Analysis of mass spectrometry data for 1st large library

Ute Hoffmann (Science For Life Laboratory (KTH), Stockholm, Sweden)

juli 23, 2024

Contents

1 Objective of analysis

Mass spectrometry data was collected to check if *Synechocystis* Rubisco (RbcL, RbcS) remains repressed after 3.7 and 8.6 generations and if dCas9 remains induced. The following samples were collected (compare also 2024-04-24_MSsamples.ods): Quadruplicates of - 1% CO2, air, generation 0/4/8 (actual generations: 0, 3.8, 7.3) - 5% CO2, N2, generation 0/4/8 (actual generations: 0, 3.7, 8.6)

2 Read in data

```
protein_slr_ID <- read.delim("2024-01-25_Protein_to_slrIdentifier.tsv", sep="\t", header=T)
names(protein_slr_ID) <- c("locus_tag", "Entry")


highCO2_data <- read.delim("dia_msstats_20240614_highco2_pcc6803.tsv", header=TRUE, dec=".", sep="\t")
highCO2_data <- left_join(highCO2_data, protein_slr_ID)
highCO2_data <- subset(highCO2_data, highCO2_data$ambig == "no" & !is.na(highCO2_data$log2FC) & !is.na(
lowCO2_comparison <- read.delim("dia_msstats_20240614_lowco2_pcc6803.tsv", header=TRUE, dec=".", sep="\
lowCO2_comparison <- left_join(lowCO2_comparison, protein_slr_ID)
lowCO2_comparison <- subset(lowCO2_comparison, lowCO2_comparison$ambig == "no" & !is.na(lowCO2_compariso
```

Separate Tables by comparison:

```
columns <- c("log2FC", "SE", "Tvalue", "DF", "pvalue", "adj.pvalue", "issue", "MissingPercentage", "Ent
highCO2_0_4 <- subset(highCO2_data, highCO2_data$Label=="highco2_4")
highCO2_0_8 <- subset(highCO2_data, highCO2_data$Label=="highco2_8")
write_csv(highCO2_0_4[,columns], "tables/MS_data_comparison_highCO2_gen0-4.csv")
