

Differential Expression Analysis for bulk RNA-seq data

Vehicle contition: 22q vs Control

Ximing Ran

2025-02-28

Contents

1. Read the count data	2
2. Differential expression analysis	2
3. Visualization for results	4
(1) Sample information	4
(2) DEG visualization - Volcano plot and Heatmap	6
(3) 22q gene visualization - Heatmap	16
4. Pathway Enrichment Analysis	18

Session information	26
----------------------------	-----------

```
library(tibble)
library(tidyr)
library(dplyr)
library(rtracklayer)

# load function from local files
source(here::here("source", "DEG_functions.R"))
```

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 results/02-DEG-Vehicle/synaptosomes_bulkRNA_counts_cleaned.csv

```
input_count <- read.csv(here::here("data", "synaptosomes_bulkRNA",
                                    "synaptosomes_bulkRNA_counts_cleaned.csv"))
counts <- as.data.frame(input_count) %>%
  column_to_rownames(var = "gene")
colnames(counts) <- gsub("_", "-", colnames(counts))

# sort the columns by the colname
condition_list <- data.frame(
  group = c(rep("CTRL_Vehicle", 10),
            rep("22q_Vehicle", 10),
            rep("CTRL_TTX", 10),
            rep("22q_TTX", 10))
)
row.names(condition_list) <- c(
  paste0("V", rep(1:10, each = 2), "-", 1:2),
  paste0("T", rep(1:10, each = 2), "-", 1:2)
)

counts <- counts[, rownames(condition_list)]

# load the 22q gene list
target_22q_gene <- read.csv(here::here("data", "ref", "22q_gene_2024_10_17.csv"))
target_gene <- target_22q_gene$gene
target_gene <- target_gene[1:(length(target_gene) - 4)]

gene_name_mapping <- readRDS(here::here("data", "ref", "gene_name_mapping.rds"))
```

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-Vehicle/DESeq2_results.csv.

```
# Init the result folder structure for the result
result_folder = file.path("results", "02-DEG-Vehicle")
Result_folder_structure(result_folder)

# load the comparison group information
reference_group <- "CTRL_Vehicle"
compare_group <- "22q_Vehicle"

filter_sample_info <- condition_list %>%
  filter(group %in% c(reference_group, compare_group))
filter_counts <- counts[, rownames(filter_sample_info)]

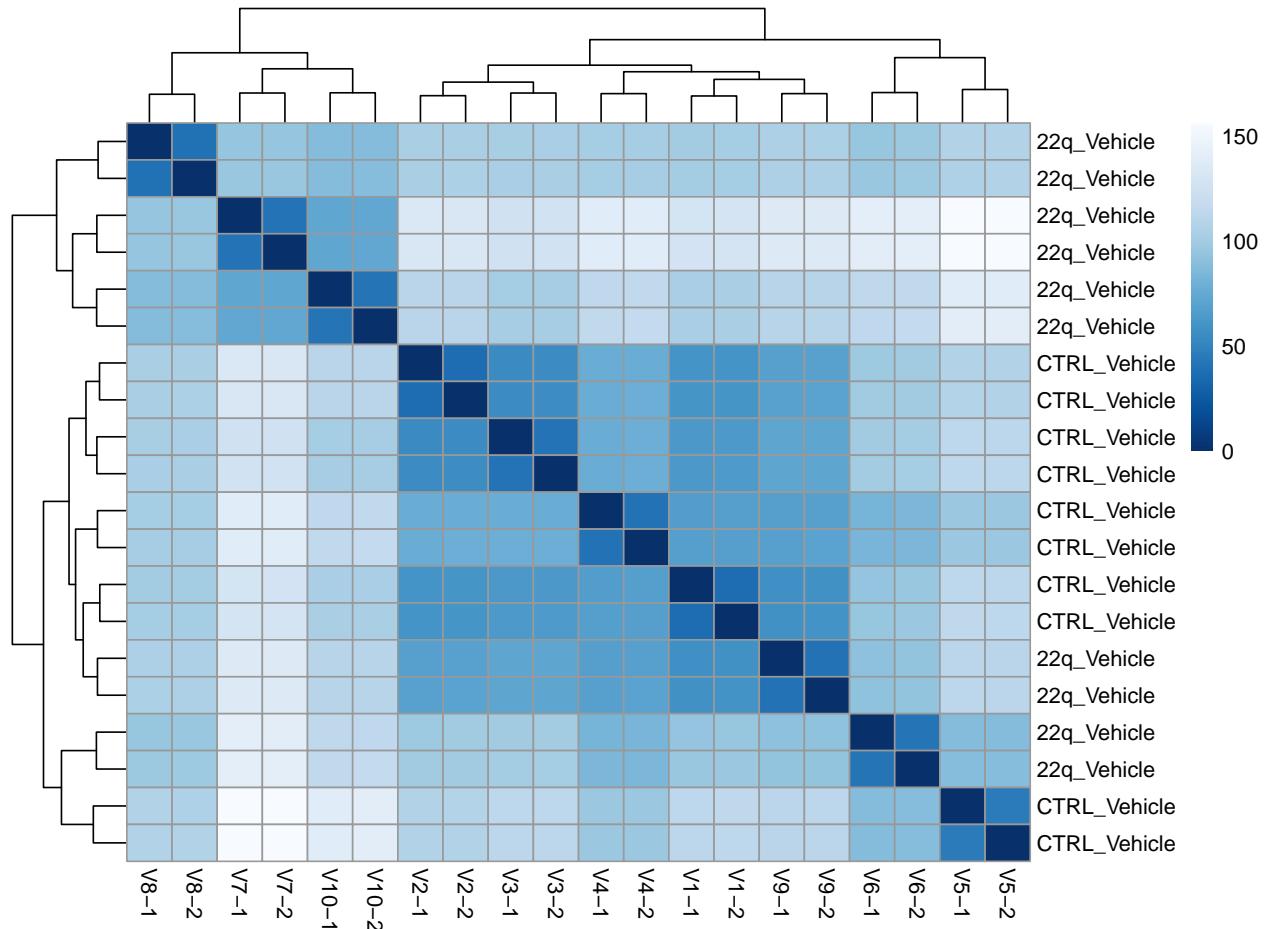
# Run the DESeq2 analysis
dds_obj <- DEAnalysis(counts = filter_counts,
                      reference_group = reference_group,
                      compare_group = compare_group,
```

```
    condition_list = filter_sample_info,  
    target_gene = target_gene,  
    result_folder = result_folder)  
  
## [1] "DEG analysis is done"
```

3. Visualization for result

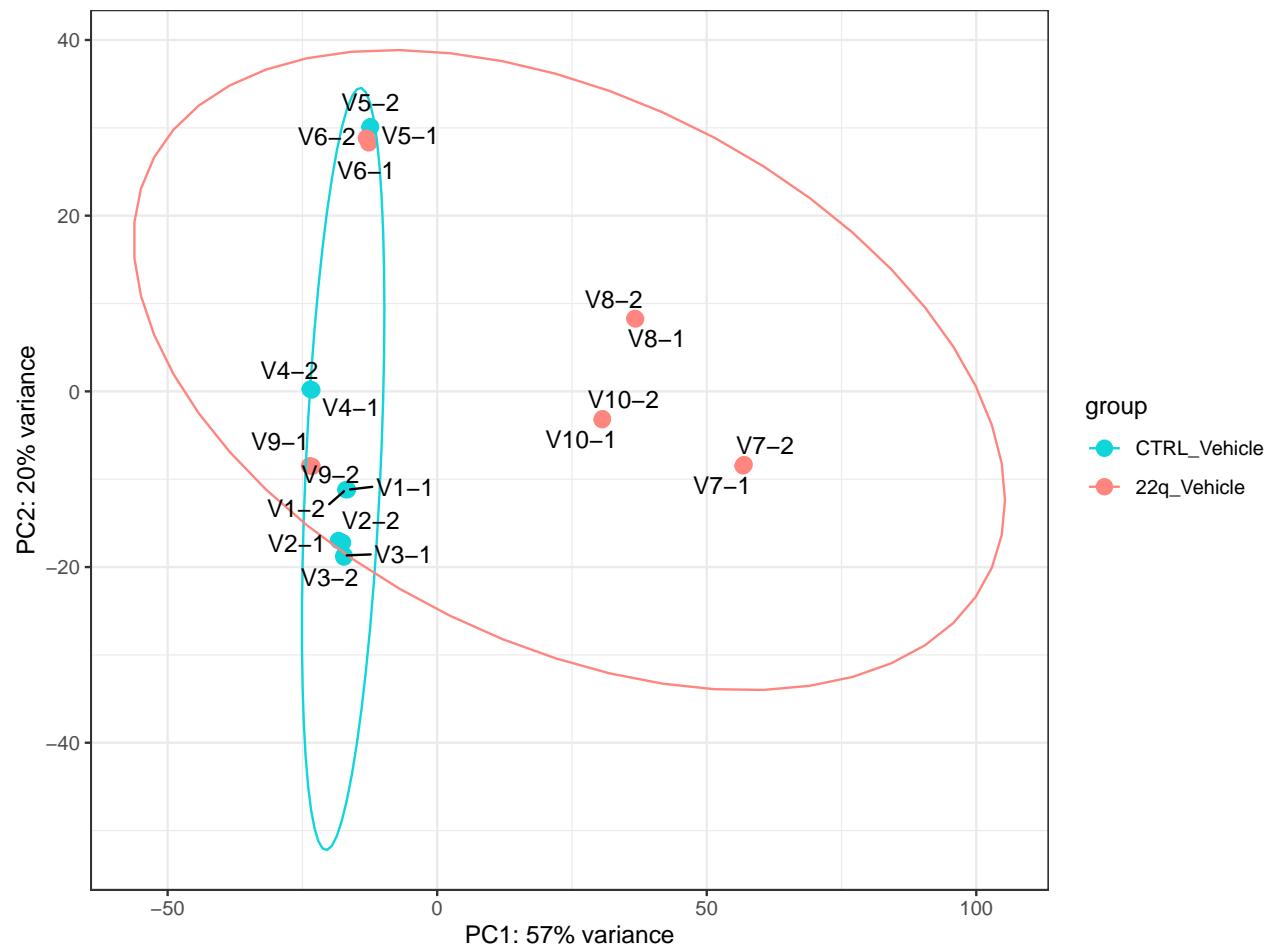
(1) Sample information

```
# Plot the sample distance
plot_sample_heatmap(dds_obj, figure_folder =
    file.path(result_folder, "01-Sample_info"),
    file_name = "01_sample_distance_heatmap")
```



```
## [1] "Sample distance heatmap is done"
# Plot the PCA plot for the sample
plot_sample_PCA_plot(dds_obj, figure_folder = file.path(result_folder, "01-Sample_info"),
    file_name = "02_sample_PCA_plot",
    reference_group, compare_group)

## [1] "PCA plot is done"
```

