# canaryGO Analysis

Carson Stacy

2022-09-29

```
# ensure BiocManager is installed, install if not
if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install(version = "3.16")

#ensure all required packages are installed
if(!"biomaRt" %in% installed.packages()) BiocManager::install("biomaRt")
if(!"topGO" %in% installed.packages()) BiocManager::install("topGO")
if(!"edgeR" %in% installed.packages()) BiocManager::install("edgeR")
if(!"tidyverse" %in% installed.packages()) install.packages("tidyverse")

library(tidyverse)
library(topGO)
library(biomaRt)
library(here)
library(edgeR)
# library(ggpubr)
```

#TODO: CHECK HOW MANY lncRNA show up in our edgeR outputs. # if they aren't interesting, let's get rid of them and redo our edgeR output. # also probably throw out the tRNAs anyway. # I'll check if BiomaRt works for this.

```
#load in edgeR output to get gene names
# load in edgeR output
MGLpost0 <- readr::read_csv("~/Desktop/canary/contrasts/MGLpost0.csv")
MGpost0 <- readr::read_csv("~/Desktop/canary/contrasts/MGpost0.csv")
MGd140 <- readr::read_csv("~/Desktop/canary/contrasts/MGd140.csv")
MGPpost0 <- readr::read_csv("~/Desktop/canary/contrasts/MGPpost0.csv")
MGd2114 <- readr::read_csv("~/Desktop/canary/contrasts/MGd2114.csv")
Diet0 <- readr::read_csv("~/Desktop/canary/contrasts/Diet0.csv")
MGDietpost0 <- readr::read_csv("~/Documents/GitHub/CanarySeq/contrasts/MGDietpost0.csv")
CtlDietpost0 <- readr::read_csv("~/Documents/GitHub/CanarySeq/contrasts/CtlDietpost0.csv")
CtlLpost0 <- readr::read_csv("~/Documents/GitHub/CanarySeq/contrasts/CtlLpost0.csv")
CtlPpost0 <- readr::read_csv("~/Documents/GitHub/CanarySeq/contrasts/CtlPpost0.csv")

# need to add this contrast for analysis!
######
# MGDietpost0 <- (((groupMG.lipid.14 + groupMG.lipid.21)/2) - (groupMG.lipid.0)) -
#   (((groupMG.protein.14 + groupMG.protein.21)/2) - (groupMG.protein.0))
#######
# DietMGpost0
```

```
contrasts <- list(
  MGLpost0 = MGLpost0,
  MGpost0 = MGpost0,
  MGd140 = MGd140,
  MGPpost0 = MGPpost0,
  MGd2114 = MGd2114,
  Diet0 = Diet0,
  MGDietpost0 = MGDietpost0,
  # CtlDietpost0 = CtlDietpost0#, # I am checking if this was the contrast with the cool GO term enrich
  CtlLpost0 = CtlLpost0,
  CtlPpost0 = CtlPpost0
)
```

```
# get biomart ensembl dataset for scanaria (has most GO terms)
mart <- if(exists('mart') == TRUE){
  mart
} else {
  useMart(biomart = "ensembl", dataset = "scanaria_gene_ensembl")
}
```

```
# loop through each dataset
for(i in 1:length(contrasts)) {
  # get names of genes for which to get annotations
  transcript_ids <- contrasts[i][[1]][1] #names of all genes in edgeR output

  res <- getBM(attributes = c('ensembl_transcript_id',
                              'ensembl_peptide_id',
                              'ensembl_gene_id',
                              'uniprot_gn_symbol',
                              'wikigene_name',
                              'entrezgene_id',
                              'external_gene_name',
                              'go_id',
                              'description'
                              ),
               # filters = 'ensembl_transcript_id',
               values = transcript_ids,
               mart = mart)

  # changing format of the output for subsequent analysis
  # getting GO terms grouped to gene names
  res_min <- res %>%
    mutate(wikigene_name = ifelse(wikigene_name %in% "",
                                  external_gene_name,
                                  wikigene_name))%>% # get gene names
    group_by(ensembl_gene_id) %>% # I think I need to change this...
    mutate(GOterms = paste0(go_id, collapse = ",")) %>%
    slice_head(n=1) %>% # get only one row per gene
    ungroup()


  res_min <- res_min %>%
    mutate(GOterms = gsub("^,+|,+$|(,)+", "\\1", GOterms, perl=TRUE)) %>%  # tidy up the extra ;;; in G
```

```
  mutate_all(na_if,"")


geneID2GO_combine_names <- res_min[,c(5,7,10)] %>%
  mutate(...1 = ifelse(is.na(res_min$wikigene_name), res_min$external_gene_name, res_min$wikigene_name
  dplyr::select(-c(wikigene_name, external_gene_name))

geneID_names <- left_join(contrasts[i][[1]][1], geneID2GO_combine_names, by = "...1")
colnames(geneID_names)[1] <- "seq_name"

# build the table of original GO terms with gene IDs
geneID2GO_build_BM <- tibble(
                          # canary_annotations$geneID_ncbi,
                          geneID_names$seq_name,
                          "\t",
                          geneID_names$GOterms
                          )

colnames(geneID2GO_build_BM) <- c("seq_name", "\t", "go_i_ds")

# load in Blast2GO results
Blast2GO <- read_delim("~/Desktop/canary/newGOterms_nomatchDEgenes.tsv",
    delim = "\t", escape_double = FALSE,
    trim_ws = TRUE) %>%
  janitor::clean_names() %>%
  dplyr::select(seq_name, go_i_ds) %>%
#change formatting of the GO term seperators to ,
  dplyr::mutate(go_i_ds = str_replace_all(go_i_ds, "; ", ",")) %>%
  dplyr::mutate(go_i_ds = str_replace_all(go_i_ds, "[PFC]:", ""))

combinedGOterms <- left_join(geneID2GO_build_BM, Blast2GO, by = "seq_name")

# now condense the two go_i_ds columns into single column of GO terms.
combinedGOterms <-
  tidyr::unite(combinedGOterms,
               GOterms,
               go_i_ds.x:go_i_ds.y,
                 na.rm = TRUE, remove = FALSE,
                 sep = ",")
geneID2GO_build_BM <- tibble(
                        # canary_annotations$geneID_ncbi,
                        seq_name = combinedGOterms$seq_name,
                        "\t",
                        GOterms = combinedGOterms$GOterms
                        )

# write out the geneID2GO_BM.map file
write.table(geneID2GO_build_BM, file=here("geneID2GO_BM.map"),
            quote = F, row.names = F, col.names = F)
save(geneID2GO_build_BM,
     file = paste( "~/Documents/GitHub/CanarySeq/geneAnnotations/geneID2GO_", names(contrasts)[i], ".

# read in with readMappings
```

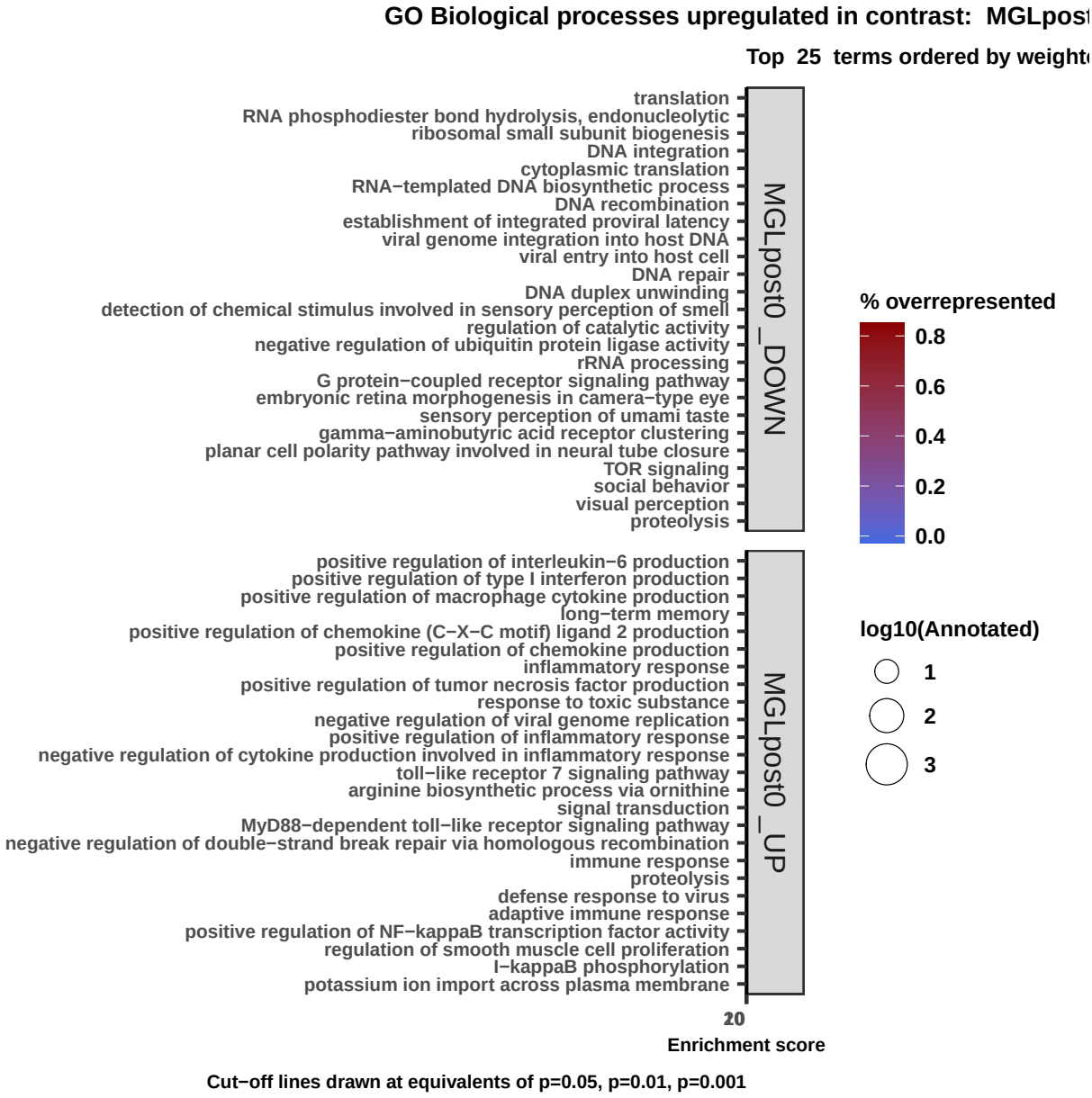
```r
canary_geneID2GO_BM <- topGO::readMappings(file = here("geneid2GO_BM.map"))

# save these mappings
save(canary_geneID2GO_BM, file = paste( "~/Documents/GitHub/CanarySeq/geneAnnotations/canary_geneID2G
# "~/Desktop/canary/canaryGeneID2GOMappings_BM.RData")


#load in canary geneID2GO map
load(file = paste( "~/Documents/GitHub/CanarySeq/geneAnnotations/canary_geneID2GO_BM_", names(contras

# Numeric ID values (in edgeR output) of genes whose names match Gene Ontology Estimates with edgeR o
matches <- which(toupper(names(canary_geneID2GO_BM)) %in% contrasts[i][[1]][[1]])
          # these names came from `edgeR diff expression.R` file

# # Numeric ID values (in eggNOG output) of genes whose names match Gene Ontology Estimates with edge
# matches_rnaIndex <- which(
#   contrasts[i][[1]][[1]] %in% toupper(names(canary_geneID2GO_BM))
#   )

#genes that are in both.
# used unique because there are some duplicates in output. (per isoform)
genesMatched <- unique(toupper(names(canary_geneID2GO_BM))[matches])

# loop through different lfc cutoffs I want to compare (and up v down)
# for (j in c(1.1,1.2,1.5)) {
for (j in c(1.2)) {
# for (j in c(1.1)) {
# for (j in c(1.5)) {
# for (j in c(2)) {
# for (j in c(1.01,1.1,1.2,1.5)) {
  DEgenes_up <- filter(as.data.frame(contrasts[[i]]), logFC > log2(j)) %>% pull(`...1`)
  DEgenes_down <- filter(contrasts[[i]], logFC < -log2(j)) %>% pull(`...1`)

  # DEgenes_up <- filter(as.data.frame(contrasts[[i]]), logFC > j) %>% pull(`...1`)
  # DEgenes_down <- filter(contrasts[[i]], logFC < -j) %>% pull(`...1`)

  # genes names with GO terms and DE.
  DEgeneswithGO_up <- which(DEgenes_up %in% genesMatched)
  DEgeneswithGO_down <- which(DEgenes_down %in% genesMatched)

  # select only genes that are DE and have GO terms
  DEgenesGO_up <- DEgenes_up[DEgeneswithGO_up]
  DEgenesGO_down <- DEgenes_down[DEgeneswithGO_down]

  #first, let's subset our canary geneIDtoGO to just these genes.
  canary_geneID2GO_BM_DE_up <-
    canary_geneID2GO_BM[which(toupper(names(canary_geneID2GO_BM)) %in% DEgenesGO_up)]

  names(canary_geneID2GO_BM_DE_up)[!(DEgenesGO_up %in% names(canary_geneID2GO_BM_DE_up))]

  canary_geneID2GO_BM_DE_down <-
    canary_geneID2GO_BM[which(toupper(names(canary_geneID2GO_BM)) %in% DEgenesGO_down)]
```