write_csv(highCO2_0_8[,columns], "tables/MS_data_comparison_highCO2_gen0-8")

lowCO2_0_4 <- subset(lowCO2_comparison, lowCO2_comparison$Label=="lowco2_4")
lowCO2_0_8 <- subset(lowCO2_comparison, lowCO2_comparison$Label=="lowco2_8")
```

```
write_csv(subset(lowCO2_comparison, lowCO2_comparison$Label=="lowco2_4")[,columns], "tables/MS_data_comp
write_csv(subset(lowCO2_comparison, lowCO2_comparison$Label=="lowco2_8")[,columns], "tables/MS_data_comp
```

3   Plot Volcano plots

```
small_height <- 12
small_width <- 16
xlims <- c(-6,5.0)
ylims <- c(0,6)
text_toggle <- TRUE
p <- volcanoPlot_ggplot(highCO2_0_4, text=text_toggle, colours=c("#199e77ff", "#da5f02ff", "#d3d3d3b2")
p <- p + xlim(xlims) + ylim(ylims)
p
```
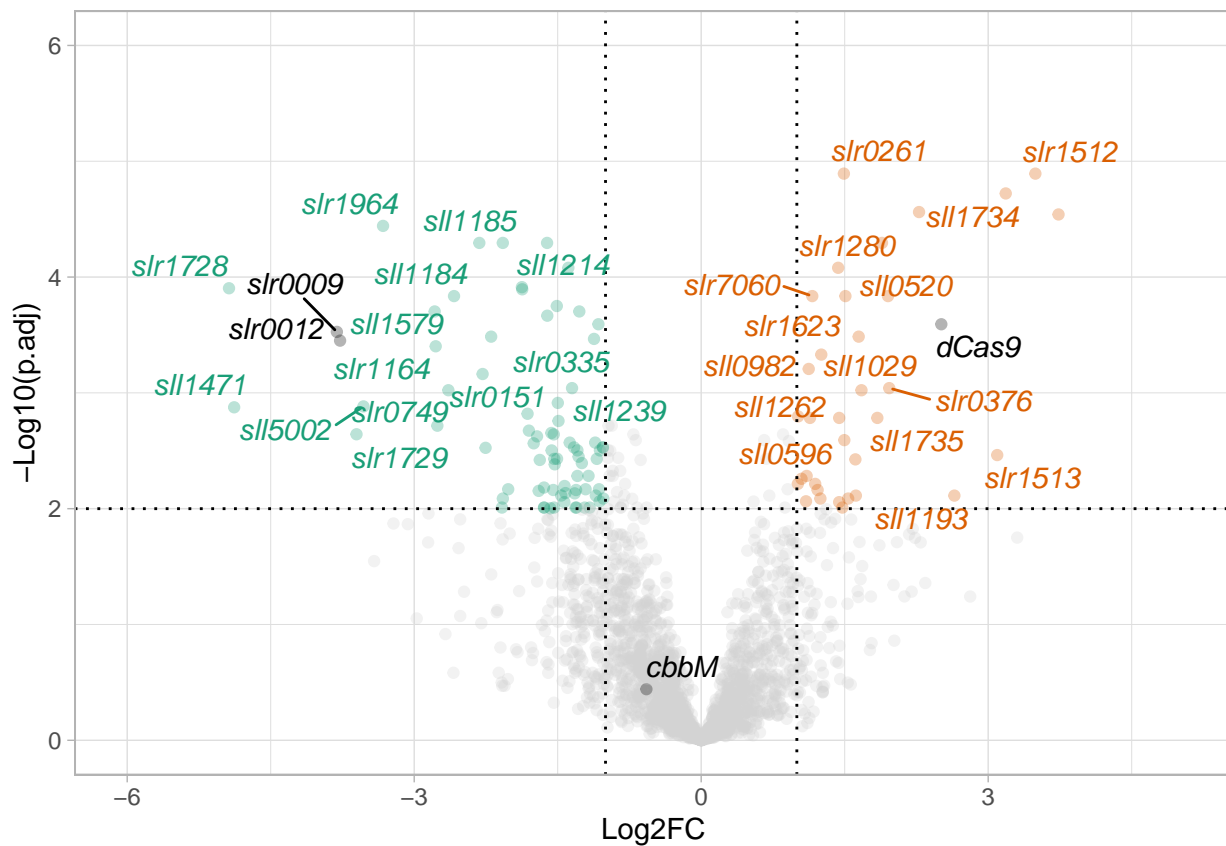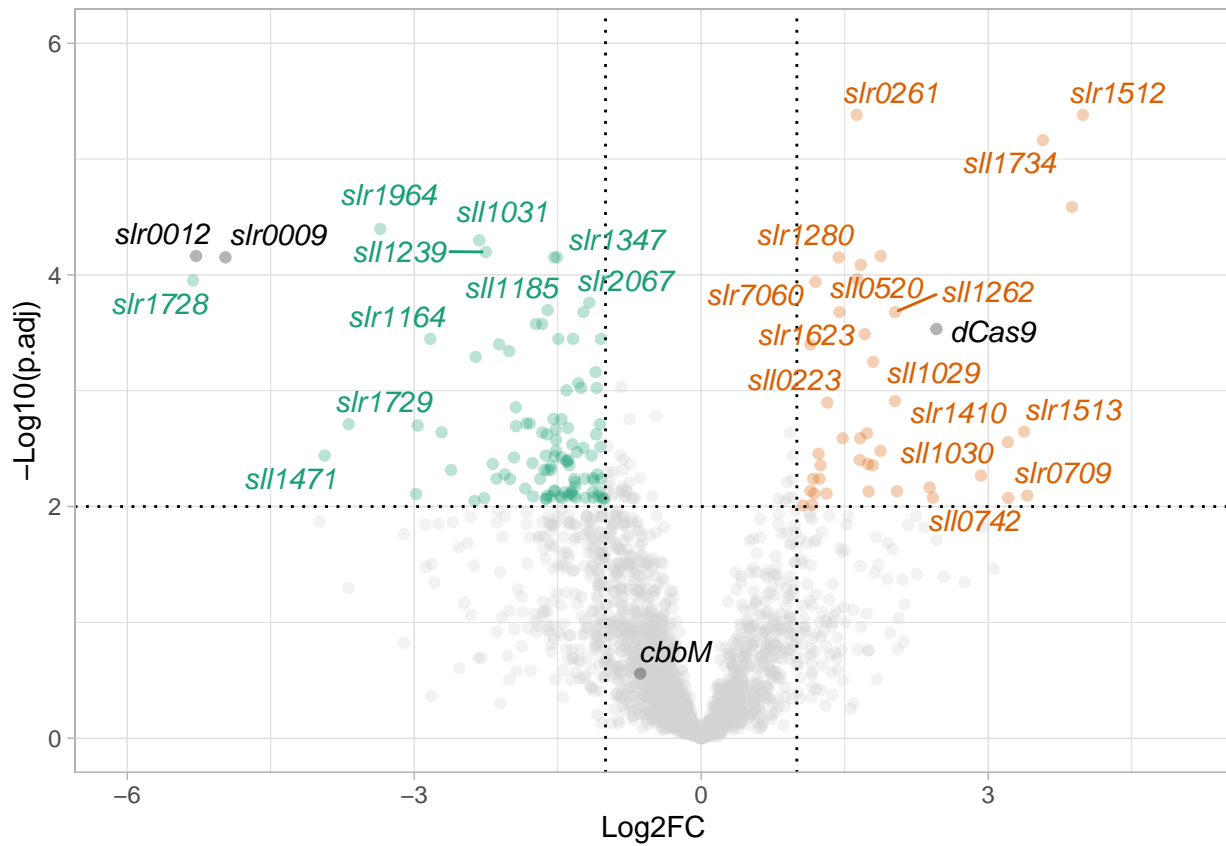


```
ggsave("Figures/Volcano_highCO2_gen0-4_vers1.pdf", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_highCO2_gen0-4_vers2.pdf", plot=p, width=small_width, height=small_height, unit
ggsave("Figures/Volcano_highCO2_gen0-4_vers1.png", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_highCO2_gen0-4_vers2.png", plot=p, width=small_width, height=small_height, unit
nrow(subset(highCO2_0_4, highCO2_0_4$adj.pvalue < 0.01 & highCO2_0_4$log2FC > 1))
```

```
## [1] 35
```

```
nrow(subset(highCO2_0_4, highCO2_0_4$adj.pvalue < 0.01 & highCO2_0_4$log2FC < -1))
```

```
## [1] 77
```

```
p <- volcanoPlot_ggplot(highCO2_0_8, text=text_toggle, colours=c("#199e77ff", "#da5f02ff", "#d3d3d3b2")
p <- p + xlim(xlims) + ylim(ylims)
p
```
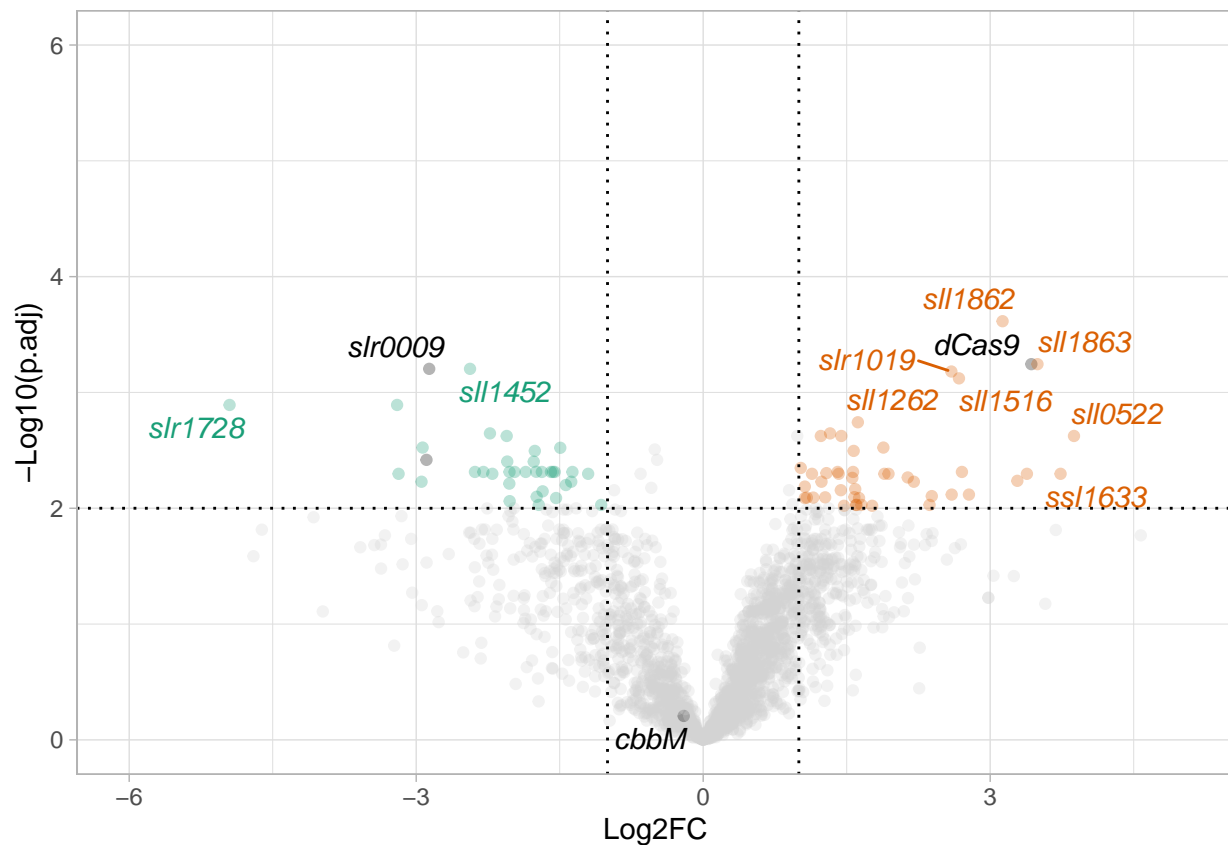


```
