    BiocParallel_1.38.0
## [11] edgeR_4.2.2
                                    limma_3.60.6
## [13] GenomicFeatures_1.56.0
                                    biomaRt_2.60.1
## [15] gprofiler2_0.2.3
                                    RColorBrewer_1.1-3
## [17] data.table_1.16.4
                                    org.Hs.eg.db_3.19.1
## [19] AnnotationDbi_1.66.0
                                    clusterProfiler_4.12.6
## [21] ggfortify_0.4.17
                                    pheatmap_1.0.12
## [23] EnhancedVolcano_1.22.0
                                    ggrepel_0.9.6
## [25] apeglm_1.26.1
                                    DESeq2_1.44.0
## [27] SummarizedExperiment_1.34.0 Biobase_2.64.0
## [29] MatrixGenerics_1.16.0
                                    matrixStats_1.5.0
## [31] reshape2_1.4.4
                                    Matrix_1.7-2
## [33] Signac_1.14.0
                                    Seurat_5.2.1
## [35] SeuratObject_5.0.2
                                    sp_2.2-0
## [37] rtracklayer_1.64.0
                                    GenomicRanges_1.56.2
## [39] GenomeInfoDb_1.40.1
                                    IRanges_2.38.1
## [41] S4Vectors_0.42.1
                                    BiocGenerics_0.50.0
## [43] knitr_1.49
                                    lubridate_1.9.4
## [45] forcats_1.0.0
                                    stringr_1.5.1
## [47] dplyr_1.1.4
                                    purrr_1.0.4
## [49] readr_2.1.5
                                    tidyr_1.3.1
## [51] tibble_3.2.1
                                    ggplot2_3.5.1
## [53] tidyverse_2.0.0
```

LAPACK v

R.methodsS3\_1.8.2 goftest\_1.2-3

## loaded via a namespace (and not attached):

[1] SpatialExperiment\_1.14.0

[3] progress\_1.2.3

##

```
[5] HDF5Array_1.32.1
##
                                     Biostrings_2.72.1
##
     [7] vctrs_0.6.5
                                     spatstat.random_3.3-2
##
     [9] digest_0.6.37
                                     png 0.1-8
   [11] deldir_2.0-4
                                     parallelly_1.42.0
##
##
    [13] magick_2.8.5
                                     MASS_7.3-64
##
   [15] httpuv 1.6.15
                                     qvalue 2.36.0
   [17] withr 3.0.2
                                     xfun 0.51
##
   [19] ggfun_0.1.8
                                     survival_3.8-3
##
    [21] memoise_2.0.1
                                     gson_0.1.0
##
   [23] systemfonts_1.2.1
                                     ragg_1.3.3
   [25] tidytree_0.4.6
                                     zoo_1.8-12
##
   [27] pbapply_1.7-2
                                     R.oo_1.27.0
##
   [29] prettyunits_1.2.0
                                     KEGGREST_1.44.1
  [31] promises_1.3.2
##
                                     httr_1.4.7
##
  [33] restfulr_0.0.15
                                     rhdf5filters_1.16.0
##
   [35] globals_0.16.3
                                     fitdistrplus_1.2-2
##
   [37] rhdf5_2.48.0
                                     rstudioapi_0.17.1
   [39] UCSC.utils 1.0.0
                                     miniUI 0.1.1.1
##
   [41] generics_0.1.3
                                     DOSE_3.30.5
##
   [43] curl_6.2.1
                                     zlibbioc_1.50.0
##
   [45] ScaledMatrix_1.12.0
                                     ggraph_2.2.1
                                     GenomeInfoDbData_1.2.12
  [47] polyclip_1.10-7
##
  [49] SparseArray 1.4.8
                                     xtable_1.8-4
##
   [51] evaluate 1.0.3
                                     S4Arrays_1.4.1
##
  [53] BiocFileCache 2.12.0
                                     hms 1.1.3
  [55] irlba_2.3.5.1
                                     colorspace_2.1-1
##
  [57] filelock_1.0.3
                                     ROCR_1.0-11
##
  [59] reticulate_1.40.0
                                     spatstat.data_3.1-4
##
  [61] magrittr_2.0.3
                                     lmtest_0.9-40
## [63] later_1.4.1
                                     viridis_0.6.5
##
   [65] ggtree_3.12.0
                                     lattice_0.22-6
##
   [67] spatstat.geom_3.3-5
                                     future.apply_1.11.3
##
   [69] scattermore_1.2
                                     shadowtext_0.1.4
##
   [71] cowplot_1.1.3
                                     RcppAnnoy_0.0.22