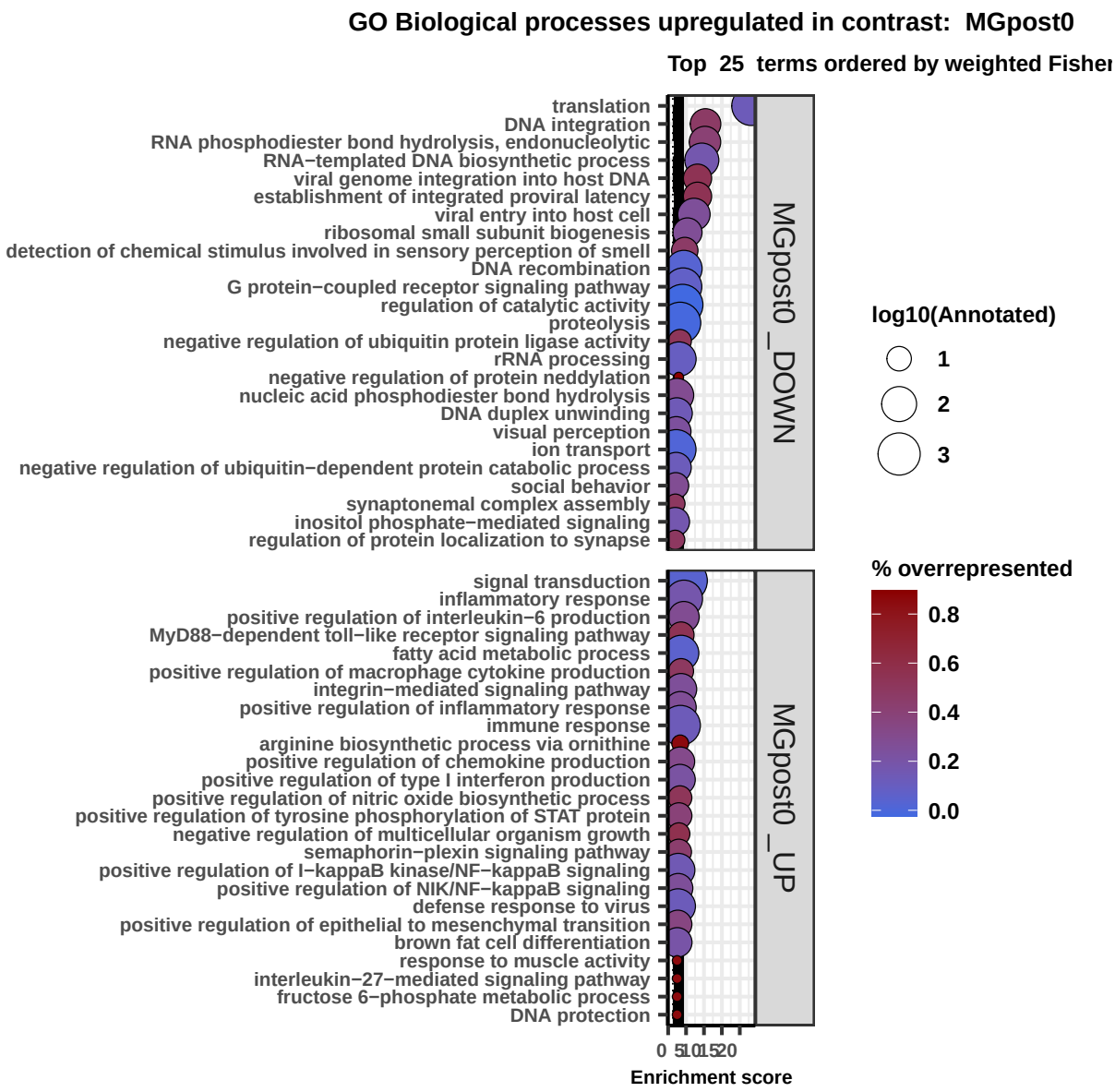
```
    names(canary_geneID2GO_BM_DE_down)[!(DEgenesGO_down %in% names(canary_geneID2GO_BM_DE_down))]


    #Let me also get my full edgeR gene list.
    canary_geneID2GO_BM_edgeR <- canary_geneID2GO_BM[which(toupper(names(canary_geneID2GO_BM)) %in% gene


    # weird formatting creating the geneList for topGO. this code seemed to work
    all_genes_up <- factor(as.integer(toupper(contrasts[i][[1]][[1]]) %in% DEgenesGO_up)) # create 0 1
    names(all_genes_up) <- toupper(contrasts[i][[1]][[1]]) # add gene names

    all_genes_down <- factor(as.integer(toupper(contrasts[i][[1]][[1]]) %in% DEgenesGO_down)) # create
    names(all_genes_down) <- toupper(contrasts[i][[1]][[1]]) # add gene names


    GOdata_up=new('topGOdata',
           ontology='BP', # 'BP' = Biological Process. Also tried 'ALL' but doesn't work.
           allGenes = all_genes_up, # all 10188 genes that are in MGpost0 data from edgeR. factor=1 for
           annot = annFUN.gene2GO, # annotation function for custom annotation
           # gene2GO = canary_geneID2GO_BM # annotation db from eggNOG output.
# I am thinking about switching this for one filtered to just edgeR genes:
           gene2GO = canary_geneID2GO_BM_edgeR # annotation db from BiomaRt.
           # gene2GO = canary_geneID2GO_BM_DE#,
           # geneSel = DE_genes
           )

    GOdata_down=new('topGOdata',
           ontology='BP', # 'BP' = Biological Process. Also tried 'ALL' but doesn't work.
           allGenes = all_genes_down, # all 10188 genes that are in MGpost0 data from edgeR. factor=1 f
           annot = annFUN.gene2GO, # annotation function for custom annotation
           # gene2GO = canary_geneID2GO_BM # annotation db from eggNOG output.
# I am thinking about switching this for one filtered to just edgeR genes:
           gene2GO = canary_geneID2GO_BM_edgeR # annotation db from BiomaRt.
           # gene2GO = canary_geneID2GO_BM_DE#,
           # geneSel = DE_genes
           )

    #analysis
    # define test using the weight01 algorithm (default) with fisher
    weight_fisher_result_up=runTest(GOdata_up, algorithm='weight01', statistic='fisher')
    weight_fisher_result_down=runTest(GOdata_down, algorithm='weight01', statistic='fisher')

    # generate a table of results: we can use the GenTable function to generate a summary table with th
    allGO_up=usedGO(GOdata_up)
    all_res_up=GenTable(GOdata_up, weightFisher=weight_fisher_result_up, orderBy='weightFisher', topNod

    allGO_down=usedGO(GOdata_down)
    all_res_down=GenTable(GOdata_down, weightFisher=weight_fisher_result_down, orderBy='weightFisher',

    #performing BH correction on our p values
    p.adj_up=round(p.adjust(all_res_up$weightFisher,method="BH"),digits = 4)
    p.adj_down=round(p.adjust(all_res_down$weightFisher,method="BH"),digits = 4)
```