ggsave("Figures/Volcano_highCO2_gen0-8_vers1.pdf", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_highCO2_gen0-8_vers2.pdf", plot=p, width=small_width, height=small_height, units
ggsave("Figures/Volcano_highCO2_gen0-8_vers1.png", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_highCO2_gen0-8_vers2.png", plot=p, width=small_width, height=small_height, units
nrow(subset(highCO2_0_8, highCO2_0_8$adj.pvalue < 0.01 & highCO2_0_4$log2FC > 1))
```

```
## [1] 12
```

```
nrow(subset(highCO2_0_8, highCO2_0_8$adj.pvalue < 0.01 & highCO2_0_4$log2FC < -1))
```

```
## [1] 26
```

```
p <- volcanoPlot_ggplot(lowCO2_0_4, text=text_toggle, colours=c("#199e77ff", "#da5f02ff", "#d3d3d3b2"))
p <- p + xlim(xlims) + ylim(ylims)
p
```

```
ggsave("Figures/Volcano_lowCO2_gen0-4_vers1.pdf", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_lowCO2_gen0-4_vers2.pdf", plot=p, width=small_width, height=small_height, units=
nrow(subset(lowCO2_0_4, lowCO2_0_4$adj.pvalue < 0.01 & highCO2_0_4$log2FC > 1))
```

```
## [1] 5
```

```
nrow(subset(lowCO2_0_4, lowCO2_0_4$adj.pvalue < 0.01 & highCO2_0_4$log2FC < -1))
```

```
## [1] 13
```

```
p <- volcanoPlot_ggplot(lowCO2_0_8, text=text_toggle, colours=c("#199e77ff", "#da5f02ff", "#d3d3d3b2"))
p <- p + xlim(xlims) + ylim(ylims)
p
```
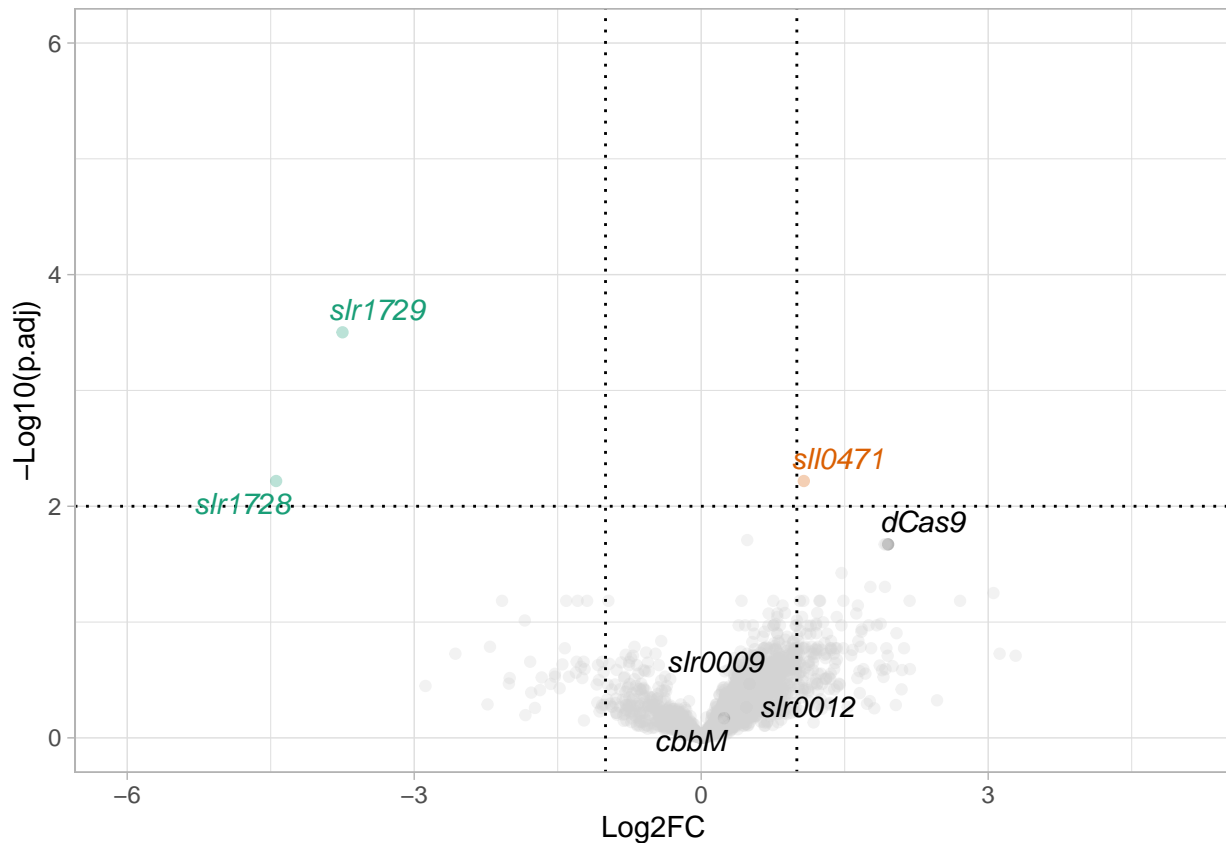
```r
ggsave("Figures/Volcano_lowCO2_gen0-8_vers1.pdf", plot=p, width=24, height=18, units="cm")
ggsave("Figures/Volcano_lowCO2_gen0-8_vers2.pdf", plot=p, width=small_width, height=small_height, units=
nrow(subset(lowCO2_0_8, lowCO2_0_8$adj.pvalue < 0.01 & highCO2_0_4$log2FC > 1))
```

```
## [1] 0
```

```r
nrow(subset(lowCO2_0_8, lowCO2_0_8$adj.pvalue < 0.01 & highCO2_0_4$log2FC < -1))