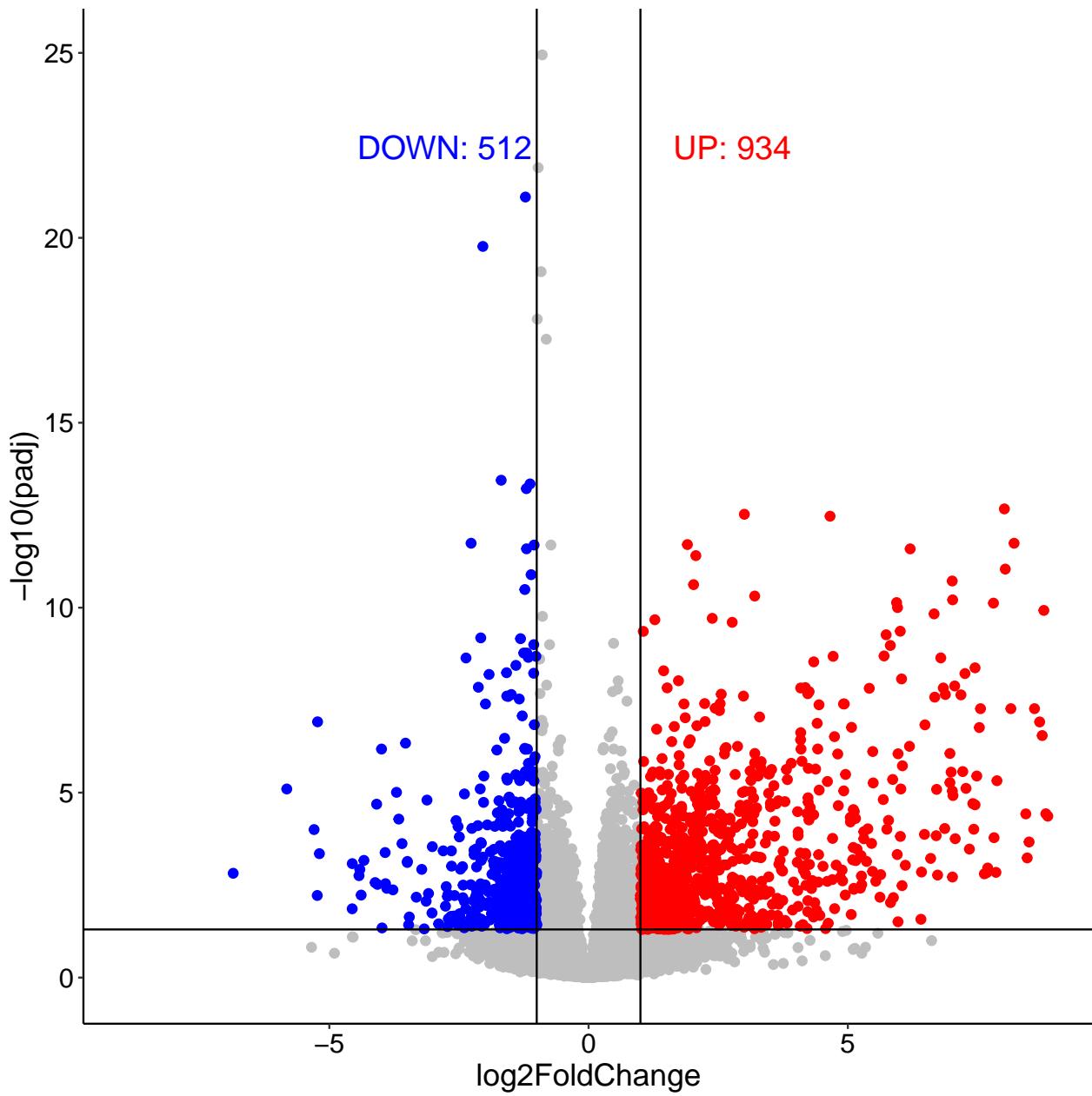


(2) DEG visualization - Volcano plot and Heatmap

```
result_df <- results(dds_obj) %>%
  as.data.frame() %>%
  rownames_to_column(var = "GeneName") %>%
  dplyr::select(GeneName, everything()) %>%
  filter(!is.na(padj)) %>% # Correct way to filter non-NA values
  arrange(padj)

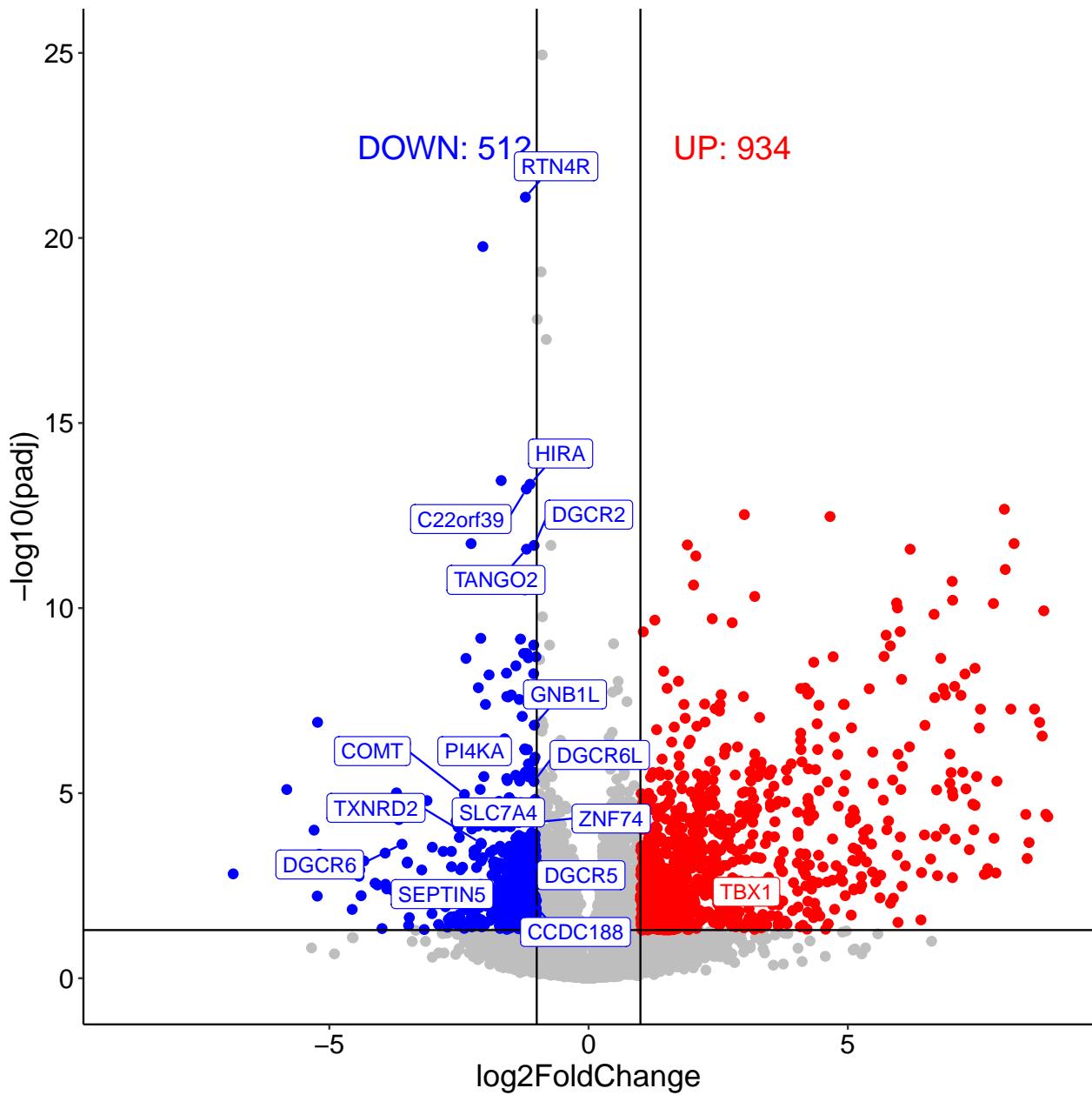
# Plot the volcano plot for the DEG
plot_volcano_plot(result_df,
  figure_folder = file.path(result_folder, "02-DEG"),
  file_name = "02_volcano_plot_log2fc_1",
  thread = 1, label_gene = NULL)

## [1] "Volcano plot for 02_volcano_plot_log2fc_1"
```



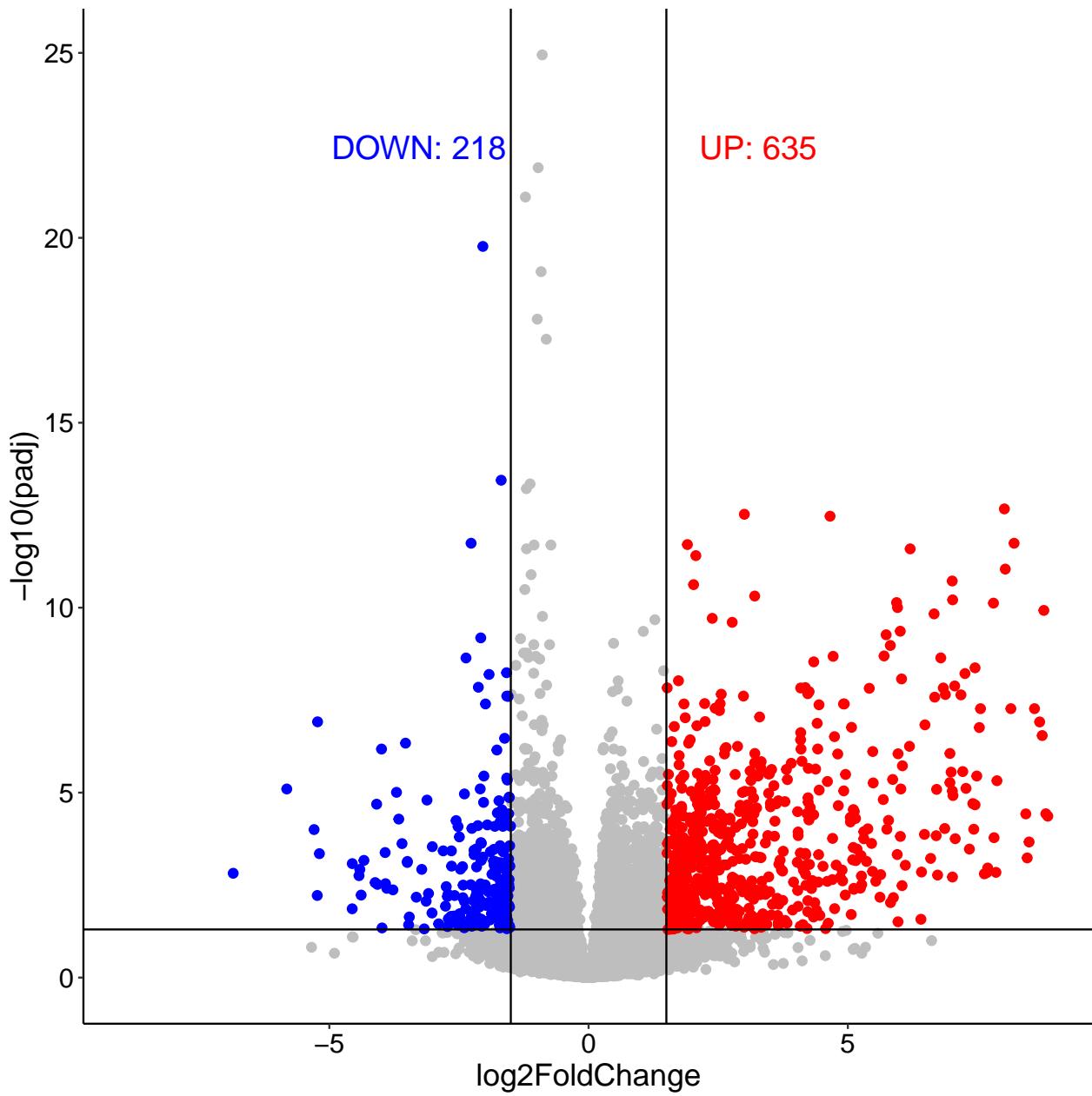
```
plot_volcano_plot(result_df,
                   figure_folder = file.path(result_folder, "02-DEG"),
                   file_name = "02_volcano_plot_log2fc_1_with_22q_gene",
                   thread = 1, dot_size = 2, label_gene = target_gene)
```

```
## [1] "Volcano plot for 02_volcano_plot_log2fc_1_with_22q_gene"
```