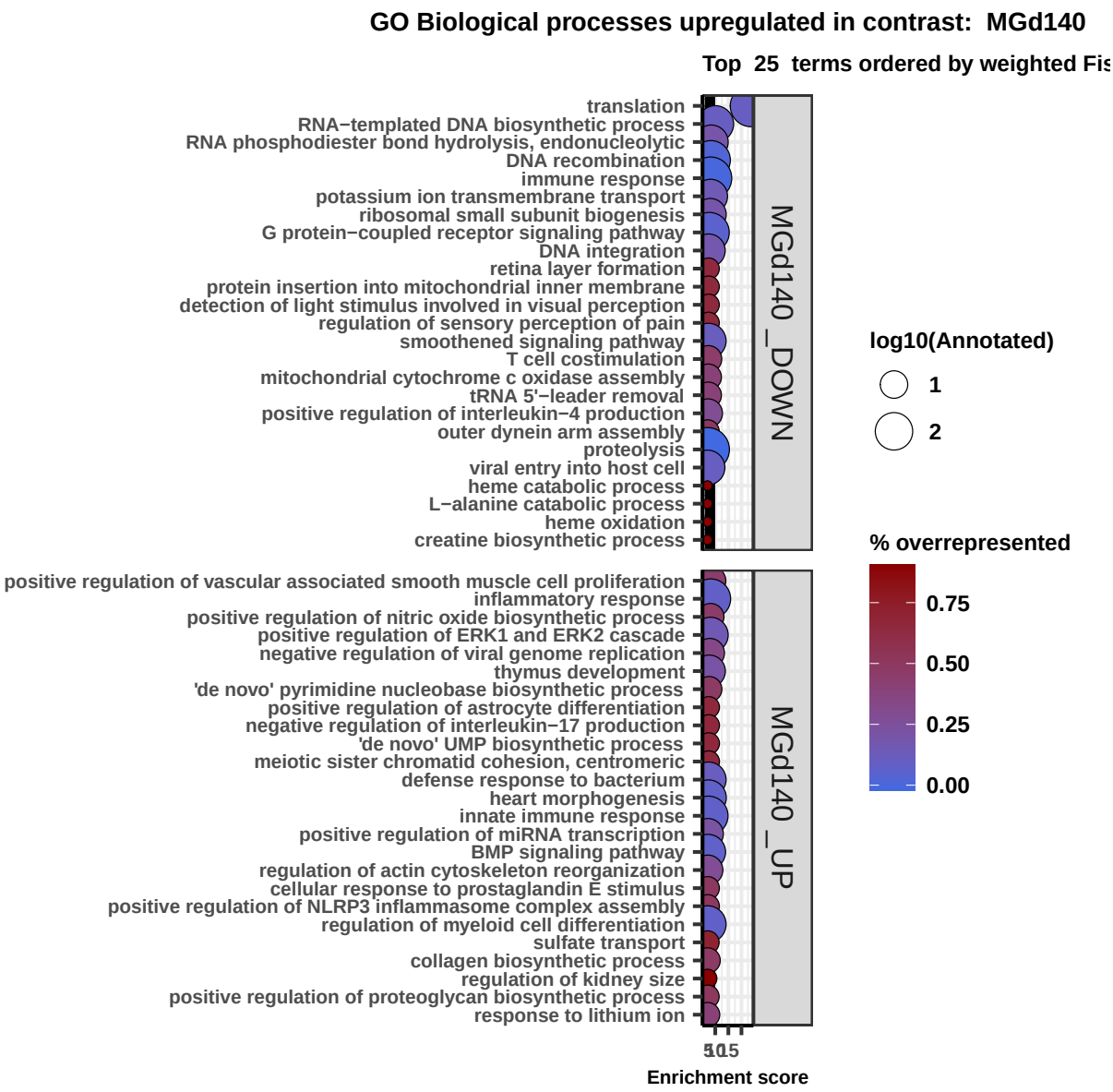
```
  # create the file with all the statistics from GO analysis
  all_res_final_up=cbind(all_res_up,p.adj_up)
  all_res_final_up=all_res_final_up[order(all_res_final_up$p.adj),] %>%
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))

  all_res_final_down=cbind(all_res_down,p.adj_down)
  all_res_final_down=all_res_final_down[order(all_res_final_down$p.adj),] %>%
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))

  #get list of significant GO before multiple testing correction
  results.table.p_up= all_res_final_up[which(as.numeric(all_res_final_up$weightFisher)<=0.1),] %>%
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))

  results.table.p_down= all_res_final_down[which(as.numeric(all_res_final_down$weightFisher)<=0.1),] %
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))

  #get list of significant GO after multiple testing correction
  results.table.bh_up=all_res_final_up[which(all_res_final_up$p.adj<=0.1),] %>%
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))

  results.table.bh_down=all_res_final_down[which(all_res_final_down$p.adj<=0.1),] %>%
    mutate(across(everything(), gsub, pattern = ",", replacement = ";"))


  #save first top 50 ontolgies sorted by adjusted pvalues
  write.table(all_res_final_up[1:50,],
              file = paste(
                "~/Documents/GitHub/CanarySeq/geneAnnotations/geneID2GO_",
                names(contrasts)[i], "log2of", as.character(j), "_up.csv", sep= ""),
              # "~/Desktop/canary/summary_topGO_analysis.csv",
              sep=",",
              quote=FALSE,
              row.names=FALSE)

  write.table(all_res_final_down[1:50,],
              file = paste(
                "~/Documents/GitHub/CanarySeq/geneAnnotations/geneID2GO_",
                names(contrasts)[i], "log2of", as.character(j), "_down.csv", sep= ""),
              # "~/Desktop/canary/summary_topGO_analysis.csv",
              sep=",",
              quote=FALSE,
              row.names=FALSE)




  # PLOT the GO hierarchy plot: the enriched GO terms are colored in yellow/red according to signific

  # pdf(file='~/Desktop/canary/topGOPlot_fullnames.pdf', height=12, width=12, paper='special', points
  # showSigOfNodes(GOdata_down, score(weight_fisher_result_up), useInfo = "none", sigForAll=FALSE, fi
  # dev.off()
}
```

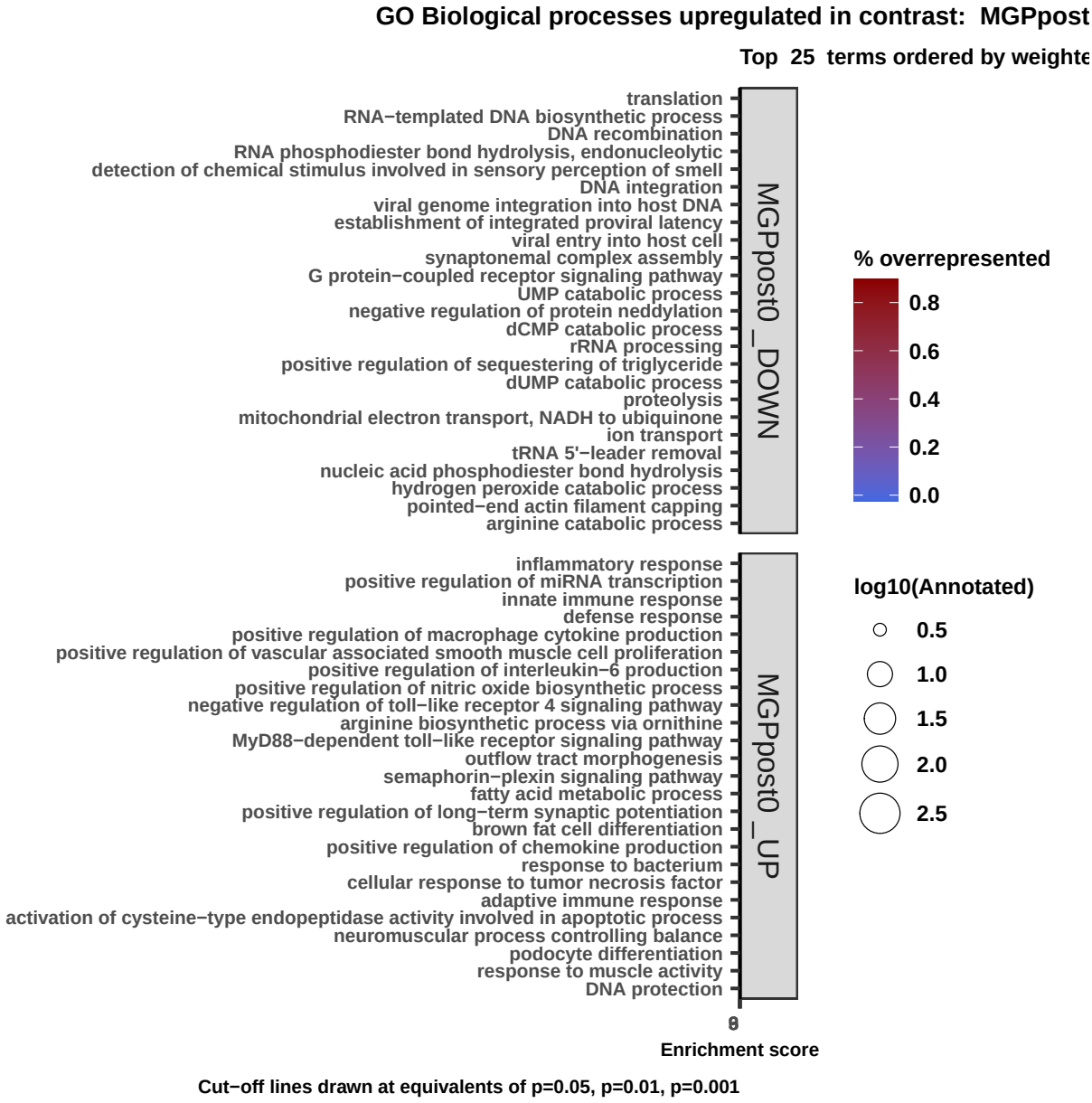
```r
# visualize results at final j value
ntop <- 25

ggdata_up <- all_res_up[1:ntop,] %>%
  mutate(enrichment = paste(names(contrasts[i]),"_UP"))
ggdata_up$Term <- factor(ggdata_up$Term, levels = rev(ggdata_up$Term)) # fixes order

ggdata_down <- all_res_down[1:ntop,] %>%
  mutate(enrichment = paste(names(contrasts[i]),"_DOWN"))
ggdata_down$Term <- factor(ggdata_down$Term, levels = rev(ggdata_down$Term))


ggdata <- bind_rows(ggdata_up, ggdata_down) %>%
  mutate(`% overrepresented` = (Significant-Expected)/Annotated)

gg <- ggplot(ggdata,
  aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = `% overrepresen

  expand_limits(y = 1) +
  geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
    linetype = c("dotted", "longdash", "solid", "dotted", "longdash", "solid"),
    colour = "black",
    size = c(0.5, 1.5, 3, 0.5, 1.5, 3)) +
  geom_point(shape = 21) +
  scale_size(range = c(2.5,12.5)) +
  scale_fill_continuous(low = 'royalblue', high = 'red4') +

  xlab('') +
  ylab('Enrichment score') +
  labs(
    title = paste('GO Biological processes upregulated in contrast: ', names(contrasts[i])),
    subtitle = paste('Top ',ntop,' terms ordered by weighted Fisher p-value'),
    caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +

  theme_bw(base_size = 24) +
  theme(
    legend.position = 'right',
    legend.background = element_rect(),
    plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust = 0.5),
    plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
    plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),

    axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
    axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
    axis.title = element_text(size = 12, face = 'bold'),
    axis.title.x = element_text(size = 12, face = 'bold'),
    axis.title.y = element_text(size = 12, face = 'bold'),
    axis.line = element_line(colour = 'black'),

    #Legend
    legend.key = element_blank(), # removes the border
    legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
    legend.text = element_text(size = 14, face = "bold"), # Text size
```

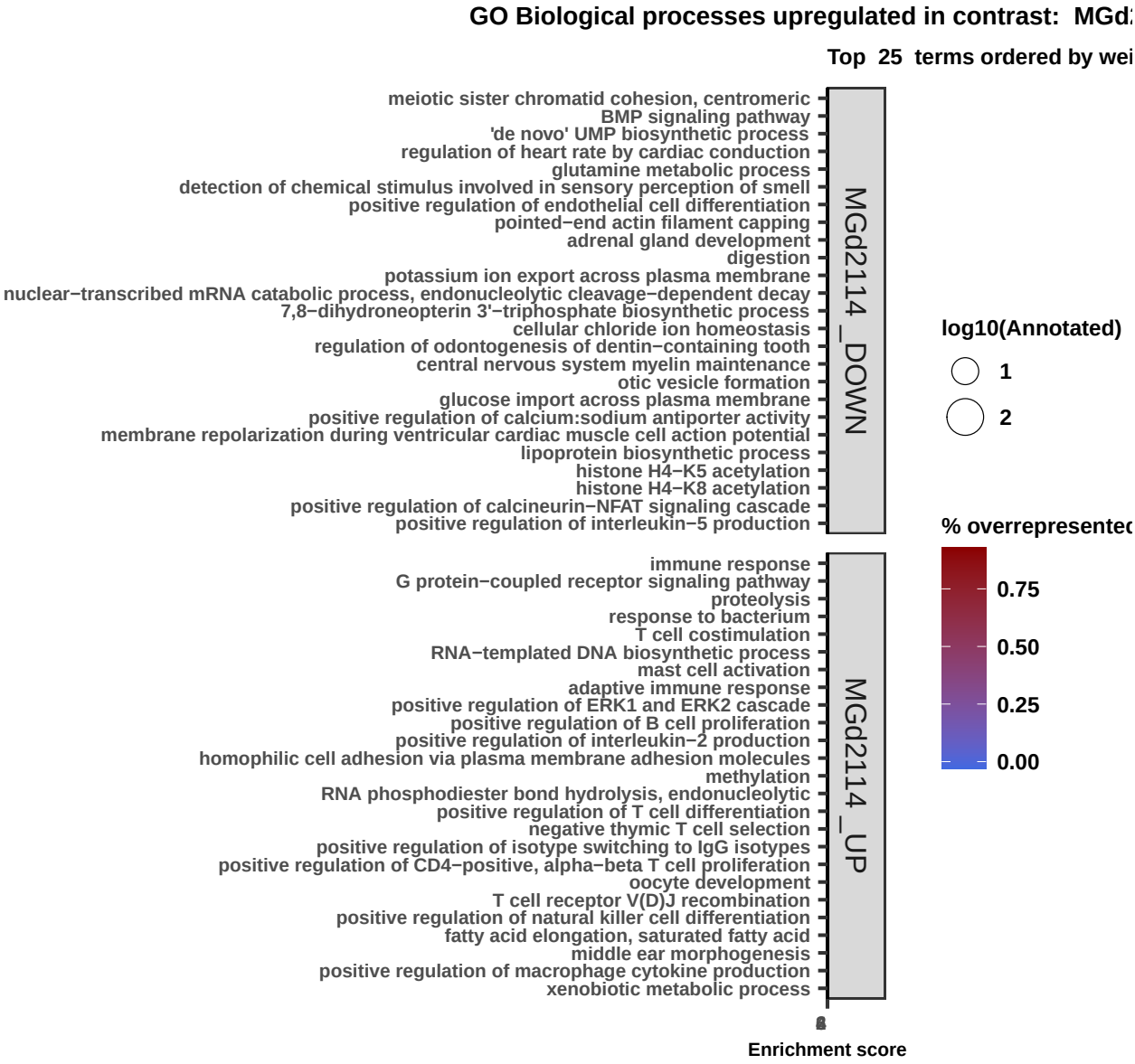
```
      title = element_text(size = 14, face = "bold")) +

    facet_grid(rows = vars(enrichment), scales = "free") +
    coord_flip()

  print(gg)

  ggsave(gg,
         device = 'png',
         filename =  paste("~/Documents/GitHub/CanarySeq/GOanalysis/GO_VIZ_", names(contrasts)[i], ".png
  )

  # save GOdata_up and down for i^th contrast at last j cutoff threshold
  save(GOdata_up, GOdata_down, DEgenes_up, DEgenes_down,
       file = paste("~/Documents/GitHub/CanarySeq/GOanalysis/GOdataUPDN_", names(contrasts)[i], ".RData
       )

}
```