```

```
## [1] 0
```

4 Extract genes which might be affected by 2PG

```r
ndhR_regulon <- c("sll1594", "sll1732", "sll1733", "sll1734", "sll1735", "slr1512", "slr1513", "slr2006
ndhR_regulon_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% ndhR_regulon)$Entry

cmpr_regulon <- c("slr0040", "slr0041", "slr0042", "slr0043", "slr0044")
cmpr_regulon_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% cmpr_regulon)$Entry

carboxysome <- c("slr1911", "sll1029", "sll1028", "slr1838", "slr1839", "sll1031", "sll1032", "sll1030"
carboxysome_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% carboxysome)$Entry

cyabrB2_regulon <- c("sll1594", "sll0030")
cyabrB2_regulon_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% cyabrB2_regulon)$Entr

rubisco <- c("slr0009", "slr0012")
rubisco_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% rubisco)$Entry
```

```
ccaA <- "slr1347"
ccaA_proteinID <- subset(protein_slr_ID, protein_slr_ID$locus_tag %in% ccaA)$Entry

regulons_uniprot <- c(ndhR_regulon_proteinID, cmpr_regulon_proteinID, carboxysome_proteinID, rubisco_pro

CCM_subset_comp1_uniprot <- subset(highCO2_0_4, highCO2_0_4$Entry %in% regulons_uniprot)
unique(CCM_subset_comp1_uniprot$Entry)
```

```
##  [1] "P73106" "P54205" "P54206" "P73231" "P73393" "P73394" "P73395" "P73862"
##  [9] "P73953" "P72757" "P72758" "P72759" "P72760" "P72761" "Q55517" "Q59991"
## [17] "P73392" "P73406" "P73407" "P73954" "Q54735" "Q55121" "Q55415" "Q55459"
## [25] "Q55460" "Q55461" "Q55462" "Q55463"
```

```
CCM_subset_comp2_uniprot <- subset(highCO2_0_8, highCO2_0_8$Entry %in% regulons_uniprot)
unique(CCM_subset_comp2_uniprot$Entry)