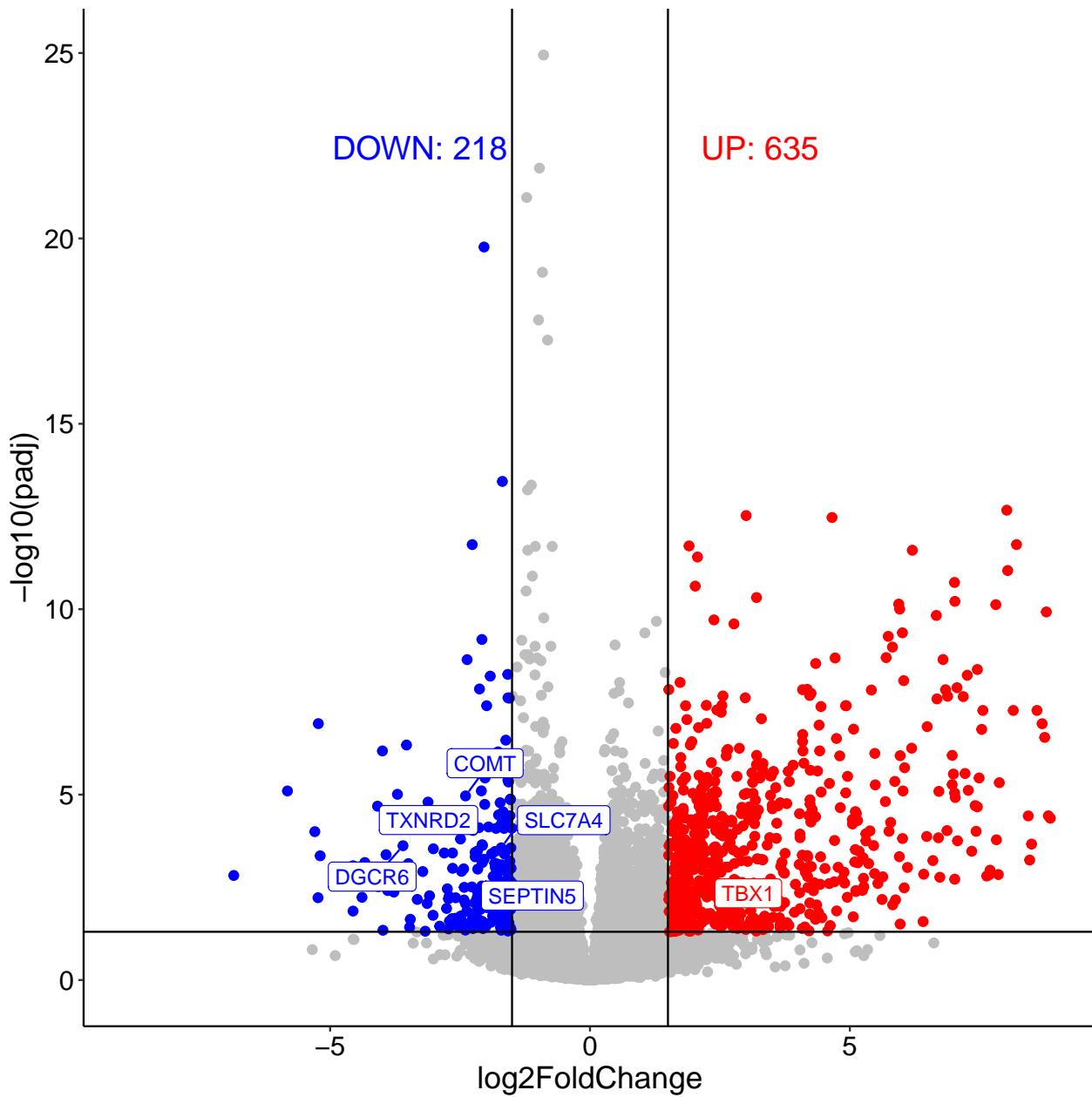
```
plot_volcano_plot(result_df,
                   figure_folder = file.path(result_folder, "02-DEG"),
                   file_name = "03_volcano_plot_log2fc_1.5",
                   thread = 1.5 , dot_size = 2,label_gene = NULL)
```

```
## [1] "Volcano plot for 03_volcano_plot_log2fc_1.5"
```



```
plot_volcano_plot(result_df,
                   figure_folder = file.path(result_folder, "02-DEG"),
                   file_name = "03_volcano_plot_log2fc_1.5_with_22q_gene",
                   thread = 1.5 , dot_size =2, label_gene = target_gene)
```

```
## [1] "Volcano plot for 03_volcano_plot_log2fc_1.5_with_22q_gene"
```

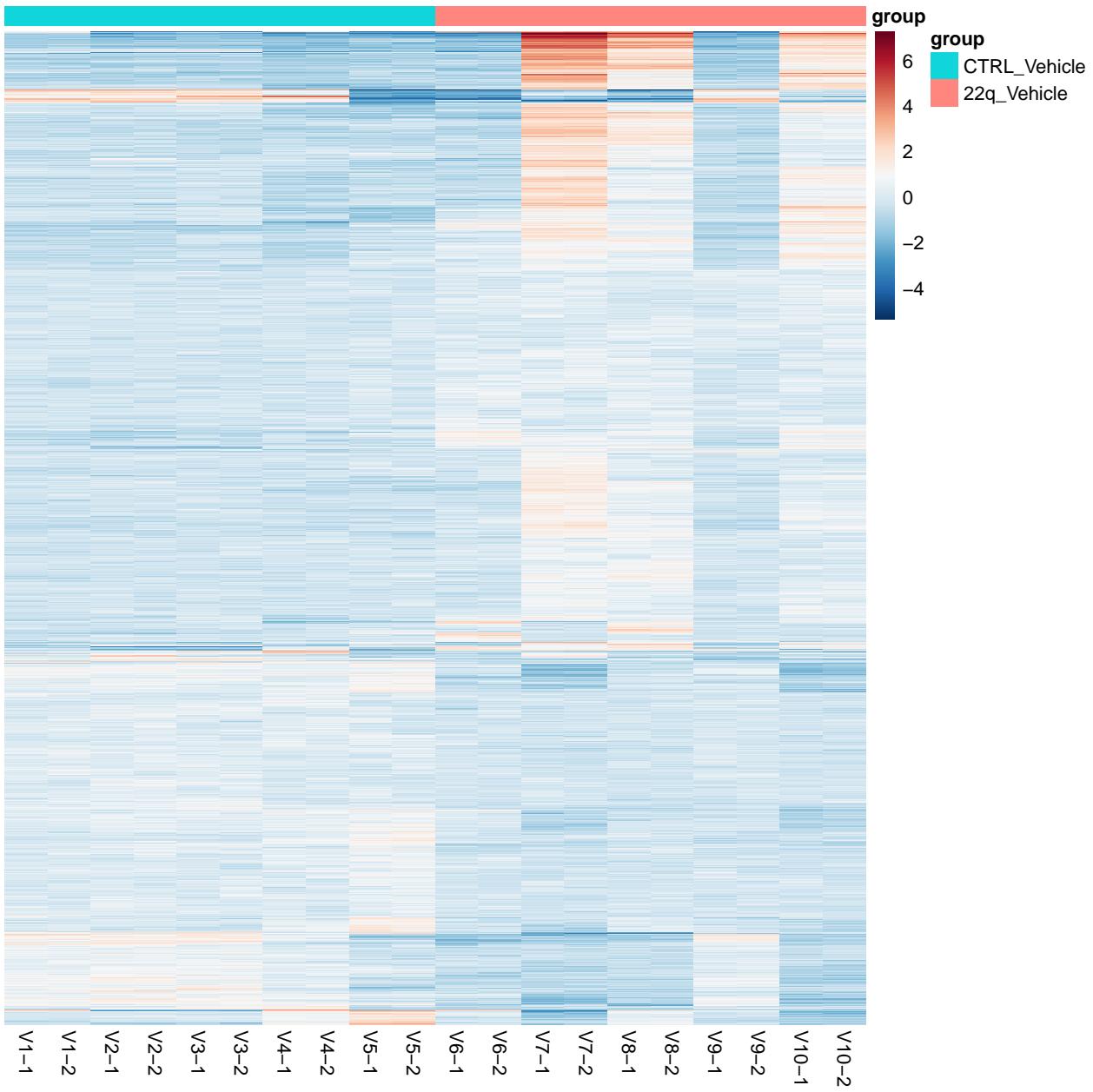


```
# Plot the heatmap for the DEG
vsd_obj <- varianceStabilizingTransformation(dds_obj, blind = TRUE)

DEG_gene_1 <- result_df %>% filter(abs(log2FoldChange) > 1) %>% pull(GeneName)
DEG_gene_1.5 <- result_df %>% filter(abs(log2FoldChange) > 1.5) %>% pull(GeneName)

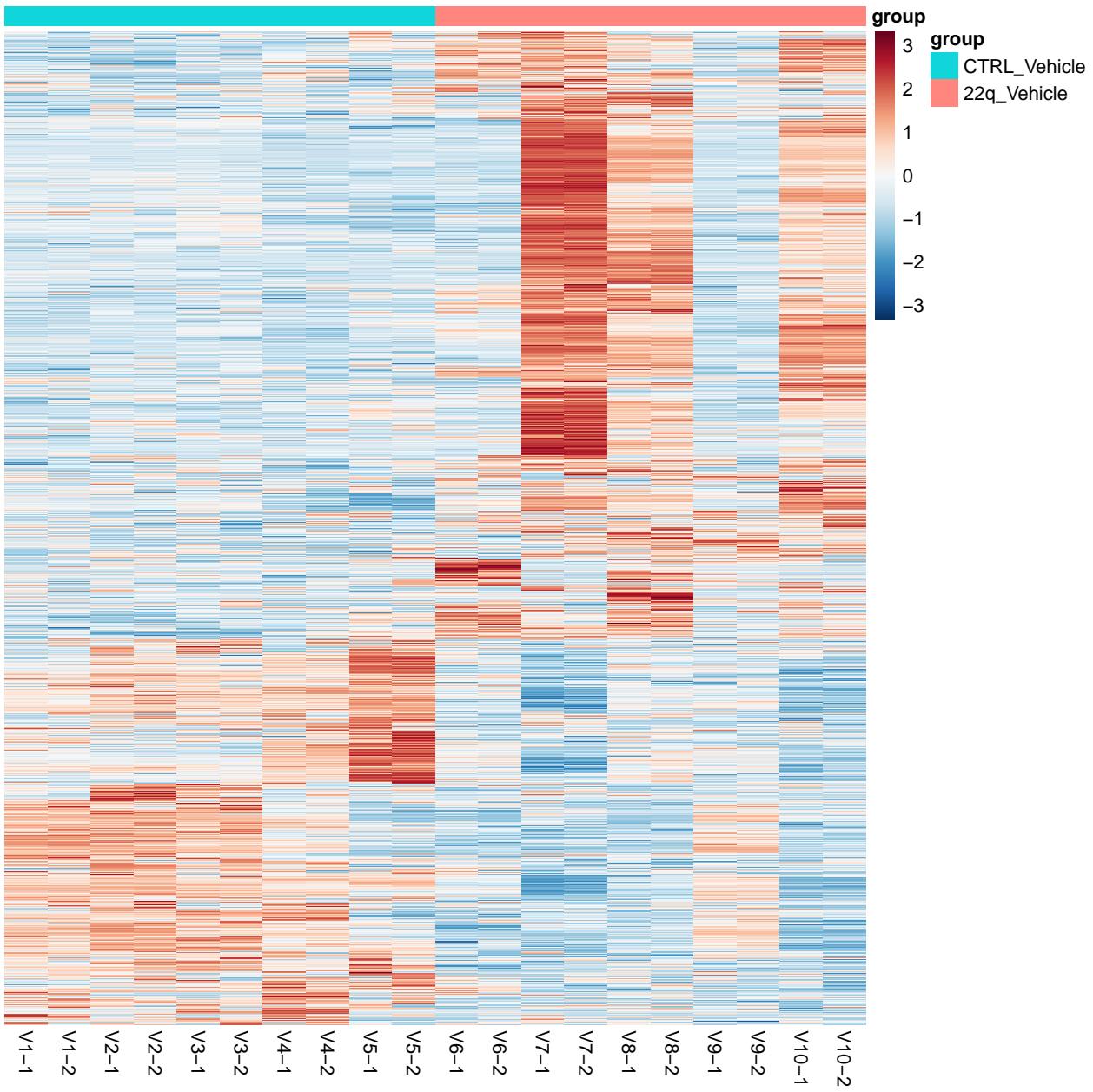
plot_gene_heatmap(vsd_obj, gene_list = DEG_gene_1,
                  figure_folder = file.path(result_folder, "02-DEG"),
                  file_name = "02_heatmap_log2fc_1",
                  reference_group, compare_group,
                  cluster_rows = TRUE, cluster_cols = FALSE,
                  scale = "none")

## [1] "Heatmap for 02_heatmap_log2fc_1 "
```