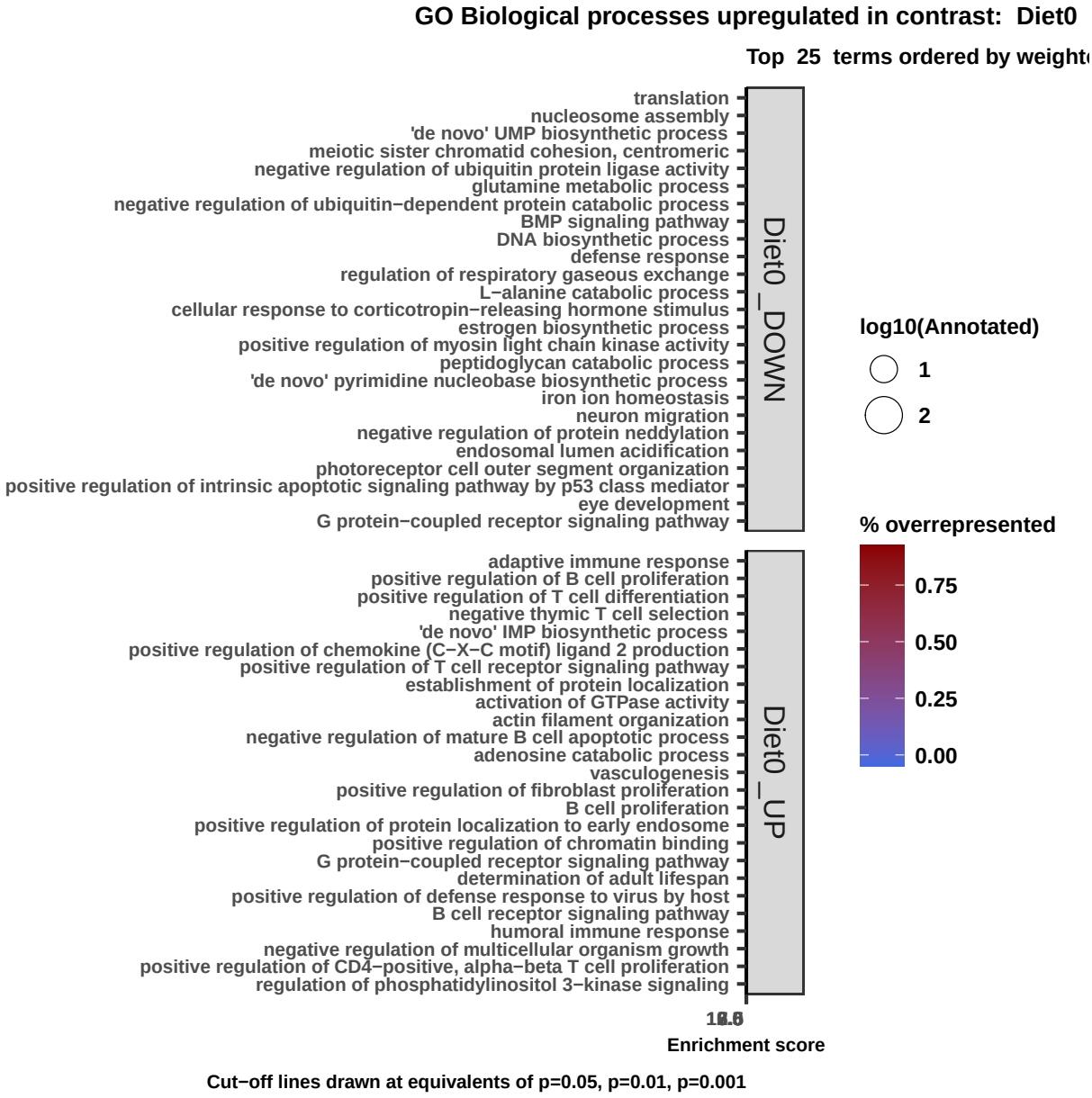

GO Biological processes upregulated in contrast: MGLpost
Top 25 terms ordered by weight
translation
RNA phosphodiester bond hydrolysis, endonucleolytic
ribosomal small subunit biogenesis
DNA integration
cytoplasmic translation
RNA−templated DNA biosynthetic process
DNA recombination
establishment of integrated proviral latency
viral genome integration into host DNA
viral entry into host cell
DNA repair
DNA duplex unwinding
detection of chemical stimulus involved in sensory perception of smell
regulation of catalytic activity
negative regulation of ubiquitin protein ligase activity
rRNA processing
G protein−coupled receptor signaling pathway
embryonic retina morphogenesis in camera−type eye
sensory perception of umami taste
gamma−aminobutyric acid receptor clustering
planar cell polarity pathway involved in neural tube closure
TOR signaling
social behavior
visual perception
proteolysis
MGLpost0_DOWN
positive regulation of interleukin−6 production
positive regulation of type I interferon production
positive regulation of macrophage cytokine production
long−term memory
positive regulation of chemokine (C−X−C motif) ligand 2 production
positive regulation of chemokine production
inflammatory response
positive regulation of tumor necrosis factor production
response to toxic substance
negative regulation of viral genome replication
positive regulation of inflammatory response
negative regulation of cytokine production involved in inflammatory response
toll−like receptor 7 signaling pathway
arginine biosynthetic process via ornithine
signal transduction
MyD88−dependent toll−like receptor signaling pathway
negative regulation of double−strand break repair via homologous recombination
immune response
proteolysis
defense response to virus
adaptive immune response
positive regulation of NF−kappaB transcription factor activity
regulation of smooth muscle cell proliferation
I−kappaB phosphorylation
potassium ion import across plasma membrane
MGLpost0_UP
20
Enrichment score
% overrepresented
0.8
0.6
0.4
0.2
0.0
log10(Annotated)
1
2
3
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001