```

```
##  [1] "P73106" "P54205" "P54206" "P73231" "P73393" "P73394" "P73395" "P73862"
##  [9] "P73953" "P72757" "P72758" "P72759" "P72760" "P72761" "Q55517" "Q59991"
## [17] "P73392" "P73406" "P73407" "P73954" "Q54735" "Q55121" "Q55415" "Q55459"
## [25] "Q55460" "Q55462" "Q55463"
```

Create table etc. for that:

```
columns_of_interest <- c("Label", "log2FC", "adj.pvalue", "Entry", "gene_names_ordered_locus")
CCM_subset_comp1_uniprot <- CCM_subset_comp1_uniprot[,columns_of_interest]
CCM_subset_comp1_uniprot <- CCM_subset_comp1_uniprot %>% left_join(protein_slr_ID)
```

```
## Joining with `by = join_by(Entry)`
```

```
CCM_subset_comp2_uniprot <- CCM_subset_comp2_uniprot[,columns_of_interest]
CCM_subset_comp2_uniprot <- CCM_subset_comp2_uniprot %>% left_join(protein_slr_ID)
```

```
## Joining with `by = join_by(Entry)`
```

```
wider_CCM_comp1 <- pivot_wider(CCM_subset_comp1_uniprot, values_from=c(log2FC, adj.pvalue), names_from=
wider_CCM_comp1 <- wider_CCM_comp1 %>% left_join(annotation_2)
```

```
## Joining with `by = join_by(gene_names_ordered_locus)`
```

```
write_csv(wider_CCM_comp1, "tables/CCM_subset_comparison1.csv")

#intermed <- subset(CCM_subset_comp2_uniprot, !CCM_subset_comp2_uniprot$gene_names_ordered_locus%in%c("
wider_CCM_comp2 <- pivot_wider(CCM_subset_comp2_uniprot, values_from=c(log2FC, adj.pvalue), names_from=
wider_CCM_comp2 <- wider_CCM_comp2 %>% left_join(annotation_2)
```

```
## Joining with `by = join_by(gene_names_ordered_locus)`
```

```
write_csv(wider_CCM_comp2, "tables/CCM_subset_comp2.csv")
```

Session info

```