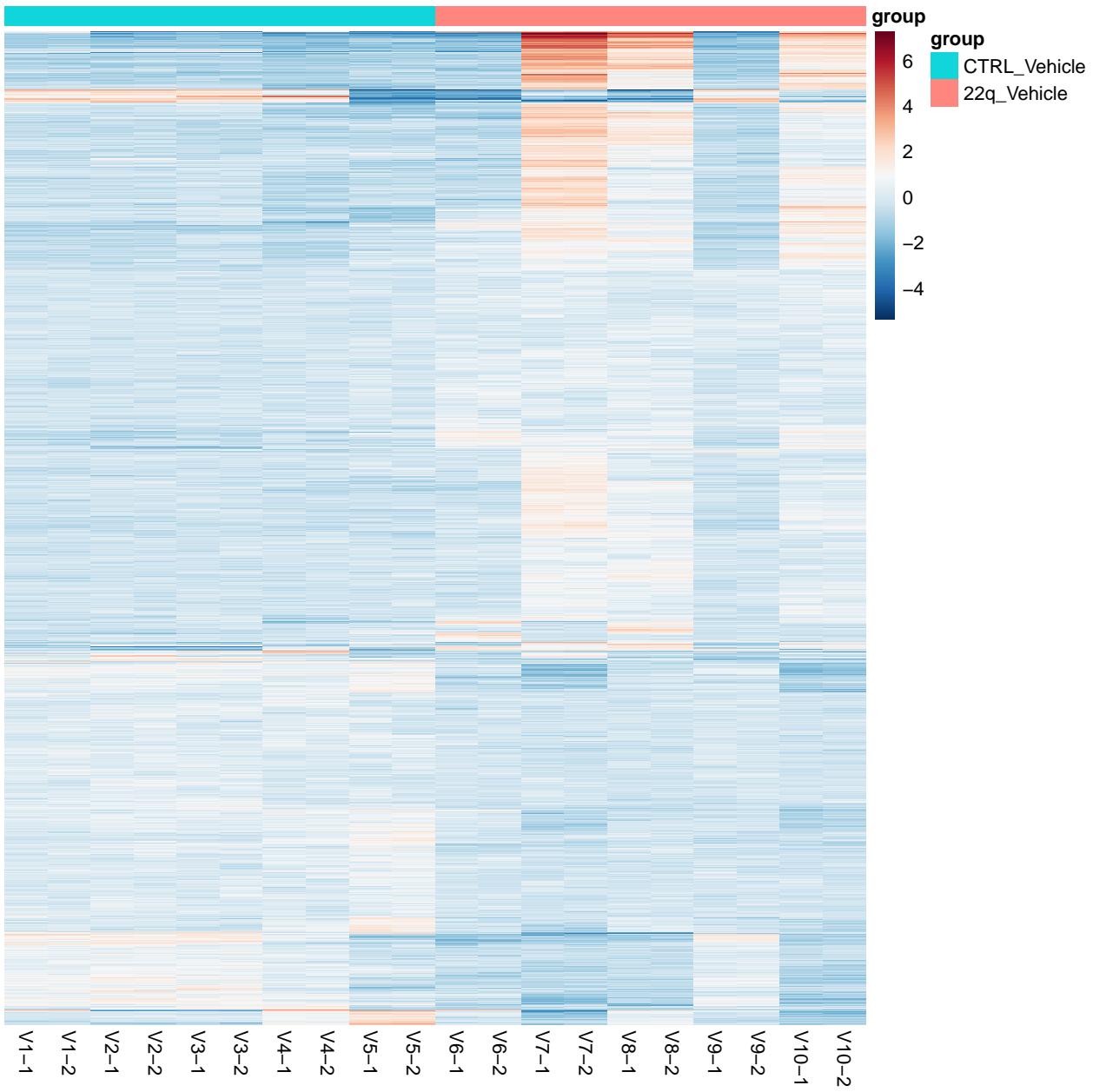
```
plot_gene_heatmap(vsd_obj, gene_list = DEG_gene_1,
                  figure_folder = file.path(result_folder, "02-DEG"),
                  file_name = "02_heatmap_log2fc_1_row",
                  reference_group, compare_group,
                  cluster_rows = TRUE, cluster_cols = FALSE,
                  scale = "row")
```

```
## [1] "Heatmap for 02_heatmap_log2fc_1_row "
```



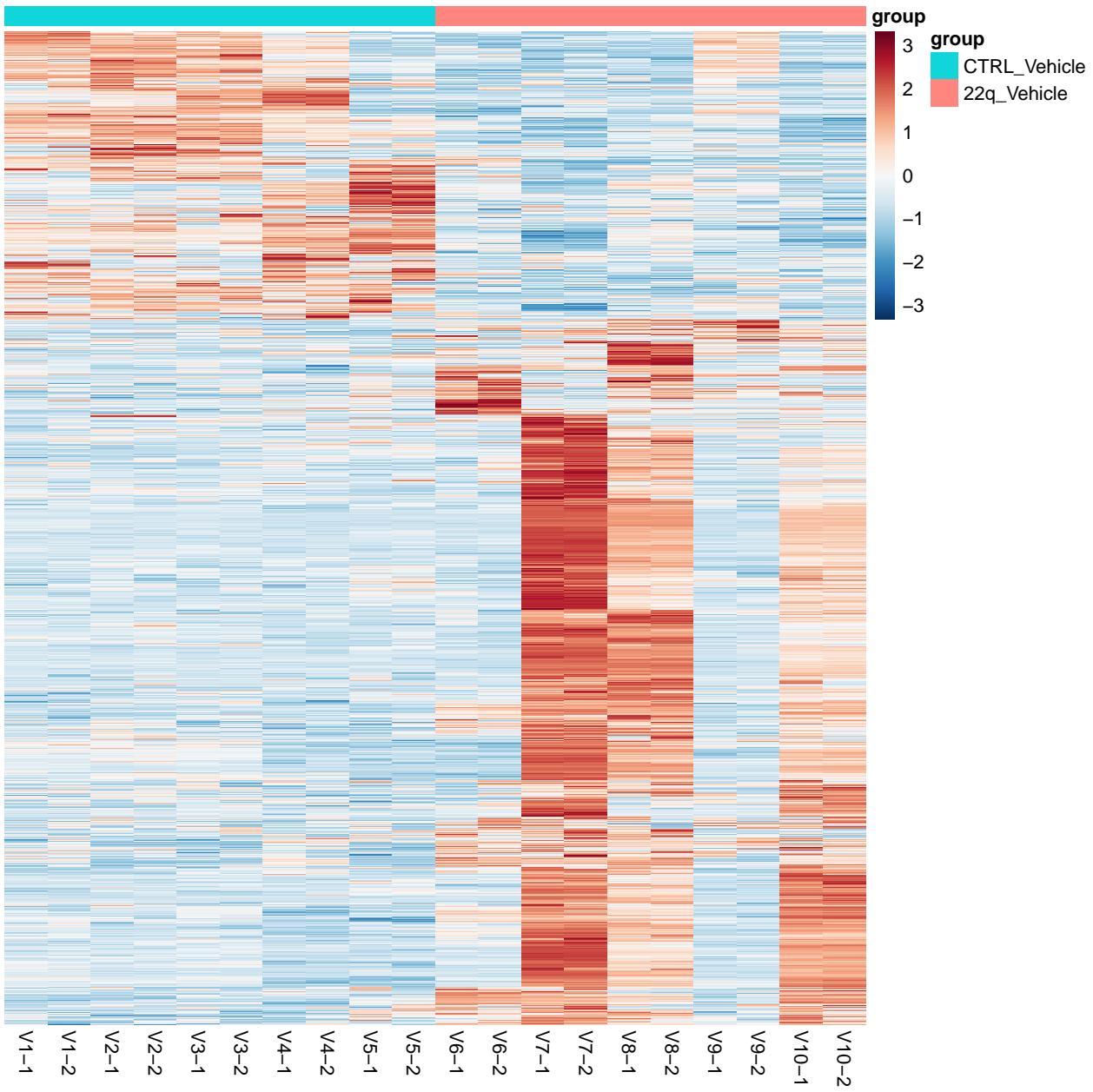
```
plot_gene_heatmap(vsd_obj, gene_list = DEG_gene_1,
                  figure_folder = file.path(result_folder, "02-DEG"),
                  file_name = "02_heatmap_log2fc_1",
                  reference_group, compare_group,
                  cluster_rows = TRUE, cluster_cols = FALSE,
                  scale = "none")
```

```
## [1] "Heatmap for 02_heatmap_log2fc_1 "
```



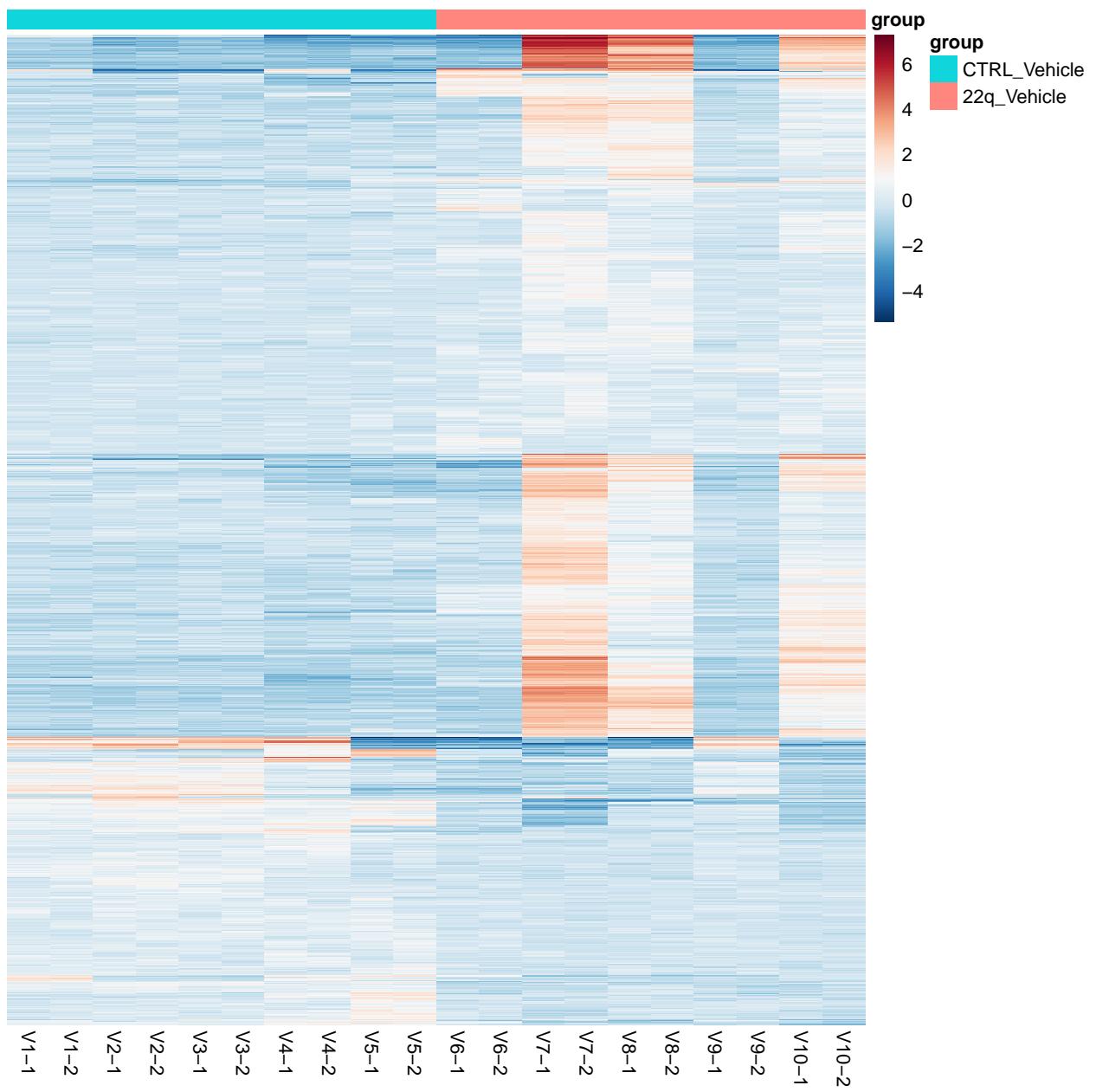
```
plot_gene_heatmap(vsd_obj, gene_list = DEG_gene_1.5,
                  figure_folder = file.path(result_folder, "02-DEG"),
                  file_name = "03_heatmap_log2fc_1.5_row",
                  reference_group, compare_group,
                  cluster_rows = TRUE, cluster_cols = FALSE,
                  scale = "row")
```

```
## [1] "Heatmap for 03_heatmap_log2fc_1.5_row "
```



```
plot_gene_heatmap(vsd_obj, gene_list = DEG_gene_1.5,
                  figure_folder = file.path(result_folder, "02-DEG"),
                  file_name = "03_heatmap_log2fc_1.5",
                  reference_group, compare_group,
                  cluster_rows = TRUE, cluster_cols = FALSE,
                  scale = "none")
```

```
## [1] "Heatmap for 03_heatmap_log2fc_1.5 "
```