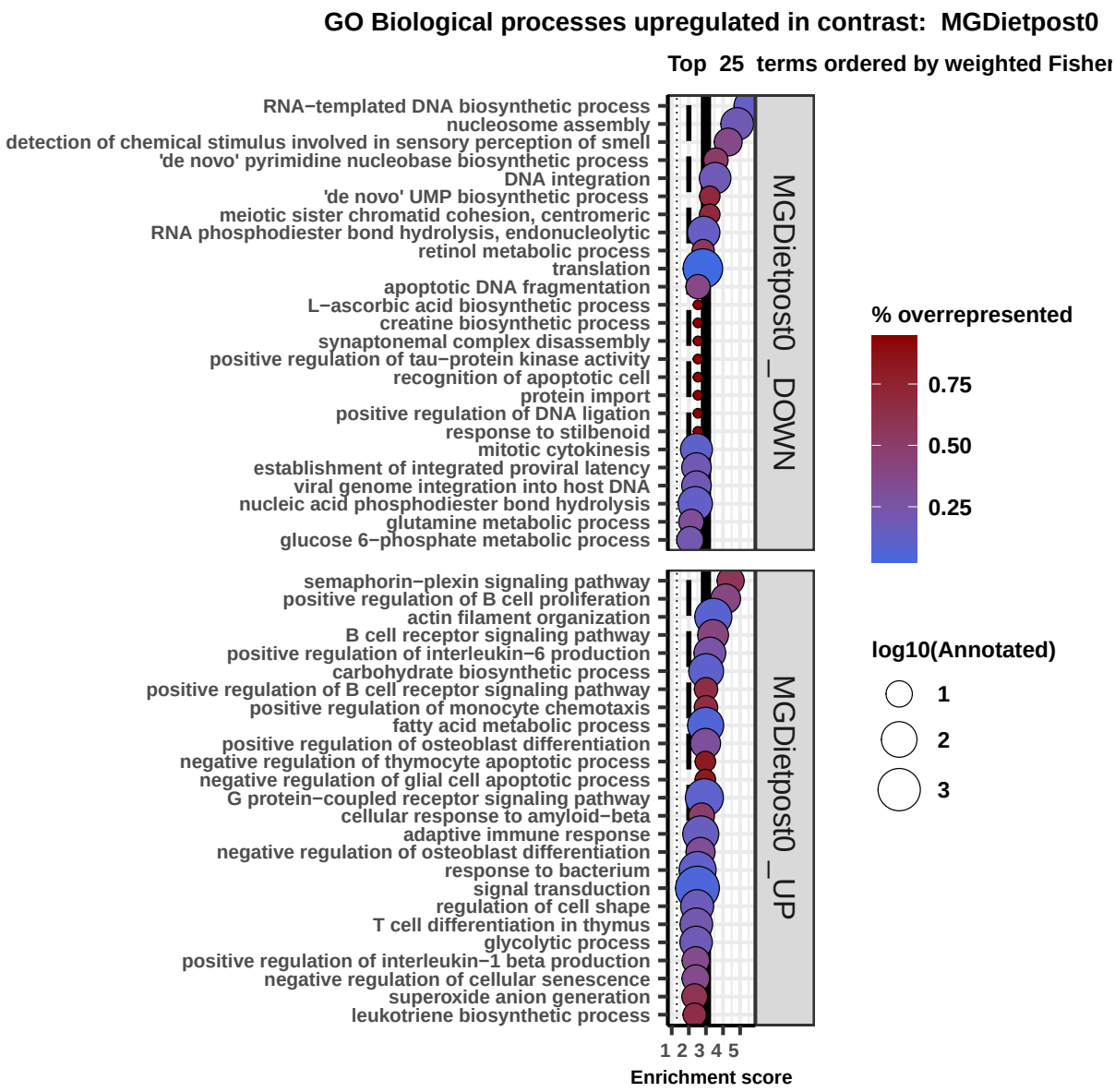

GO Biological processes upregulated in contrast: MGpost0
Top 25 terms ordered by weighted Fisher
translation
DNA integration
RNA phosphodiester bond hydrolysis, endonucleolytic
RNA−templated DNA biosynthetic process
viral genome integration into host DNA
establishment of integrated proviral latency
viral entry into host cell
ribosomal small subunit biogenesis
detection of chemical stimulus involved in sensory perception of smell
DNA recombination
G protein−coupled receptor signaling pathway
regulation of catalytic activity
proteolysis
negative regulation of ubiquitin protein ligase activity
rRNA processing
negative regulation of protein neddylation
nucleic acid phosphodiester bond hydrolysis
DNA duplex unwinding
visual perception
ion transport
negative regulation of ubiquitin−dependent protein catabolic process
social behavior
synaptonemal complex assembly
inositol phosphate−mediated signaling
regulation of protein localization to synapse
MGpost0_DOWN
signal transduction
inflammatory response
positive regulation of interleukin−6 production
MyD88−dependent toll−like receptor signaling pathway
fatty acid metabolic process
positive regulation of macrophage cytokine production
integrin−mediated signaling pathway
positive regulation of inflammatory response
immune response
arginine biosynthetic process via ornithine
positive regulation of chemokine production
positive regulation of type I interferon production
positive regulation of nitric oxide biosynthetic process
positive regulation of tyrosine phosphorylation of STAT protein
negative regulation of multicellular organism growth
semaphorin−plexin signaling pathway
positive regulation of I−kappaB kinase/NF−kappaB signaling
positive regulation of NIK/NF−kappaB signaling
defense response to virus
positive regulation of epithelial to mesenchymal transition
brown fat cell differentiation
response to muscle activity
interleukin−27−mediated signaling pathway
fructose 6−phosphate metabolic process
DNA protection
MGpost0_UP
0 5 10 15 20
Enrichment score
log10(Annotated)
1
2
3
% overrepresented
0.8
0.6
0.4
0.2
0.0
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001

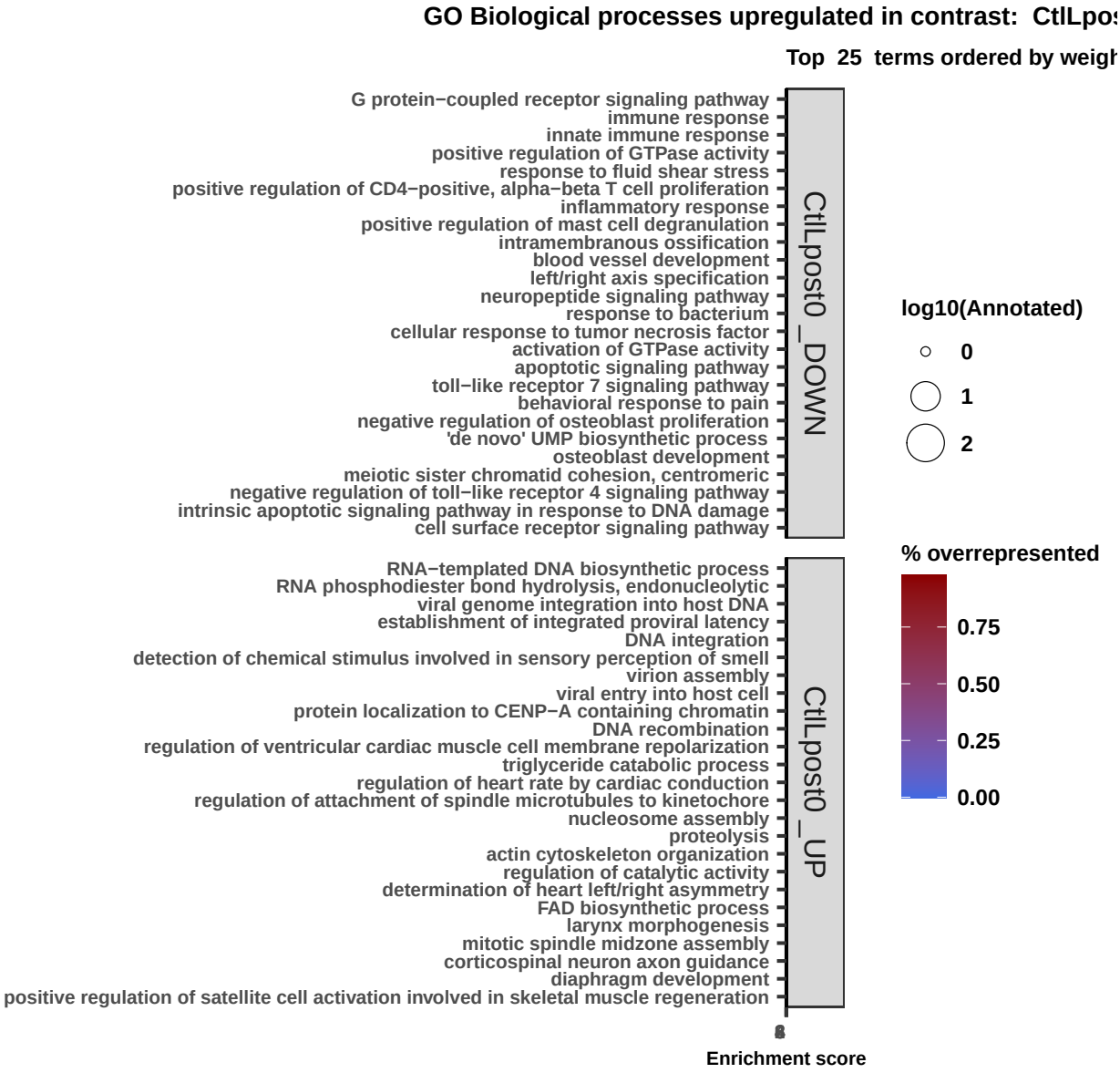

GO Biological processes upregulated in contrast: MGd140
Top 25 terms ordered by weighted Fis
translation
RNA−templated DNA biosynthetic process
RNA phosphodiester bond hydrolysis, endonucleolytic
DNA recombination
immune response
potassium ion transmembrane transport
ribosomal small subunit biogenesis
G protein−coupled receptor signaling pathway
DNA integration
retina layer formation
protein insertion into mitochondrial inner membrane
detection of light stimulus involved in visual perception
regulation of sensory perception of pain
smoothened signaling pathway
T cell costimulation
mitochondrial cytochrome c oxidase assembly
tRNA 5'−leader removal
positive regulation of interleukin−4 production
outer dynein arm assembly
proteolysis
viral entry into host cell
heme catabolic process
L−alanine catabolic process
heme oxidation
creatine biosynthetic process
MGd140_DOWN
positive regulation of vascular associated smooth muscle cell proliferation
inflammatory response
positive regulation of nitric oxide biosynthetic process
positive regulation of ERK1 and ERK2 cascade
negative regulation of viral genome replication
thymus development
'de novo' pyrimidine nucleobase biosynthetic process
positive regulation of astrocyte differentiation
negative regulation of interleukin−17 production
'de novo' UMP biosynthetic process
meiotic sister chromatid cohesion, centromeric
defense response to bacterium
heart morphogenesis
innate immune response
positive regulation of miRNA transcription
BMP signaling pathway
regulation of actin cytoskeleton reorganization
cellular response to prostaglandin E stimulus
positive regulation of NLRP3 inflammasome complex assembly
regulation of myeloid cell differentiation
sulfate transport
collagen biosynthetic process
regulation of kidney size
positive regulation of proteoglycan biosynthetic process
response to lithium ion
MGd140_UP
5
10
15
Enrichment score
log10(Annotated)
1
2
% overrepresented
0.75
0.50
0.25
0.00
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001