## R version 4.4.1 (2024-06-14)
## Platform: x86_64-pc-linux-gnu
## 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=sv_SE.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=sv_SE.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=sv_SE.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=sv_SE.UTF-8 LC_IDENTIFICATION=C
## 
## time zone: Europe/Stockholm
## tzcode source: system (glibc)
## 
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base     
## 
## other attached packages:
##  [1] enrichplot_1.24.0     clusterProfiler_4.12.0 ggrepel_0.9.5
##  [4] colorblindr_0.1.0     colorspace_2.1-1       lubridate_1.9.3
##  [7] forcats_1.0.0         stringr_1.5.1          dplyr_1.1.4
## [10] purrr_1.0.2           readr_2.1.5            tidyr_1.3.1
## [13] tibble_3.2.1          tidyverse_2.0.0        ggplot2_3.5.1
## [16] knitr_1.47
## 
## loaded via a namespace (and not attached):
##  [1] DBI_1.2.2             gson_0.1.0           shadowtext_0.1.3
##  [4] gridExtra_2.3         rlang_1.1.3          magrittr_2.0.3
##  [7] DOSE_3.30.1           compiler_4.4.1       RSQLite_2.3.7
## [10] systemfonts_1.1.0     png_0.1-8            vctrs_0.6.5
## [13] reshape2_1.4.4        pkgconfig_2.0.3      crayon_1.5.2
## [16] fastmap_1.2.0         XVector_0.44.0       labeling_0.4.3
## [19] ggraph_2.2.1          utf8_1.2.4           HDO.db_0.99.1
## [22] rmarkdown_2.27        tzdb_0.4.0           UCSC.utils_1.0.0
## [25] ragg_1.3.2            tinytex_0.51         bit_4.0.5
## [28] xfun_0.44             zlibbioc_1.50.0      cachem_1.1.0
## [31] aplot_0.2.2           GenomeInfoDb_1.40.1  jsonlite_1.8.8
## [34] blob_1.2.4            highr_0.11           BiocParallel_1.38.0