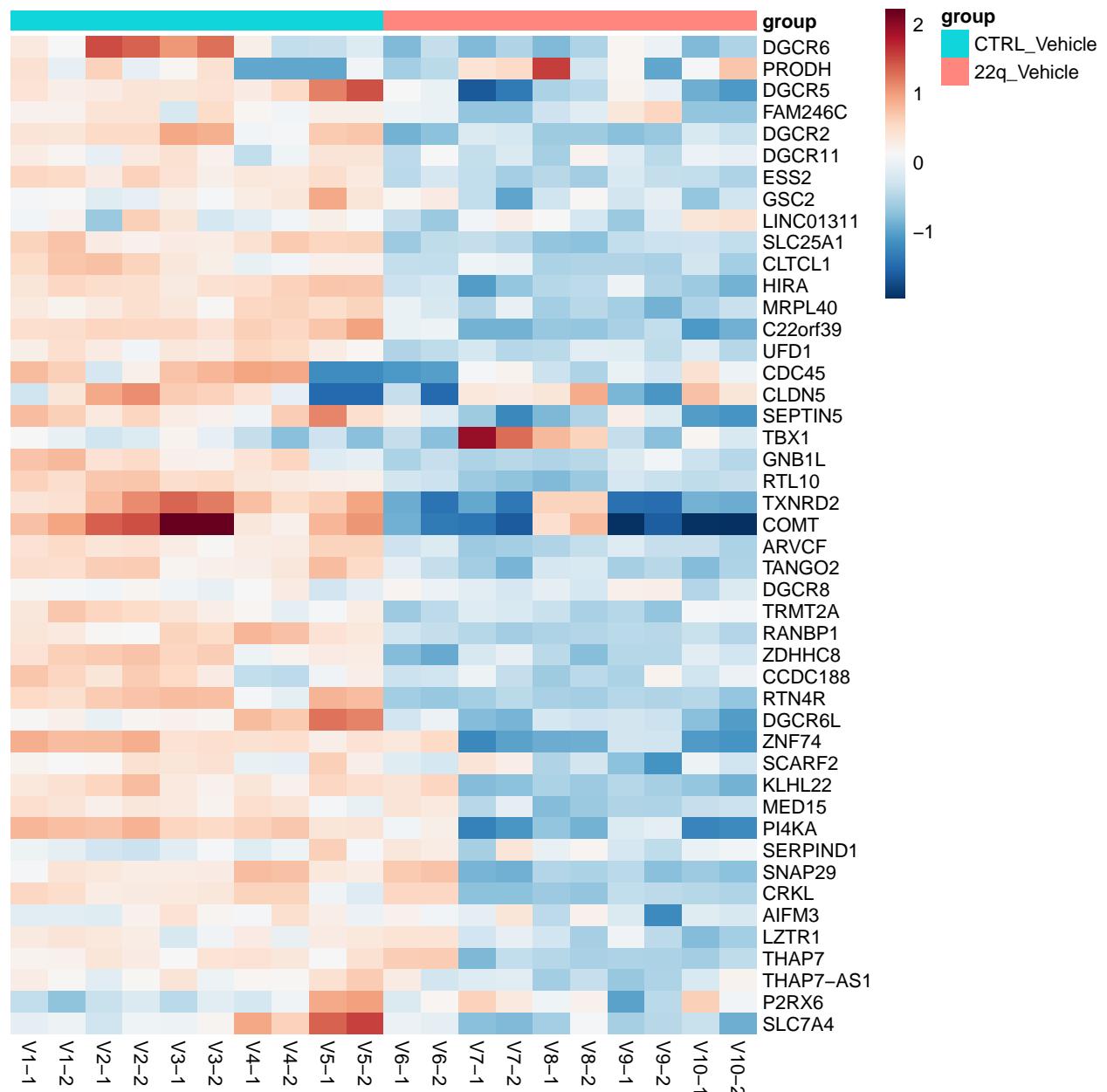
(3) 22q gene visualization - Heatmap

```
# Plot the heatmap for the 22q gene

target_gene <- intersect(target_gene, rownames(vsd_obj))

plot_gene_heatmap(vsd_obj, gene_list = target_gene ,
                  figure_folder = file.path(result_folder,"05-22q_Gene"),
                  file_name = "01_heatmap_22q_gene",
                  reference_group, compare_group,
                  cluster_rows = FALSE, cluster_cols = FALSE,
                  scale = "none", show_rownames =TRUE)
```

```
## [1] "Heatmap for 01_heatmap_22q_gene "
```

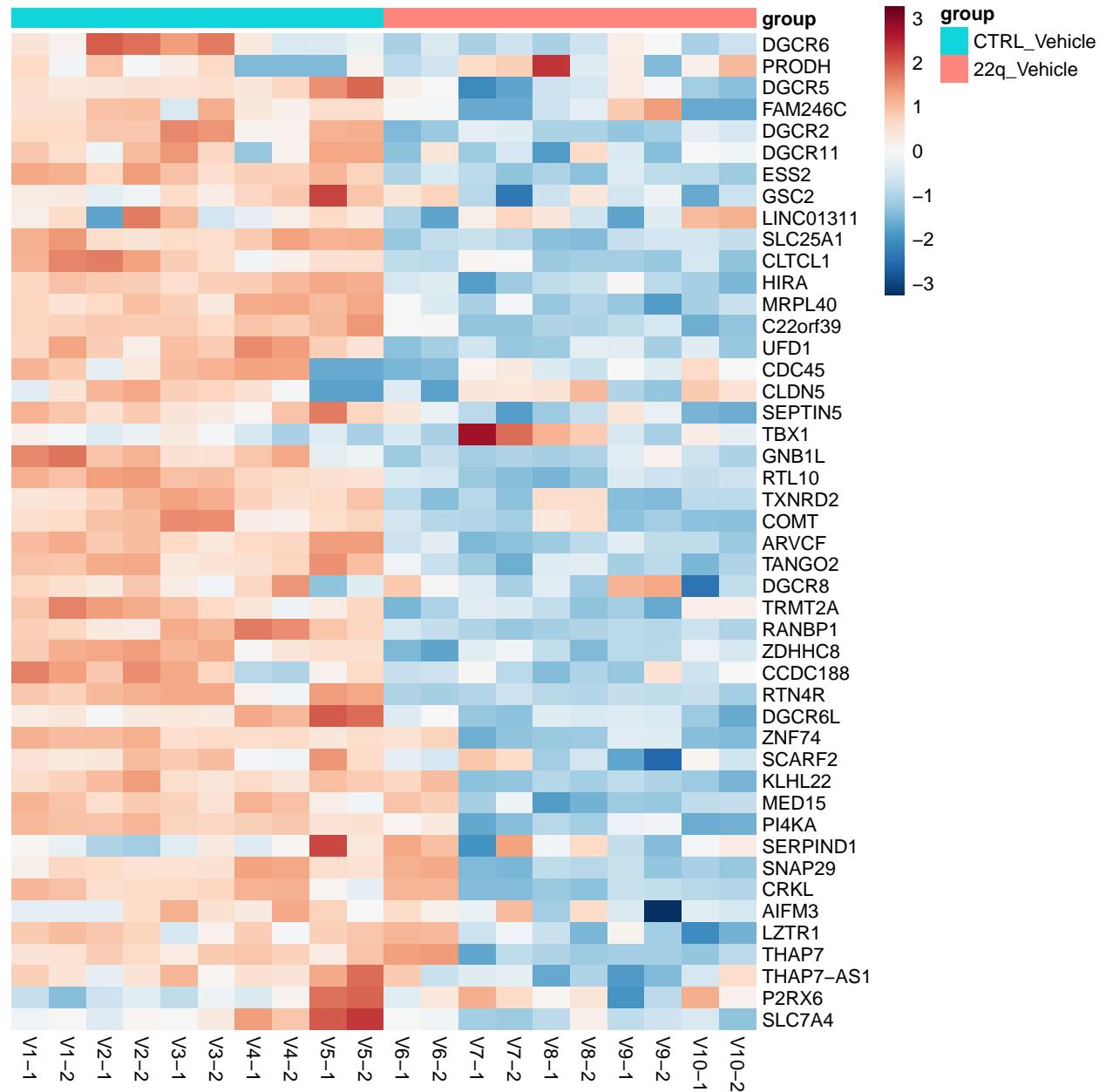


```

plot_gene_heatmap(vsd_obj, gene_list = target_gene ,
                  figure_folder = file.path(result_folder,"05-22q_Gene"),
                  file_name = "01_heatmap_22q_gene_row",
                  reference_group, compare_group,
                  cluster_rows = FALSE, cluster_cols = FALSE,
                  scale = "row", show_rownames =TRUE)

```

[1] "Heatmap for 01_heatmap_22q_gene_row "