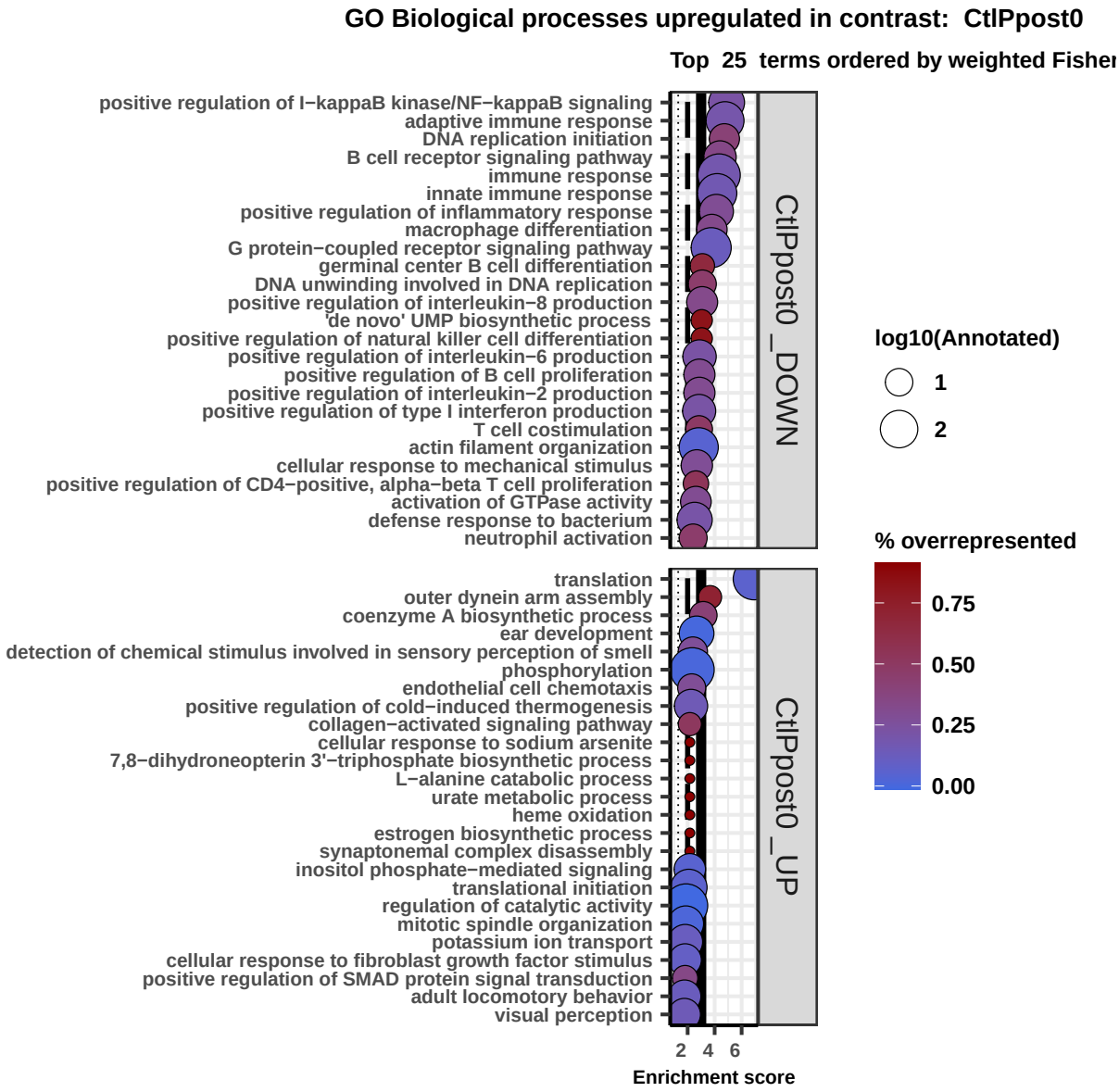

GO Biological processes upregulated in contrast: MGPpost
Top 25 terms ordered by weighte
translation
RNA−templated DNA biosynthetic process
DNA recombination
RNA phosphodiester bond hydrolysis, endonucleolytic
detection of chemical stimulus involved in sensory perception of smell
DNA integration
viral genome integration into host DNA
establishment of integrated proviral latency
viral entry into host cell
synaptonemal complex assembly
G protein−coupled receptor signaling pathway
UMP catabolic process
negative regulation of protein neddylation
dCMP catabolic process
rRNA processing
positive regulation of sequestering of triglyceride
dUMP catabolic process
proteolysis
mitochondrial electron transport, NADH to ubiquinone
ion transport
tRNA 5'−leader removal
nucleic acid phosphodiester bond hydrolysis
hydrogen peroxide catabolic process
pointed−end actin filament capping
arginine catabolic process
MGPpost0_DOWN
inflammatory response
positive regulation of miRNA transcription
innate immune response
defense response
positive regulation of macrophage cytokine production
positive regulation of vascular associated smooth muscle cell proliferation
positive regulation of interleukin−6 production
positive regulation of nitric oxide biosynthetic process
negative regulation of toll−like receptor 4 signaling pathway
arginine biosynthetic process via ornithine
MyD88−dependent toll−like receptor signaling pathway
outflow tract morphogenesis
semaphorin−plexin signaling pathway
fatty acid metabolic process
positive regulation of long−term synaptic potentiation
brown fat cell differentiation
positive regulation of chemokine production
response to bacterium
cellular response to tumor necrosis factor
adaptive immune response
activation of cysteine−type endopeptidase activity involved in apoptotic process
neuromuscular process controlling balance
podocyte differentiation
response to muscle activity
DNA protection
MGPpost0_UP
0
Enrichment score
% overrepresented
0.8
0.6
0.4
0.2
0.0
log10(Annotated)
0.5
1.0
1.5
2.0
2.5
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001


GO Biological processes upregulated in contrast: MGd
Top 25 terms ordered by wei
meiotic sister chromatid cohesion, centromeric
BMP signaling pathway
'de novo' UMP biosynthetic process
regulation of heart rate by cardiac conduction
glutamine metabolic process
detection of chemical stimulus involved in sensory perception of smell
positive regulation of endothelial cell differentiation
pointed−end actin filament capping
adrenal gland development
digestion
potassium ion export across plasma membrane
nuclear−transcribed mRNA catabolic process, endonucleolytic cleavage−dependent decay
7,8−dihydroneopterin 3'−triphosphate biosynthetic process
cellular chloride ion homeostasis
regulation of odontogenesis of dentin−containing tooth
central nervous system myelin maintenance
otic vesicle formation
glucose import across plasma membrane
positive regulation of calcium:sodium antiporter activity
membrane repolarization during ventricular cardiac muscle cell action potential
lipoprotein biosynthetic process
histone H4−K5 acetylation
histone H4−K8 acetylation
positive regulation of calcineurin−NFAT signaling cascade
positive regulation of interleukin−5 production
MGd2114 _DOWN
immune response
G protein−coupled receptor signaling pathway
proteolysis
response to bacterium
T cell costimulation
RNA−templated DNA biosynthetic process
mast cell activation
adaptive immune response
positive regulation of ERK1 and ERK2 cascade
positive regulation of B cell proliferation
positive regulation of interleukin−2 production
homophilic cell adhesion via plasma membrane adhesion molecules
methylation
RNA phosphodiester bond hydrolysis, endonucleolytic
positive regulation of T cell differentiation
negative thymic T cell selection
positive regulation of isotype switching to IgG isotypes
positive regulation of CD4−positive, alpha−beta T cell proliferation
oocyte development
T cell receptor V(D)J recombination
positive regulation of natural killer cell differentiation
fatty acid elongation, saturated fatty acid
middle ear morphogenesis
positive regulation of macrophage cytokine production
xenobiotic metabolic process
MGd2114 _UP
Enrichment score
log10(Annotated)
1
2
% overrepresented
0.75
0.50
0.25
0.00
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001


GO Biological processes upregulated in contrast: Diet0
Top 25 terms ordered by weight
translation
nucleosome assembly
'de novo' UMP biosynthetic process
meiotic sister chromatid cohesion, centromeric
negative regulation of ubiquitin protein ligase activity
glutamine metabolic process
negative regulation of ubiquitin−dependent protein catabolic process
BMP signaling pathway
DNA biosynthetic process
defense response
regulation of respiratory gaseous exchange
L−alanine catabolic process
cellular response to corticotropin−releasing hormone stimulus
estrogen biosynthetic process
positive regulation of myosin light chain kinase activity
peptidoglycan catabolic process
'de novo' pyrimidine nucleobase biosynthetic process
iron ion homeostasis
neuron migration
negative regulation of protein neddylation
endosomal lumen acidification
photoreceptor cell outer segment organization
positive regulation of intrinsic apoptotic signaling pathway by p53 class mediator
eye development
G protein−coupled receptor signaling pathway
Diet0_DOWN
adaptive immune response
positive regulation of B cell proliferation
positive regulation of T cell differentiation
negative thymic T cell selection
'de novo' IMP biosynthetic process
positive regulation of chemokine (C−X−C motif) ligand 2 production
positive regulation of T cell receptor signaling pathway
establishment of protein localization
activation of GTPase activity
actin filament organization
negative regulation of mature B cell apoptotic process
adenosine catabolic process
vasculogenesis
positive regulation of fibroblast proliferation
B cell proliferation
positive regulation of protein localization to early endosome
positive regulation of chromatin binding
G protein−coupled receptor signaling pathway
determination of adult lifespan
positive regulation of defense response to virus by host
B cell receptor signaling pathway
humoral immune response
negative regulation of multicellular organism growth
positive regulation of CD4−positive, alpha−beta T cell proliferation
regulation of phosphatidylinositol 3−kinase signaling
Diet0_UP
Enrichment score
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001
log10(Annotated)
1
2
% overrepresented
0.75
0.50
0.25
0.00


GO Biological processes upregulated in contrast: MGDietpost0
Top 25 terms ordered by weighted Fisher
RNA−templated DNA biosynthetic process
nucleosome assembly
detection of chemical stimulus involved in sensory perception of smell
'de novo' pyrimidine nucleobase biosynthetic process
DNA integration
'de novo' UMP biosynthetic process
meiotic sister chromatid cohesion, centromeric
RNA phosphodiester bond hydrolysis, endonucleolytic
retinol metabolic process
translation
apoptotic DNA fragmentation
L−ascorbic acid biosynthetic process
creatine biosynthetic process
synaptonemal complex disassembly
positive regulation of tau−protein kinase activity
recognition of apoptotic cell
protein import
positive regulation of DNA ligation
response to stilbenoid
mitotic cytokinesis
establishment of integrated proviral latency
viral genome integration into host DNA
nucleic acid phosphodiester bond hydrolysis
glutamine metabolic process
glucose 6−phosphate metabolic process
MGDietpost0_DOWN
semaphorin−plexin signaling pathway
positive regulation of B cell proliferation
actin filament organization
B cell receptor signaling pathway
positive regulation of interleukin−6 production
carbohydrate biosynthetic process
positive regulation of B cell receptor signaling pathway
positive regulation of monocyte chemotaxis
fatty acid metabolic process
positive regulation of osteoblast differentiation
negative regulation of thymocyte apoptotic process
negative regulation of glial cell apoptotic process
G protein−coupled receptor signaling pathway
cellular response to amyloid−beta
adaptive immune response
negative regulation of osteoblast differentiation
response to bacterium
signal transduction
regulation of cell shape
T cell differentiation in thymus
glycolytic process
positive regulation of interleukin−1 beta production
negative regulation of cellular senescence
superoxide anion generation
leukotriene biosynthetic process
MGDietpost0_UP
1 2 3 4 5
Enrichment score
% overrepresented
0.75
0.50
0.25
log10(Annotated)
1
2
3
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001


GO Biological processes upregulated in contrast: CtlLpos
Top 25 terms ordered by weigh
G protein−coupled receptor signaling pathway
immune response
innate immune response
positive regulation of GTPase activity
response to fluid shear stress
positive regulation of CD4−positive, alpha−beta T cell proliferation
inflammatory response
positive regulation of mast cell degranulation
intramembranous ossification
blood vessel development
left/right axis specification
neuropeptide signaling pathway
response to bacterium
cellular response to tumor necrosis factor
activation of GTPase activity
apoptotic signaling pathway
toll−like receptor 7 signaling pathway
behavioral response to pain
negative regulation of osteoblast proliferation
'de novo' UMP biosynthetic process
osteoblast development
meiotic sister chromatid cohesion, centromeric
negative regulation of toll−like receptor 4 signaling pathway
intrinsic apoptotic signaling pathway in response to DNA damage
cell surface receptor signaling pathway
CtlLpost0_DOWN
RNA−templated DNA biosynthetic process
RNA phosphodiester bond hydrolysis, endonucleolytic
viral genome integration into host DNA
establishment of integrated proviral latency
DNA integration
detection of chemical stimulus involved in sensory perception of smell
virion assembly
viral entry into host cell
protein localization to CENP−A containing chromatin
DNA recombination
regulation of ventricular cardiac muscle cell membrane repolarization
triglyceride catabolic process
regulation of heart rate by cardiac conduction
regulation of attachment of spindle microtubules to kinetochore
nucleosome assembly
proteolysis
actin cytoskeleton organization
regulation of catalytic activity
determination of heart left/right asymmetry
FAD biosynthetic process
larynx morphogenesis
mitotic spindle midzone assembly
corticospinal neuron axon guidance
diaphragm development
positive regulation of satellite cell activation involved in skeletal muscle regeneration
CtlLpost0_UP
8
Enrichment score
log10(Annotated)
0
1
2
% overrepresented
0.75
0.50
0.25
0.00
Cut−off lines drawn at equivalents of p=0.05, p=0.01, p=0.001




```
# # Run GO enrichment on clustered genes from cluster_chunks.csv
#
# cluster_chunks_GO <- read_csv(here("cluster analysis/cluster_chunks_GO.csv"))
#
#   # Numeric ID values (in edgeR output) of genes whose names match Gene Ontology Estimates with edgeR
#   chunk_matches <- which(toupper(names(canary_geneID2GO_BM)) %in% cluster_chunks_GO$gene)
#             # these names came from `edgeR diff expression.R` file
#
# genesMatched <- unique(toupper(names(canary_geneID2GO_BM))[chunk_matches])
#
# DEgenes <- pull(cluster_chunks_GO, gene)
#
#     # genes names with GO terms and DE.
# DEgeneswithGO <- which(DEgenes %in% genesMatched)
```

```
#
#     # select only genes that are DE and have GO terms
# DEgenesGO <- DEgenes[DEgeneswithGO]
#
#
# all_genes <- factor(as.integer(toupper(cluster_chunks_GO$gene) %in% DEgenesGO)) # create 0 1 factors.
# names(all_genes) <- toupper(cluster_chunks_GO$gene) # add gene names
#
# #Do the GO mapping
# GOdata=new('topGOdata',
#            ontology='BP', # 'BP' = Biological Process. Also tried 'ALL' but doesn't work.
#            allGenes = all_genes, # all 10188 genes that are in MGpost0 data from edgeR. factor=1 for
#            annot = annFUN.gene2GO, # annotation function for custom annotation
#            # gene2GO = canary_geneID2GO_BM # annotation db from eggNOG output.
# # I am thinking about switching this for one filtered to just edgeR genes:
#            gene2GO = canary_geneID2GO_BM_edgeR # annotation db from BiomaRt.
#            # gene2GO = canary_geneID2GO_BM_DE#,
#            # geneSel = DE_genes
#            )
#
# # Now analyze the results:
#
# # define test using the weight01 algorithm (default) with fisher
# weight_fisher_result=runTest(GOdata, algorithm='weight01', statistic='fisher')
#
#
#     # generate a table of results: we can use the GenTable function to generate
#     # a summary table with the results from tests applied to the topGOdata object.
#     allGO=usedGO(GOdata)
#     all_res=GenTable(GOdata,
#                      weightFisher=weight_fisher_result,
#                      orderBy='weightFisher',
#                      topNodes=length(allGO),
#                      numChar = 200)
#
#     #performing BH correction on our p values
#     p.adj=round(p.adjust(all_res$weightFisher,method="BH"),digits = 4)
#
#     # create the file with all the statistics from GO analysis
#     all_res_final=cbind(all_res,p.adj)
#     all_res_final=all_res_final_up[order(all_res_final$p.adj),]  %>%
#       mutate(across(everything(), gsub, pattern = ",", replacement = ";"))
#
#     #get list of significant GO before multiple testing correction
#     results.table.p= all_res_final[which(as.numeric(all_res_final$weightFisher)<=0.1),] %>%
#       mutate(across(everything(), gsub, pattern = ",", replacement = ";"))
#
#
#     #get list of significant GO after multiple testing correction
#     results.table.bh=all_res_final[which(all_res_final$p.adj<=0.1),] %>%
#       mutate(across(everything(), gsub, pattern = ",", replacement = ";"))
#
#
```