## [37] tweenr_2.0.3          parallel_4.4.1       R6_2.5.1
## [40] RColorBrewer_1.1-3    stringi_1.8.4        GOSemSim_2.30.0
## [43] Rcpp_1.0.12           IRanges_2.38.0       Matrix_1.6-5
## [46] splines_4.4.1         igraph_2.0.3         timechange_0.3.0
## [49] tidyselect_1.2.1      qvalue_2.36.0        rstudioapi_0.16.0
## [52] yaml_2.3.8            viridis_0.6.5        codetools_0.2-19
## [55] lattice_0.22-5        plyr_1.8.9           treeio_1.28.0
## [58] Biobase_2.64.0        withr_3.0.0          KEGGREST_1.44.0
## [61] evaluate_0.23         gridGraphics_0.5-1   scatterpie_0.2.2
## [64] polyclip_1.10-6       Biostrings_2.72.0    ggtree_3.12.0
## [67] pillar_1.9.0          stats4_4.4.1         ggfun_0.1.5
## [70] generics_0.1.3        vroom_1.6.5          S4Vectors_0.42.0
## [73] hms_1.1.3             tidytree_0.4.6       munsell_0.5.1
## [76] scales_1.3.0          glue_1.7.0           lazyeval_0.2.2
## [79] tools_4.4.1           data.table_1.15.4    fgsea_1.30.0
## [82] fs_1.6.4              graphlayouts_1.1.1   fastmatch_1.1-4
## [85] tidygraph_1.3.1       cowplot_1.1.3.9000   grid_4.4.1
## [88] ape_5.8               AnnotationDbi_1.66.0 nlme_3.1-165
```

```
##  [91] patchwork_1.2.0        GenomeInfoDbData_1.2.12 ggforce_0.4.2
##  [94] cli_3.6.2              textshaping_0.4.0      fansi_1.0.6
##  [97] viridisLite_0.4.2      gtable_0.3.5           yulab.utils_0.1.4
## [100] digest_0.6.35          BiocGenerics_0.50.0    ggplotify_0.1.2
## [103] farver_2.1.2           memoise_2.0.1          htmltools_0.5.8.1
## [106] lifecycle_1.0.4        httr_1.4.7             GO.db_3.19.1
## [109] bit64_4.0.5            MASS_7.3-61
```