4. Pathway Enrichment Analysis

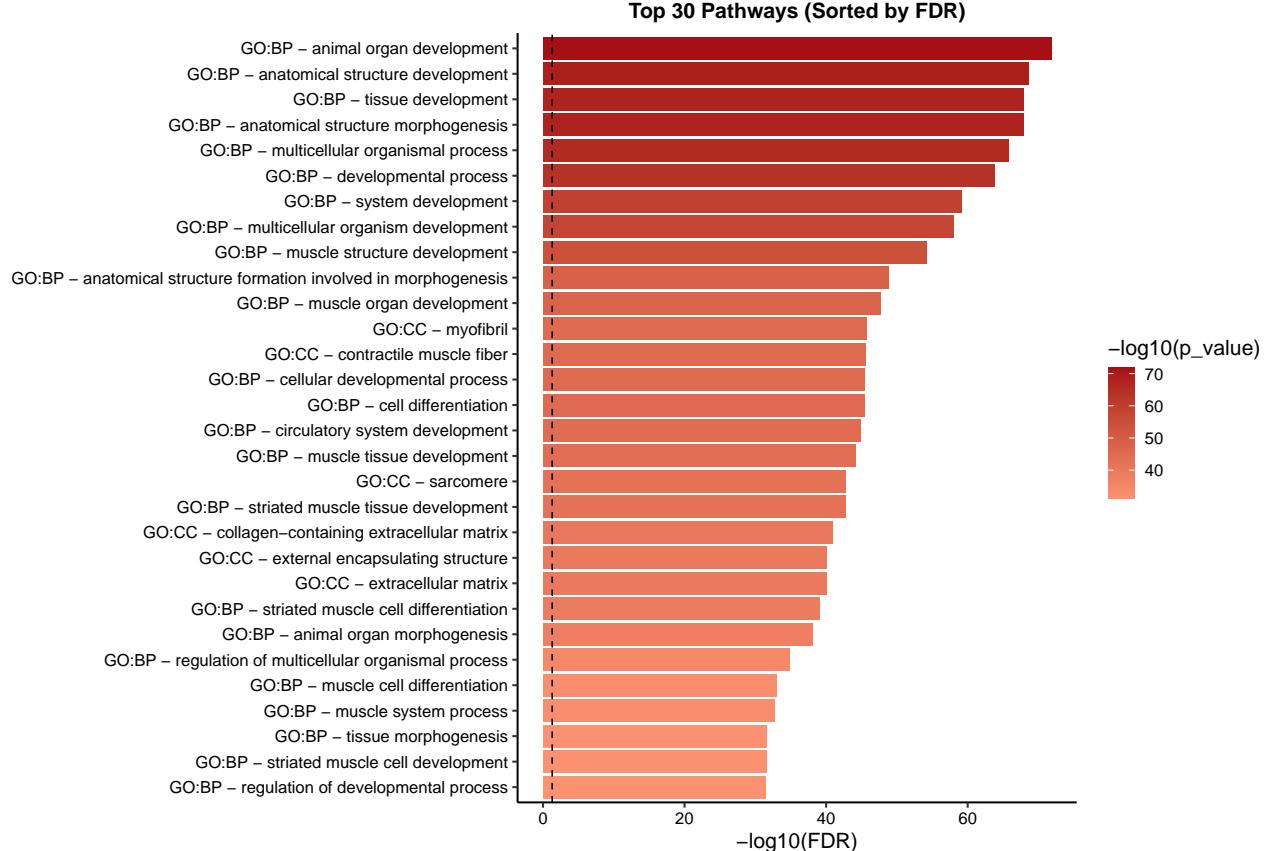
```

deg1 <- result_df %>% filter(padj < 0.05 & abs(log2FoldChange) > 1)
up_gene_1 <- deg1 %>% filter(log2FoldChange > 0) %>% pull(GeneName)
down_gene_1 <- deg1 %>% filter(log2FoldChange < 0) %>% pull(GeneName)

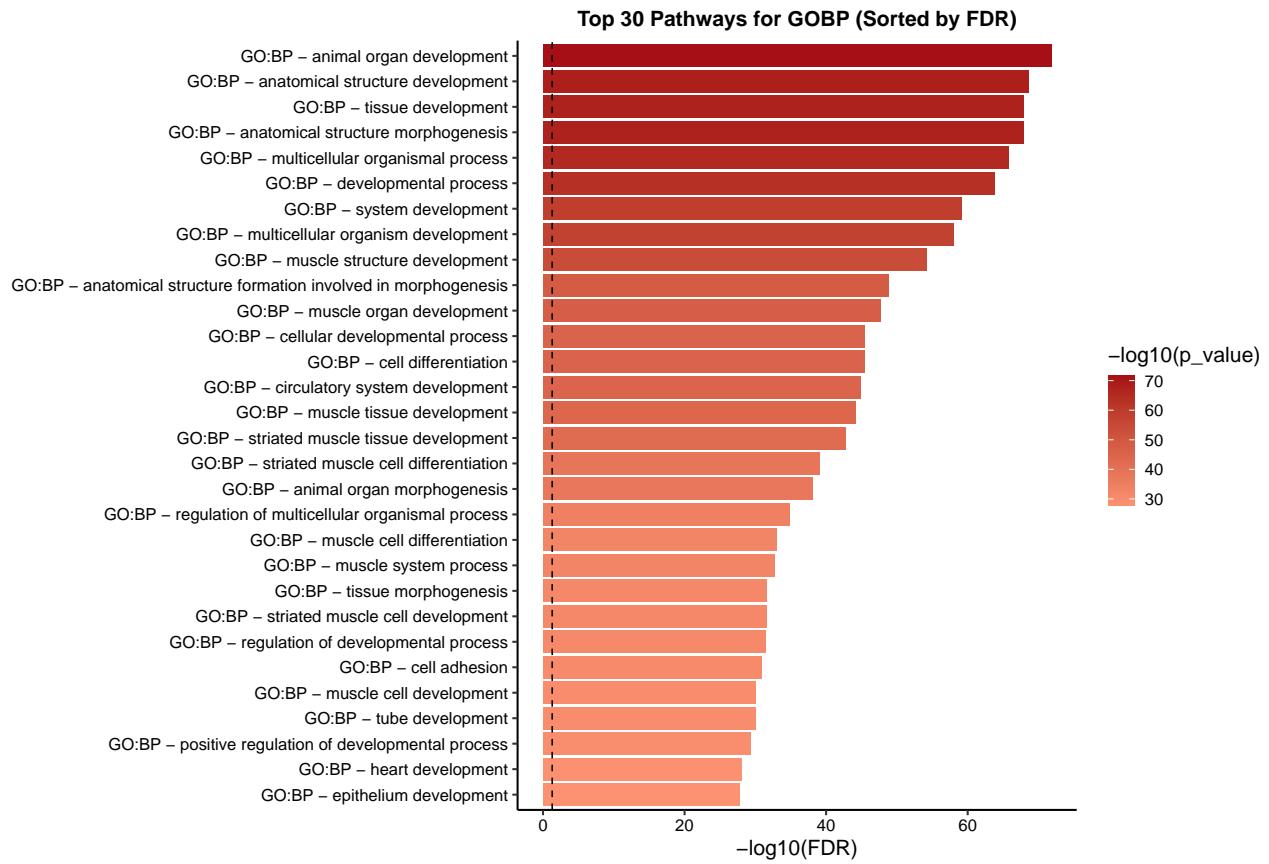
# test for the function
Enrichment_analysis(gene_list = up_gene_1,
                     result_folder = file.path(result_folder, "03-Enrichment"),
                     file_name = "01-DEG_1.0_up", gene_name_mapping, flag = "Up")

```

[1] "Enrichment analysis for 01-DEG_1.0_up "

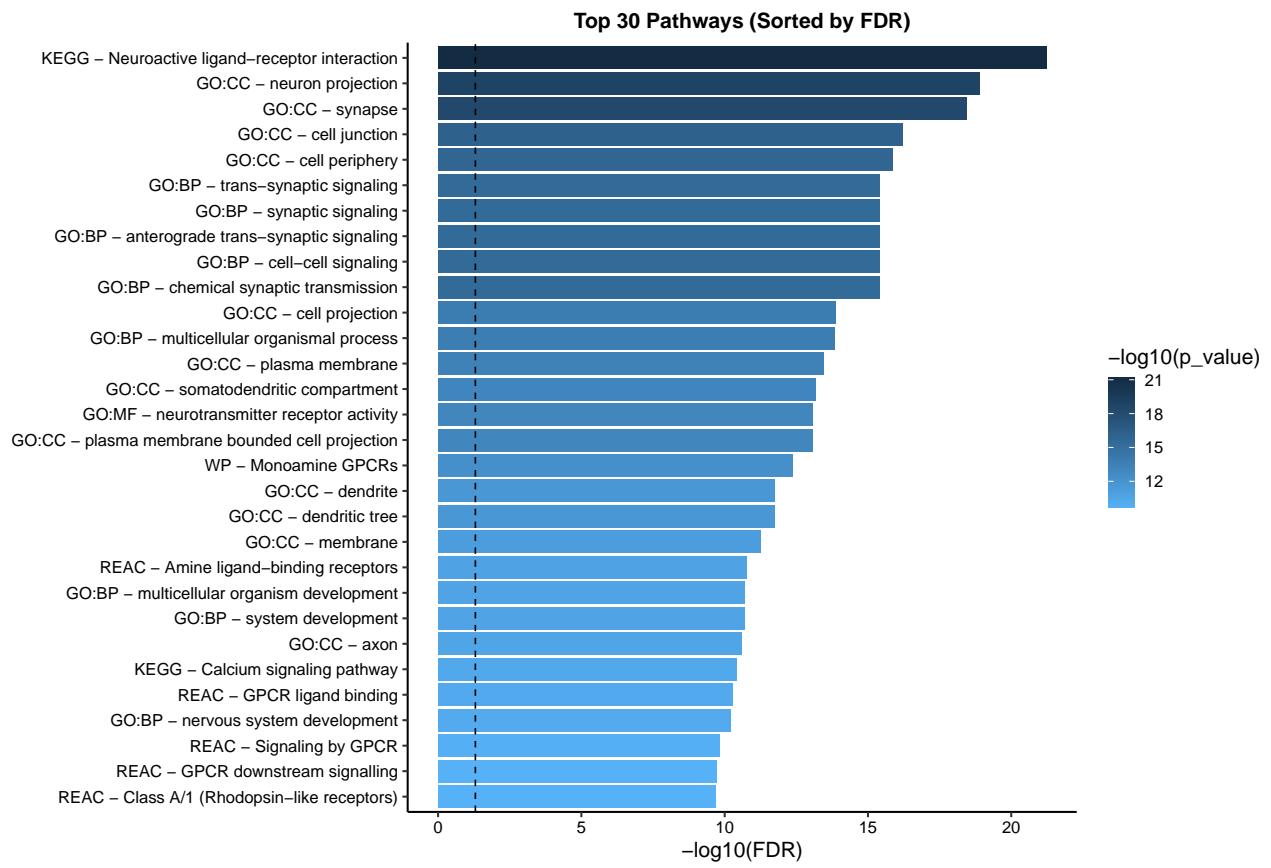


[1] "Enrichment analysis for GOBP 01-DEG_1.0_up "