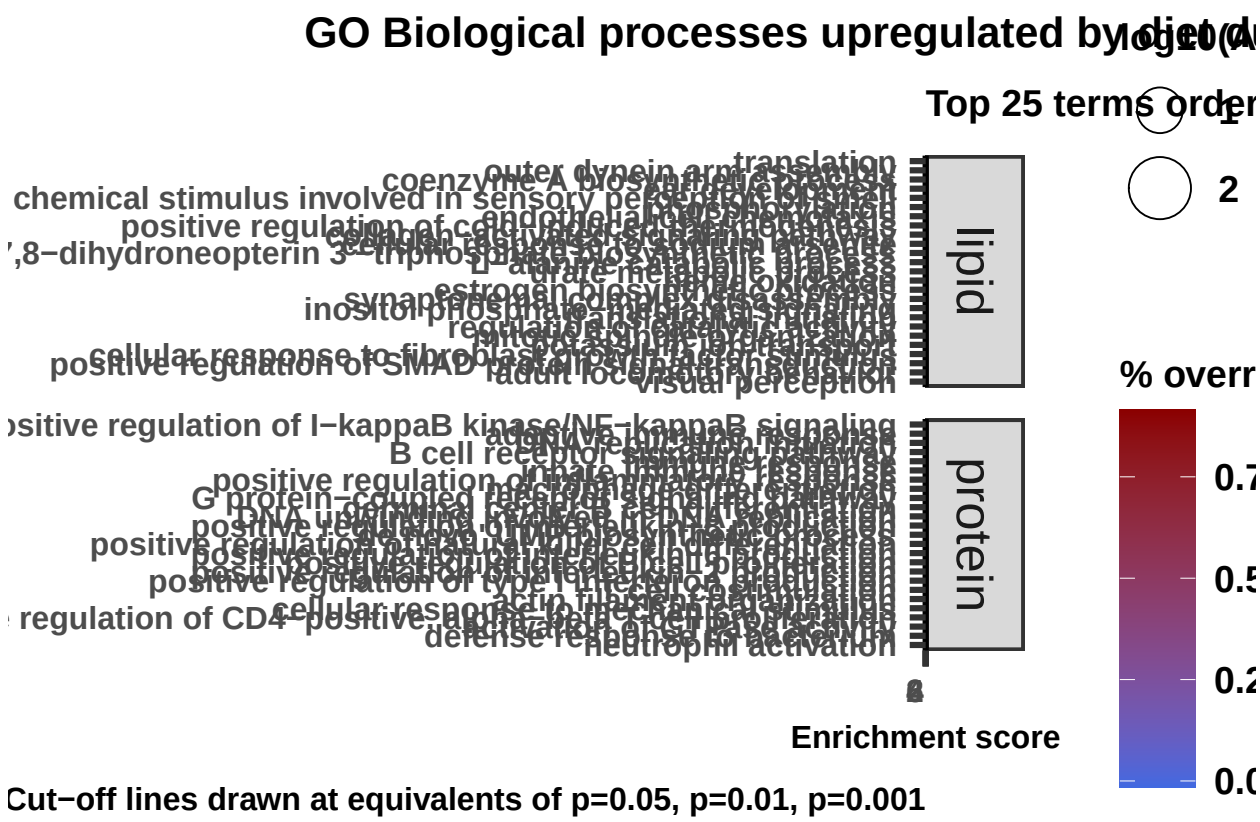
```
#     #save first top 50 ontolgies sorted by adjusted pvalues
#     write.table(all_res_final[1:50,],
#                 file = paste(
#                   "~/Documents/GitHub/CanarySeq/geneAnnotations/geneID2GO_",
#                   "cluster_chunks", ".csv", sep= ""),
#                 # "~/Desktop/canary/summary_topGO_analysis.csv",
#                 sep=",",
#                 quote=FALSE,
#                 row.names=FALSE)
#
#
```

```
library(scales)

ntop <- 25

ggdata_up <- all_res_up[1:ntop,] %>%
  mutate(enrichment = "lipid")
ggdata_up$Term <- factor(ggdata_up$Term, levels = rev(ggdata_up$Term)) # fixes order

ggdata_down <- all_res_down[1:ntop,] %>%
  mutate(enrichment = "protein")
ggdata_down$Term <- factor(ggdata_down$Term, levels = rev(ggdata_down$Term)) # fixes order
#
# gg1 <- ggplot(ggdata_up,
#   # aes(x = Term, y = -log10(as.numeric(weightFisher)), size = -log10(as.numeric(weightFisher)), fill
#   aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = ((Significant-E
#
#   geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
#     linetype = c("dotted", "longdash", "solid"),
#     colour = c("black", "black", "black"),
#     size = c(0.5, 1.5, 3)) +
#
#   expand_limits(y = 1) +
#   geom_point(shape = 21) +
#   scale_size(range = c(2.5,12.5)) +
#   scale_fill_continuous(low = 'royalblue', high = 'red4') +
#
#   xlab('') +
#   ylab('Enrichment score') +
#   labs(
#     title = 'GO Biological processes upregulated in infected + lipid',
#     subtitle = 'Top 25 terms ordered by weighted Fisher p-value',
#     caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +
#
#
#   theme_bw(base_size = 24) +
#   theme(
#     legend.position = 'right',
#     legend.background = element_rect(),
#     plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust=0.5),
#     plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
#     plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),
```

```
#
#     axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
#     axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
#     axis.title = element_text(size = 12, face = 'bold'),
#     axis.title.x = element_text(size = 12, face = 'bold'),
#     axis.title.y = element_text(size = 12, face = 'bold'),
#     axis.line = element_line(colour = 'black'),
#
#     #Legend
#     legend.key = element_blank(), # removes the border
#     legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
#     legend.text = element_text(size = 14, face = "bold"), # Text size
#     title = element_text(size = 14, face = "bold")) +
#
#   coord_flip()
#
#
#
# ggdata_down <- all_res_down[1:ntop,] %>%
#   mutate(enrichment = "protein")
# ggdata_down$Term <- factor(ggdata_down$Term, levels = rev(ggdata_down$Term)) # fixes order
#
# gg2 <- ggplot(ggdata_down,
#   aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = ((Significant-E
#
#   expand_limits(y = 1) +
#   geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
#     linetype = c("dotted", "longdash", "solid"),
#     colour = c("black", "black", "black"),
#     size = c(0.5, 1.5, 3)) +
#   geom_point(shape = 21) +
#   scale_size(range = c(2.5,12.5)) +
#   scale_fill_continuous(low = 'royalblue', high = 'red4') +
#
#   xlab('') +
#   ylab('Enrichment score') +
#   labs(
#     title = 'GO Biological processes upregulated in infected + protein',
#     subtitle = 'Top 25 terms ordered by weighted Fisher p-value',
#     caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +
#
#   theme_bw(base_size = 24) +
#   theme(
#     legend.position = 'right',
#     legend.background = element_rect(),
#     plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust = 0.5),
#     plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
#     plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),
#
#     axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
#     axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
#     axis.title = element_text(size = 12, face = 'bold'),
#     axis.title.x = element_text(size = 12, face = 'bold'),
```