##
    [73] pillar_1.10.1
                                     nlme_3.1-167
##
  [75] compiler_4.4.0
                                     beachmat_2.20.0
  [77] RSpectra 0.16-2
                                     stringi 1.8.4
##
                                     GenomicAlignments_1.40.0
  [79] tensor_1.5
##
   [81] plyr_1.8.9
                                     crayon_1.5.3
##
  [83] abind_1.4-8
                                     BiocIO_1.14.0
## [85] gridGraphics_0.5-1
                                     emdbook 1.3.13
## [87] locfit_1.5-9.11
                                     graphlayouts_1.2.2
## [89] bit_4.5.0.1
                                     fastmatch 1.1-6
## [91] textshaping_1.0.0
                                     codetools_0.2-20
## [93] BiocSingular_1.20.0
                                     plotly_4.10.4
## [95] mime_0.12
                                     splines_4.4.0
##
   [97] Rcpp_1.0.14
                                     fastDummies_1.7.5
  [99] sparseMatrixStats_1.16.0
                                     dbplyr_2.5.0
## [101] Rttf2pt1_1.3.12
                                     blob_1.2.4
## [103] here_1.0.1
                                     fs_1.6.5
## [105] listenv_0.9.1
                                     ggplotify_0.1.2
## [107] statmod 1.5.0
                                     tzdb_0.4.0
## [109] tweenr_2.0.3
                                     pkgconfig_2.0.3
## [111] tools 4.4.0
                                     cachem 1.1.0
```

```
## [113] RSQLite 2.3.9
                                     viridisLite_0.4.2
## [115] DBI_1.2.3
                                     numDeriv_2016.8-1.1
## [117] fastmap 1.2.0
                                     rmarkdown 2.29
                                     grid_4.4.0
## [119] scales_1.3.0
                                     Rsamtools_2.20.0
## [121] ica_1.0-3
## [123] coda 0.19-4.1
                                     dotCall64 1.2
## [125] RANN 2.6.2
                                     farver 2.1.2
## [127] tidygraph_1.3.1
                                     scatterpie_0.2.4
## [129] yaml_2.3.10
                                     cli_3.6.4
## [131] lifecycle_1.0.4
                                     uwot_0.2.2
## [133] mvtnorm_1.3-3
                                     timechange_0.3.0
## [135] gtable_0.3.6
                                     rjson_0.2.23
## [137] ggridges_0.5.6
                                     progressr_0.15.1
## [139] ape_5.8-1
                                     jsonlite_1.9.0
## [141] RcppHNSW_0.6.0
                                     bitops_1.0-9
## [143] bit64_4.6.0-1
                                     Rtsne_0.17
## [145] yulab.utils_0.2.0
                                     spatstat.utils_3.1-2
## [147] bdsmatrix 1.3-7
                                     GOSemSim 2.30.2
## [149] spatstat.univar_3.1-1
                                     R.utils_2.12.3
## [151] lazyeval 0.2.2
                                     shiny 1.10.0
## [153] htmltools_0.5.8.1
                                     enrichplot_1.24.4
## [155] GO.db 3.19.1
                                     sctransform_0.4.1
## [157] rappdirs_0.3.3
                                     tinytex_0.55
## [159] glue_1.8.0
                                     spam 2.11-1
                                     XVector_0.44.0
## [161] httr2 1.1.0
## [163] RCurl_1.98-1.16
                                     rprojroot_2.0.4
## [165] treeio_1.28.0
                                     gridExtra_2.3
                                     igraph_2.1.4
## [167] extrafontdb_1.0
## [169] R6_2.6.1
                                     SingleCellExperiment_1.26.0
## [171] labeling_0.4.3
                                     RcppRoll_0.3.1
## [173] cluster_2.1.8
                                     bbmle_1.0.25.1
## [175] Rhdf5lib_1.26.0
                                     aplot_0.2.4
## [177] DelayedArray_0.30.1
                                     tidyselect_1.2.1
## [179] ggforce_0.4.2
                                     xml2_1.3.6
## [181] future 1.34.0
                                     rsvd 1.0.5
## [183] munsell_0.5.1
                                     KernSmooth_2.23-26
## [185] htmlwidgets 1.6.4
                                     fgsea 1.30.0
## [187] rlang_1.1.5
                                     spatstat.sparse_3.1-0
## [189] spatstat.explore_3.3-4
```