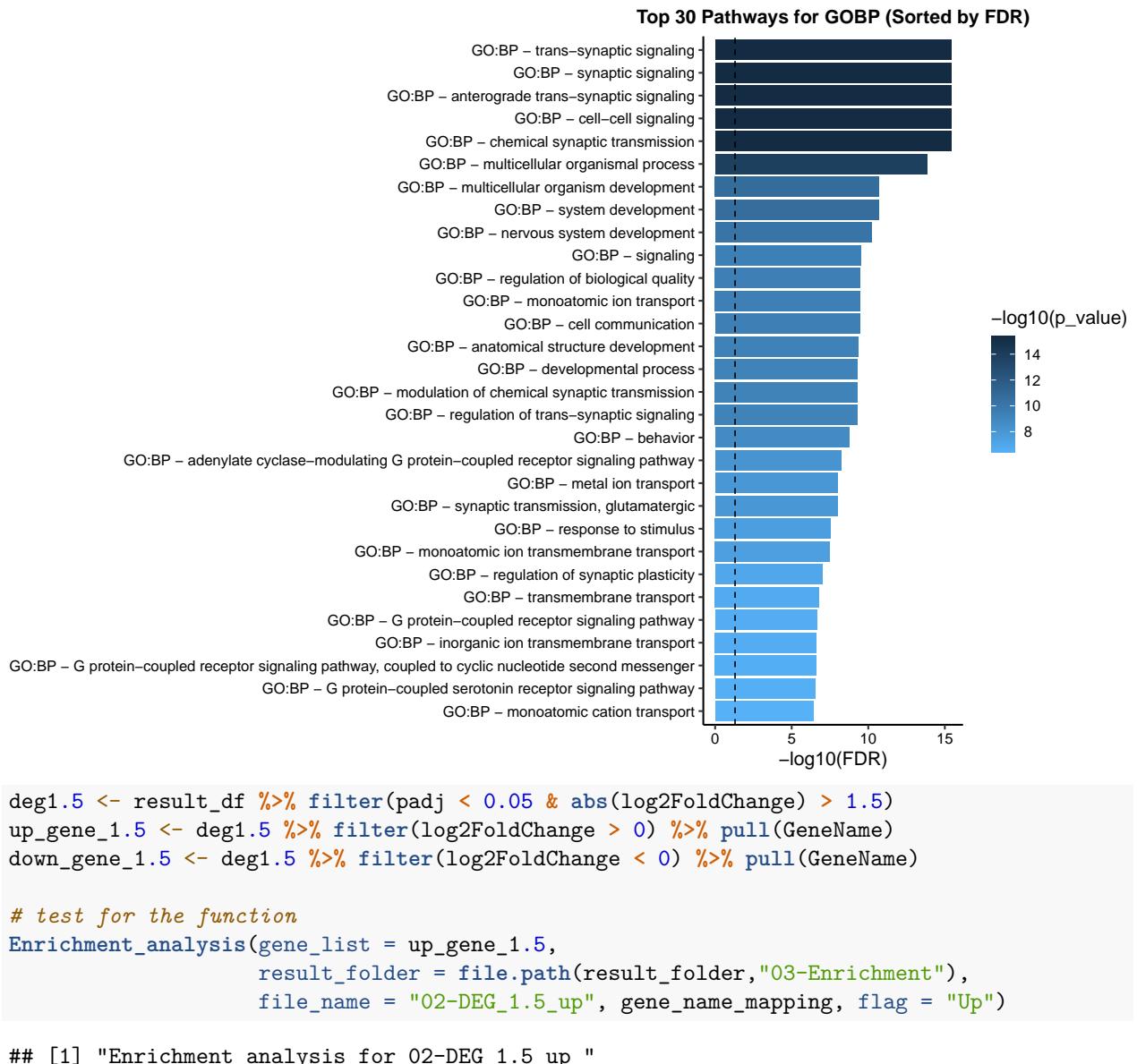


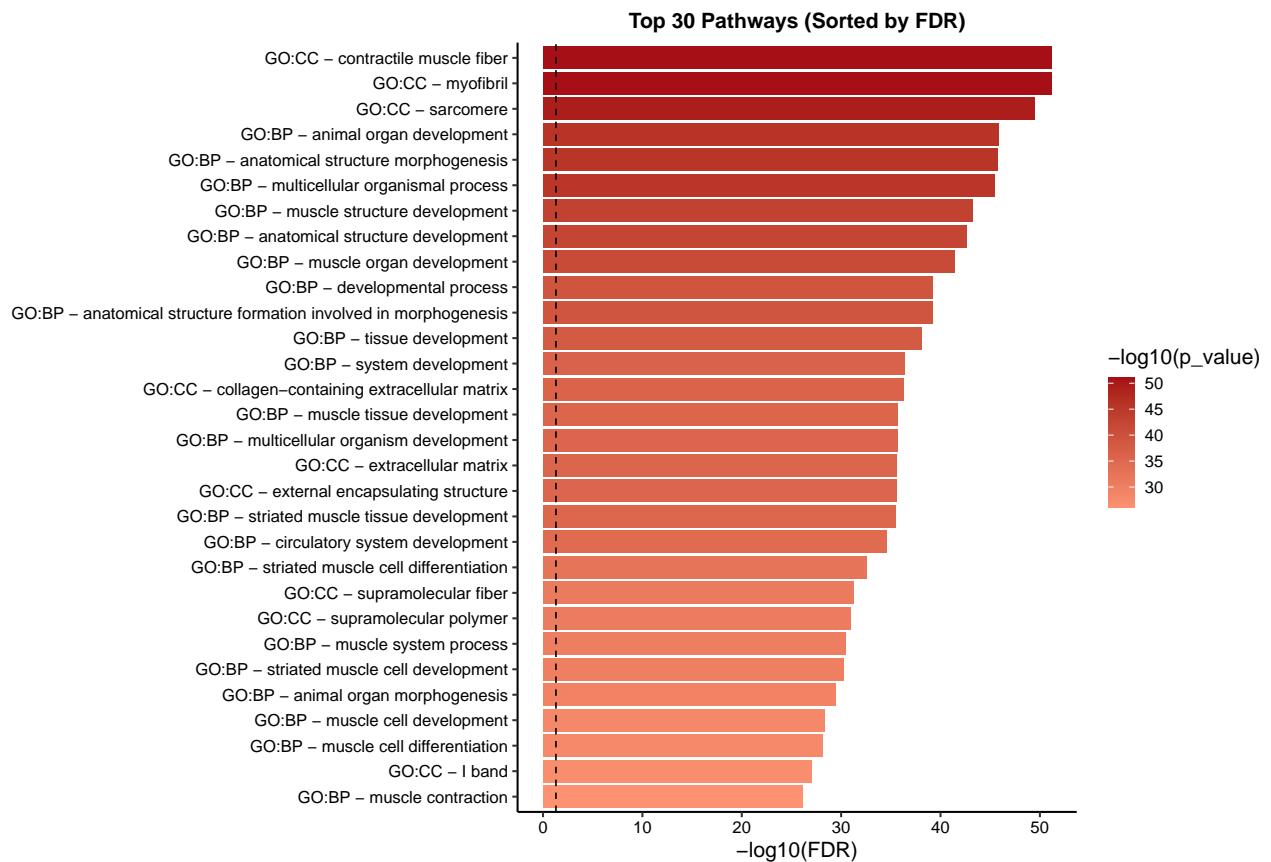
```
Enrichment_analysis(gene_list = down_gene_1,
                    result_folder = file.path(result_folder,"03-Enrichment"),
                    file_name = "01-DEG_1.0_down", gene_name_mapping, flag = "Down")

## [1] "Enrichment analysis for 01-DEG_1.0_down "
```

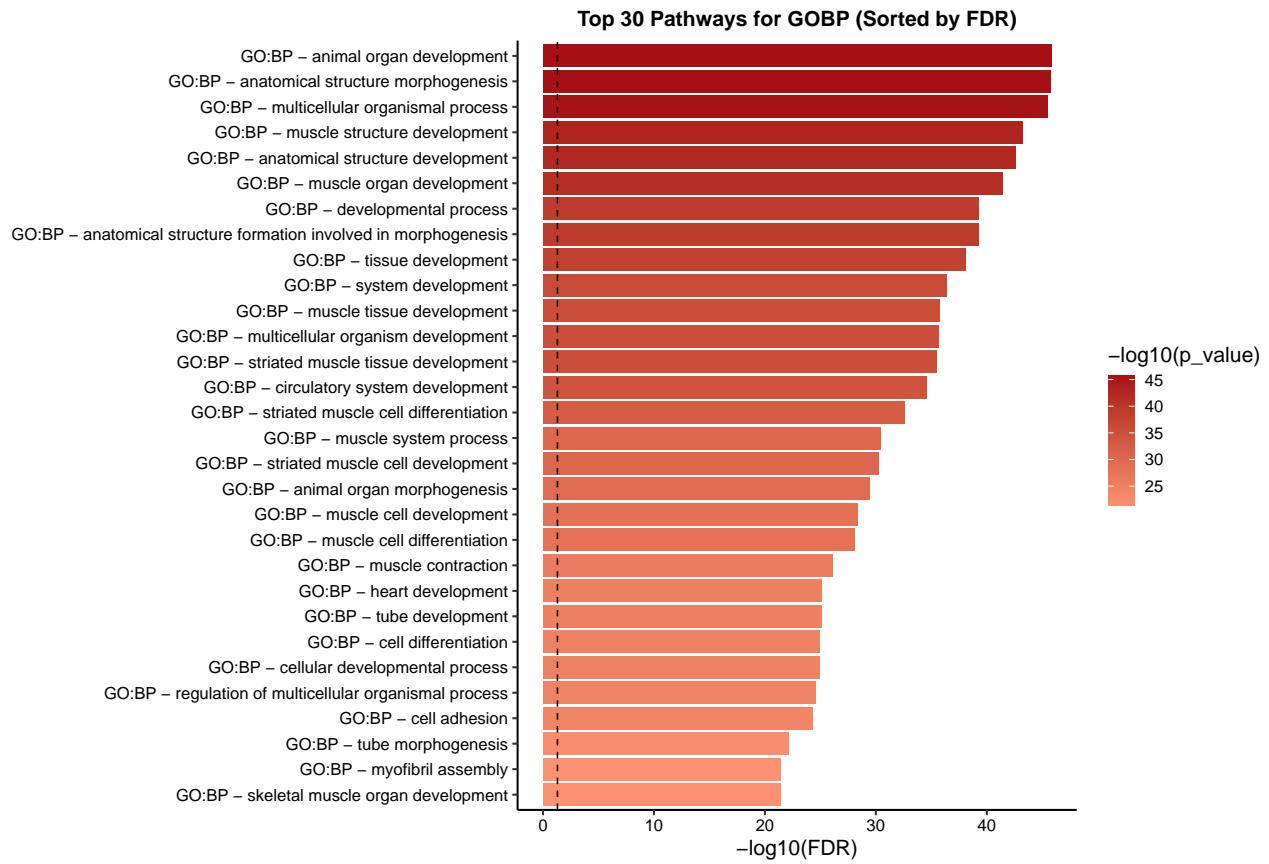


```
## [1] "Enrichment analysis for GOBP_01-DEG_1.0_down "
```



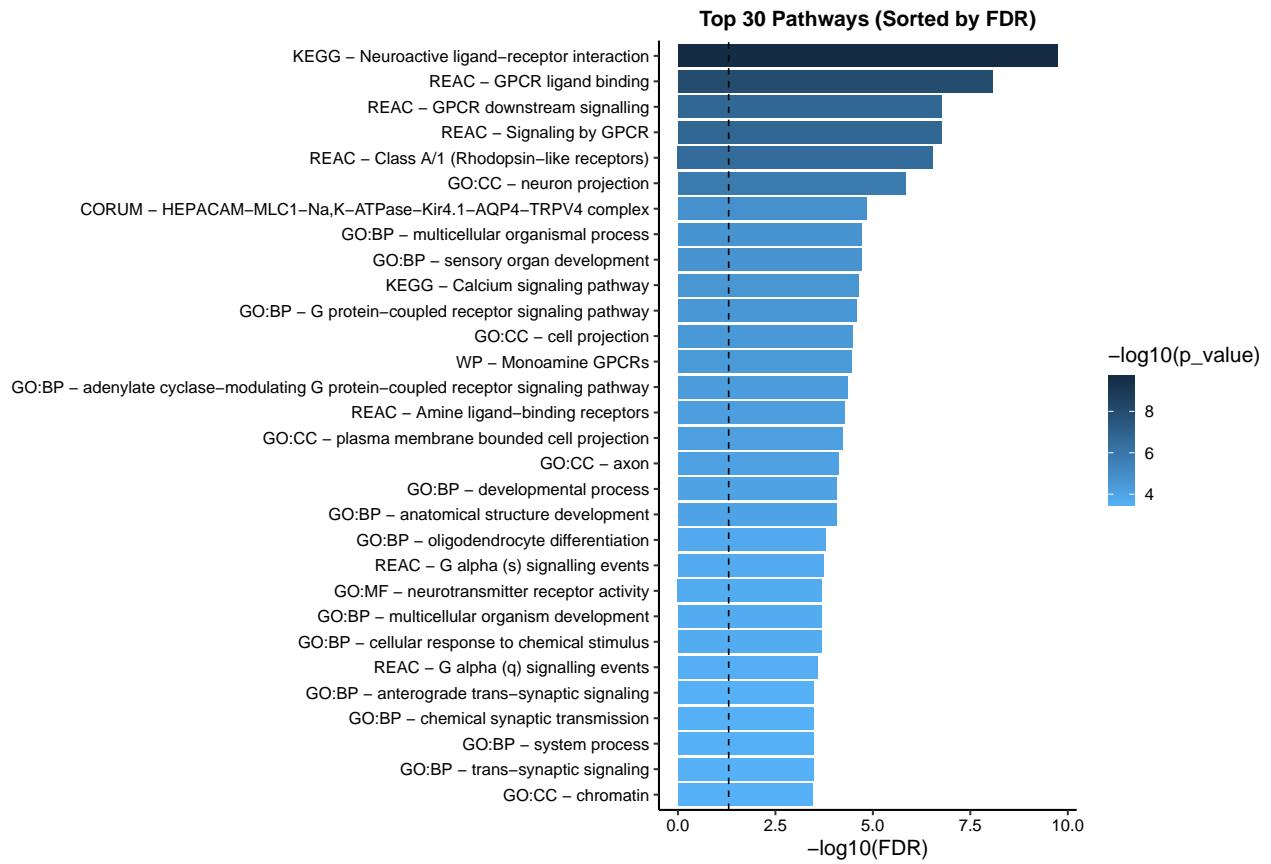


```
## [1] "Enrichment analysis for GOBP_02-DEG_1.5_up "
```

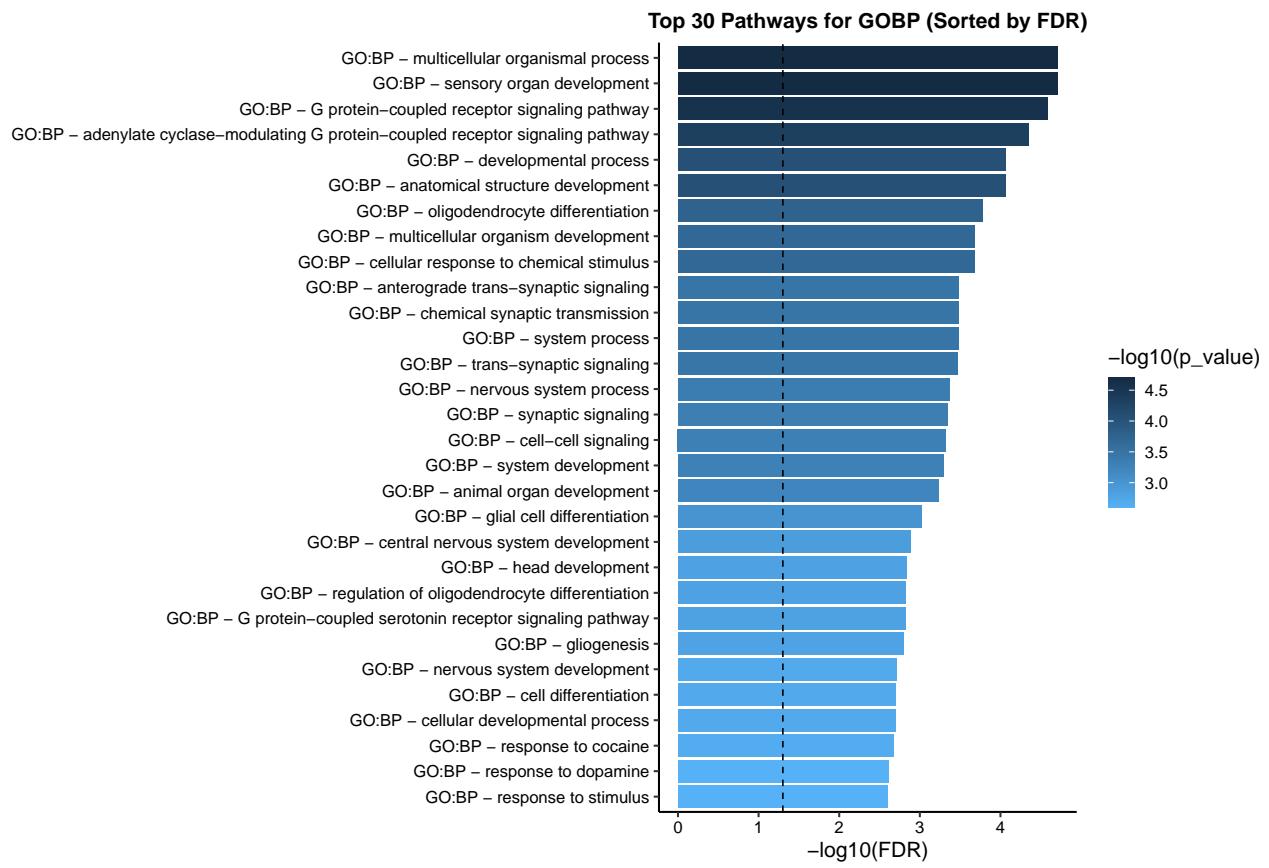


```
Enrichment_analysis(gene_list = down_gene_1.5,
                    result_folder = file.path(result_folder,"03-Enrichment"),
                    file_name = "02-DEG_1.5_down", gene_name_mapping, flag = "Down")

## [1] "Enrichment analysis for 02-DEG_1.5_down "
```



```
## [1] "Enrichment analysis for GOBP_02-DEG_1.5_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;  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] parallel  stats4   stats      graphics  grDevices utils      datasets
## [8] methods   base
##
## other attached packages:
## [1] extrafont_0.19          ggsignif_0.6.4
## [3] patchwork_1.3.0         decoupleR_2.10.0
## [5] GSVA_1.52.3            BiocParallel_1.38.0
## [7] edgeR_4.2.2             limma_3.60.6
## [9] GenomicFeatures_1.56.0  biomaRt_2.60.1
## [11] gprofiler2_0.2.3        RColorBrewer_1.1-3
## [13] data.table_1.16.4       org.Hs.eg.db_3.19.1
## [15] AnnotationDbi_1.66.0   clusterProfiler_4.12.6
## [17] ggfortify_0.4.17        pheatmap_1.0.12
## [19] EnhancedVolcano_1.22.0 ggrepel_0.9.6
## [21] apeglm_1.26.1           DESeq2_1.44.0
## [23] SummarizedExperiment_1.34.0 Biobase_2.64.0
## [25] MatrixGenerics_1.16.0   matrixStats_1.5.0
## [27] reshape2_1.4.4           Matrix_1.7-2
## [29] Signac_1.14.0           Seurat_5.2.1
## [31] SeuratObject_5.0.2      sp_2.2-0
## [33] rtracklayer_1.64.0      GenomicRanges_1.56.2
## [35] GenomeInfoDb_1.40.1     IRanges_2.38.1
## [37] S4Vectors_0.42.1        BiocGenerics_0.50.0
## [39] knitr_1.49               lubridate_1.9.4
## [41]forcats_1.0.0            stringr_1.5.1
## [43] dplyr_1.1.4              purrr_1.0.4
## [45] readr_2.1.5              tidyverse_1.3.1
## [47] tibble_3.2.1             ggplot2_3.5.1
## [49] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] SpatialExperiment_1.14.0   R.methodsS3_1.8.2
## [3] GSEABase_1.66.0           progress_1.2.3
## [5] goftest_1.2-3              HDF5Array_1.32.1
## [7] Biostrings_2.72.1          vctrs_0.6.5
```

```

## [ 9] spatstat.random_3.3-2      digest_0.6.37
## [11] png_0.1-8                  deldir_2.0-4
## [13] parallelly_1.42.0          magick_2.8.5
## [15] MASS_7.3-64                httpuv_1.6.15
## [17] qvalue_2.36.0              withr_3.0.2
## [19] xfun_0.51                 ggfun_0.1.8
## [21] survival_3.8-3            memoise_2.0.1
## [23] gson_0.1.0                systemfonts_1.2.1
## [25] ragg_1.3.3                tidytree_0.4.6
## [27] zoo_1.8-12                pbapply_1.7-2
## [29] R.oo_1.27.0               prettyunits_1.2.0
## [31] KEGGREST_1.44.1           promises_1.3.2
## [33] httr_1.4.7                restfulr_0.0.15
## [35] rhdf5filters_1.16.0        globals_0.16.3
## [37] fitdistrplus_1.2-2        rhdf5_2.48.0
## [39] rstudioapi_0.17.1         UCSC.utils_1.0.0
## [41] miniUI_0.1.1.1            generics_0.1.3
## [43] DOSE_3.30.5               curl_6.2.1
## [45] zlibbioc_1.50.0            ScaledMatrix_1.12.0
## [47] ggraph_2.2.1               polyclip_1.10-7
## [49] GenomeInfoDbData_1.2.12   SparseArray_1.4.8
## [51] xtable_1.8-4               evaluate_1.0.3
## [53] S4Arrays_1.4.1             BiocFileCache_2.12.0
## [55] hms_1.1.3                 irlba_2.3.5.1
## [57] colorspace_2.1-1          filelock_1.0.3
## [59] ROCR_1.0-11               reticulate_1.40.0
## [61] spatstat.data_3.1-4       magrittr_2.0.3
## [63] lmtest_0.9-40              later_1.4.1
## [65] viridis_0.6.5              ggtree_3.12.0
## [67] lattice_0.22-6            spatstat.geom_3.3-5
## [69] future.apply_1.11.3        scattermore_1.2
## [71] XML_3.99-0.18             shadowtext_0.1.4
## [73] cowplot_1.1.3              RcppAnnoy_0.0.22
## [75] pillar_1.10.1              nlme_3.1-167
## [77] compiler_4.4.0             beachmat_2.20.0
## [79] RSpectra_0.16-2            stringi_1.8.4
## [81] tensor_1.5                 GenomicAlignments_1.40.0
## [83] plyr_1.8.9                 crayon_1.5.3
## [85] abind_1.4-8                BiocIO_1.14.0
## [87] gridGraphics_0.5-1          emdbook_1.3.13
## [89] locfit_1.5-9.11            graphlayouts_1.2.2
## [91] bit_4.5.0.1                fastmatch_1.1-6
## [93] textshaping_1.0.0           codetools_0.2-20
## [95] BiocSingular_1.20.0         plotly_4.10.4
## [97] mime_0.12                  splines_4.4.0
## [99] Rcpp_1.0.14                 fastDummies_1.7.5
## [101] sparseMatrixStats_1.16.0   dbplyr_2.5.0
## [103] Rttf2pt1_1.3.12            blob_1.2.4
## [105] here_1.0.1                 fs_1.6.5
## [107] listenv_0.9.1              ggplotify_0.1.2
## [109] statmod_1.5.0              tzdb_0.4.0
## [111] tweenr_2.0.3               pkgconfig_2.0.3
## [113] tools_4.4.0                cachem_1.1.0
## [115] RSQLite_2.3.9               viridisLite_0.4.2

```

```
## [117] DBI_1.2.3
## [119] fastmap_1.2.0
## [121] scales_1.3.0
## [123] ica_1.0-3
## [125] coda_0.19-4.1
## [127] graph_1.82.0
## [129] farver_2.1.2
## [131] scatterpie_0.2.4
## [133] cli_3.6.4
## [135] uwot_0.2.2
## [137] annotate_1.82.0
## [139] gtable_0.3.6
## [141] ggridges_0.5.6
## [143] ape_5.8-1
## [145] RcppHNSW_0.6.0
## [147] bit64_4.6.0-1
## [149] yulab.utils_0.2.0
## [151] bdsmatrix_1.3-7
## [153] spatstat.univar_3.1-1
## [155] lazyeval_0.2.2
## [157] htmltools_0.5.8.1
## [159] GO.db_3.19.1
## [161] rappdirs_0.3.3
## [163] glue_1.8.0
## [165] httr2_1.1.0
## [167] RCurl_1.98-1.16
## [169] treeio_1.28.0
## [171] extrafontdb_1.0
## [173] R6_2.6.1
## [175] labeling_0.4.3
## [177] cluster_2.1.8
## [179] Rhdf5lib_1.26.0
## [181] DelayedArray_0.30.1
## [183] ggforce_0.4.2
## [185] future_1.34.0
## [187] munsell_0.5.1
## [189] htmlwidgets_1.6.4
## [191] rlang_1.1.5
## [193] spatstat.explore_3.3-4
numDeriv_2016.8-1.1
rmarkdown_2.29
grid_4.4.0
Rsamtools_2.20.0
dotCall64_1.2
RANN_2.6.2
tidygraph_1.3.1
yaml_2.3.10
lifecycle_1.0.4
mvtnorm_1.3-3
timechange_0.3.0
rjson_0.2.23
progressr_0.15.1
jsonlite_1.9.0
bitops_1.0-9
Rtsne_0.17
spatstat.utils_3.1-2
GOSemSim_2.30.2
R.utils_2.12.3
shiny_1.10.0
enrichplot_1.24.4
sctransform_0.4.1
tinytex_0.55
spam_2.11-1
XVector_0.44.0
rprojroot_2.0.4
gridExtra_2.3
igraph_2.1.4
SingleCellExperiment_1.26.0
RcppRoll_0.3.1
bbmle_1.0.25.1
aplot_0.2.4
tidyselect_1.2.1
xml2_1.3.6
rsvd_1.0.5
KernSmooth_2.23-26
fgsea_1.30.0
spatstat.sparse_3.1-0
```