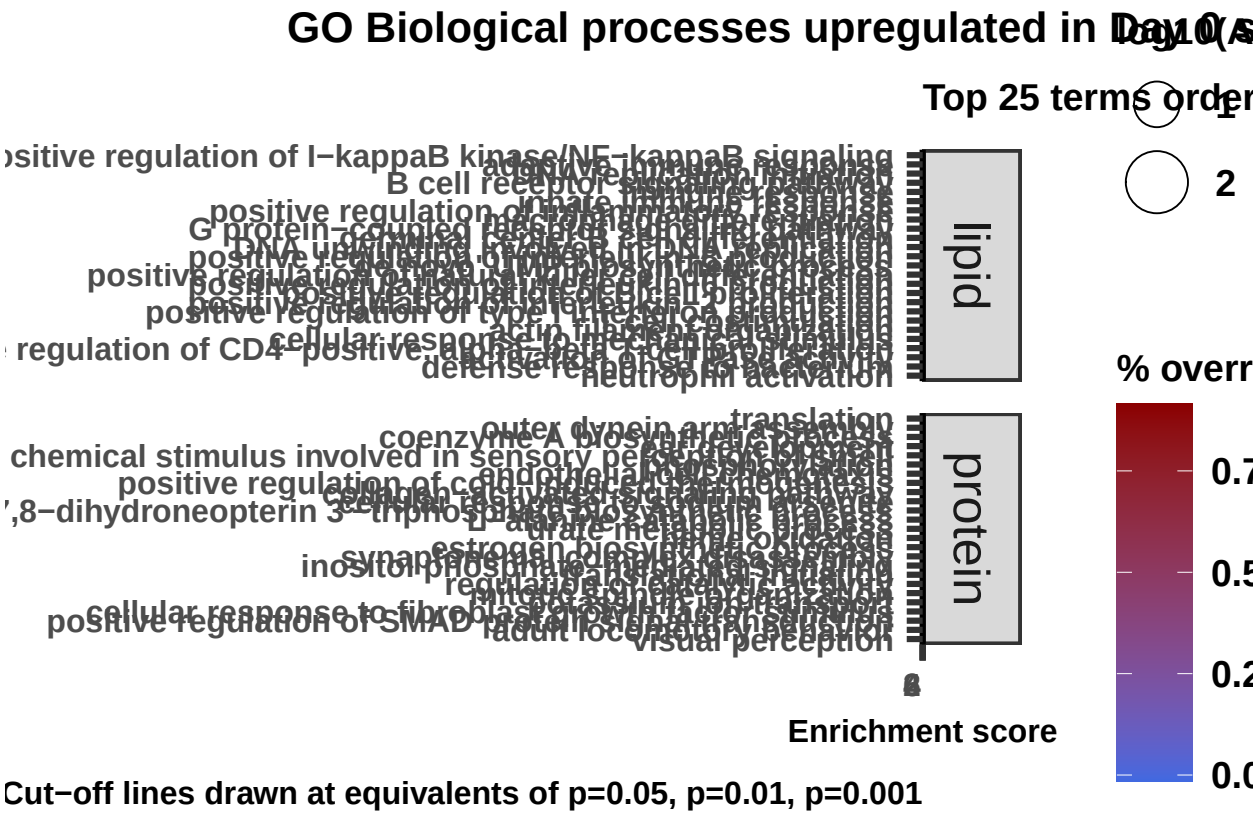
```
#     axis.title.y = element_text(size = 12, face = 'bold'),
#     axis.line = element_line(colour = 'black'),
#
#     #Legend
#     legend.key = element_blank(), # removes the border
#     legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
#     legend.text = element_text(size = 14, face = "bold"), # Text size
#     title = element_text(size = 14, face = "bold")) +
#
#   coord_flip()

# ggarrange(gg1, gg2, ncol=1)

ggdata <- bind_rows(ggdata_up, ggdata_down) %>%
  mutate(`% overrepresented` = (Significant-Expected)/Annotated)

gg <- ggplot(ggdata,
  aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = `% overrepresented

  expand_limits(y = 1) +
  geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
    linetype = c("dotted", "longdash", "solid", "dotted", "longdash", "solid"),
    colour = "black",
    size = c(0.5, 1.5, 3, 0.5, 1.5, 3)) +
  geom_point(shape = 21) +
  scale_size(range = c(2.5,12.5)) +
  scale_fill_continuous(low = 'royalblue', high = 'red4') +

  xlab('') +
  ylab('Enrichment score') +
  labs(
    title = 'GO Biological processes upregulated by diet during infection',
    subtitle = 'Top 25 terms ordered by weighted Fisher p-value',
    caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +

  theme_bw(base_size = 24) +
  theme(
    legend.position = 'right',
    legend.background = element_rect(),
    plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust = 0.5),
    plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
    plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),

    axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
    axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
    axis.title = element_text(size = 12, face = 'bold'),
    axis.title.x = element_text(size = 12, face = 'bold'),
    axis.title.y = element_text(size = 12, face = 'bold'),
    axis.line = element_line(colour = 'black'),

    #Legend
    legend.key = element_blank(), # removes the border
    legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
```

```
    legend.text = element_text(size = 14, face = "bold"), # Text size
    title = element_text(size = 14, face = "bold")) +

  facet_grid(rows = vars(enrichment), scales = "free") +
  coord_flip()


gg
```




```
ntop.c <- 25

ggdata_up.c <- all_res_up[1:ntop.c,] %>%
  # mutate(enrichment = "up.post0")
  mutate(enrichment = "protein", ID = row_number()*pi)

ggdata_up.c$Term <- factor(ggdata_up.c$Term, levels = rev(ggdata_up.c$Term)) # fixes order

ggdata_down.c <- all_res_down[1:ntop.c,] %>%
  # mutate(enrichment = "down.post0")
  mutate(enrichment = "lipid", ID = row_number()*exp(1))

ggdata_down.c$Term <- factor(ggdata_down.c$Term, levels = rev(ggdata_down.c$Term)) # fixes order
```

```r
ggdata.c <- bind_rows(ggdata_up.c, ggdata_down.c) %>%
  mutate(`% overrepresented` = (Significant-Expected)/Annotated) %>%
  group_by(enrichment)

ggdata.c$Term <- factor(ggdata.c$Term, levels = unique(rev(ggdata.c$Term)), ordered=TRUE) # fixes order

gg.c <- ggplot(ggdata.c,
  aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = `% overrepresented

  expand_limits(y = 1) +
  geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
    linetype = c("dotted", "longdash", "solid", "dotted", "longdash", "solid"),
    colour = "black",
    size = c(0.5, 1.5, 3, 0.5, 1.5, 3)) +
  geom_point(shape = 21) +
  scale_size(range = c(2.5,12.5)) +
  scale_fill_continuous(low = 'royalblue', high = 'red4') +

  xlab('') +
  ylab('Enrichment score') +
  labs(
    title = 'GO Biological processes upregulated in Day 0 samples by Diet',
    subtitle = 'Top 25 terms ordered by weighted Fisher p-value',
    caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +

  theme_bw(base_size = 24) +
  theme(
    legend.position = 'right',
    legend.background = element_rect(),
    plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust = 0.5),
    plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
    plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),

    axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
    axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
    axis.title = element_text(size = 12, face = 'bold'),
    axis.title.x = element_text(size = 12, face = 'bold'),
    axis.title.y = element_text(size = 12, face = 'bold'),
    axis.line = element_line(colour = 'black'),

    #Legend
    legend.key = element_blank(), # removes the border
    legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
    legend.text = element_text(size = 14, face = "bold"), # Text size
    title = element_text(size = 14, face = "bold")) +

  facet_grid(rows = vars(enrichment), scales = "free") +
  coord_flip()

gg.c
```

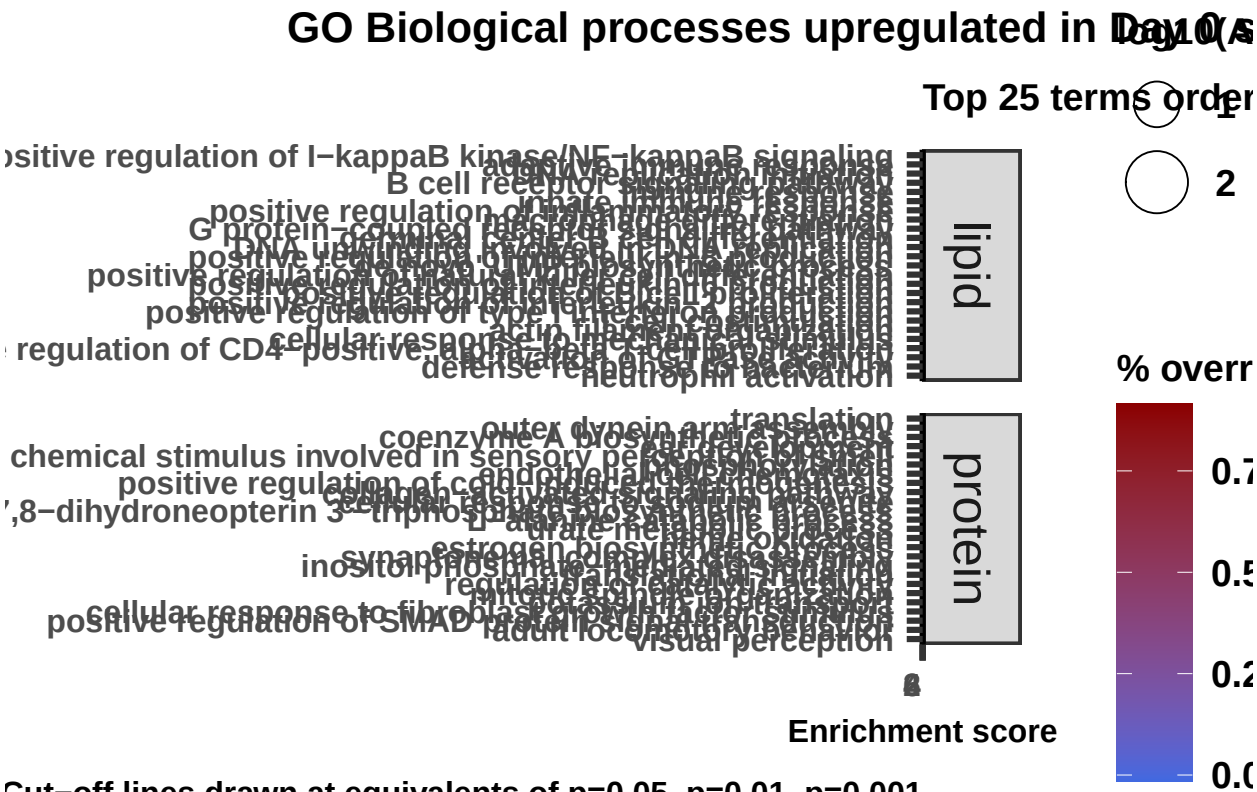



# look at the difference between all time points controls proteins vs lipids.

**NOTE: Diet0 shows really interesting BH correction sigs. Look more at it, see other changes.**

```
ntop.cpost0 <- 25

ggdata_up.cpost0 <- all_res_up[1:ntop.cpost0,] %>%
  # mutate(enrichment = "up.post0")
  mutate(enrichment = "protein")

ggdata_up.cpost0$Term <- factor(ggdata_up.cpost0$Term, levels = rev(ggdata_up.cpost0$Term)) # fixes ord

ggdata_down.cpost0 <- all_res_down[1:ntop.cpost0,] %>%
  # mutate(enrichment = "down.post0")
  mutate(enrichment = "lipid") %>%
  mutate(Term = str_replace(Term, "G protein coupled", " G protein coupled"))

ggdata_down.cpost0$Term <- factor(ggdata_down.cpost0$Term, levels = rev(ggdata_down.cpost0$Term)) # fix


ggdata.cpost0 <- bind_rows(ggdata_up.cpost0, ggdata_down.cpost0) %>%
  mutate(`% overrepresented` = (Significant-Expected)/Annotated) %>%
  group_by(enrichment)
```

```r
ggdata.cpost0$Term <- factor(ggdata.cpost0$Term, levels = unique(rev(ggdata.cpost0$Term)), ordered=TRUE)

gg.cpost0 <- ggplot(ggdata.cpost0,
  aes(x = Term, y = -log10(as.numeric(weightFisher)), size = log10(Annotated), fill = `% overrepresented`)) +

  expand_limits(y = 1) +
  geom_hline(yintercept = c(-log10(0.05), -log10(0.01), -log10(0.001)),
    linetype = c("dotted", "longdash", "solid", "dotted", "longdash", "solid"),
    colour = "black",
    size = c(0.5, 1.5, 3, 0.5, 1.5, 3)) +
  geom_point(shape = 21) +
  scale_size(range = c(2.5,12.5)) +
  scale_fill_continuous(low = 'royalblue', high = 'red4') +

  xlab('') +
  ylab('Enrichment score') +
  labs(
    title = 'GO Biological processes upregulated in Day 0 samples by Diet',
    subtitle = 'Top 25 terms ordered by weighted Fisher p-value',
    caption = 'Cut-off lines drawn at equivalents of p=0.05, p=0.01, p=0.001') +

  theme_bw(base_size = 24) +
  theme(
    legend.position = 'right',
    legend.background = element_rect(),
    plot.title = element_text(angle = 0, size = 16, face = 'bold', vjust = 1, hjust = 0.5),
    plot.subtitle = element_text(angle = 0, size = 14, face = 'bold', vjust = 1),
    plot.caption = element_text(angle = 0, size = 12, face = 'bold', vjust = 1),

    axis.text.x = element_text(angle = 0, size = 12, face = 'bold', hjust = 1.10),
    axis.text.y = element_text(angle = 0, size = 12, face = 'bold', vjust = 0.5),
    axis.title = element_text(size = 12, face = 'bold'),
    axis.title.x = element_text(size = 12, face = 'bold'),
    axis.title.y = element_text(size = 12, face = 'bold'),
    axis.line = element_line(colour = 'black'),

    #Legend
    legend.key = element_blank(), # removes the border
    legend.key.size = unit(1, "cm"), # Sets overall area/size of the legend
    legend.text = element_text(size = 14, face = "bold"), # Text size
    title = element_text(size = 14, face = "bold")) +

  facet_grid(rows = vars(enrichment), scales = "free") +
  coord_flip()

gg.cpost0
```



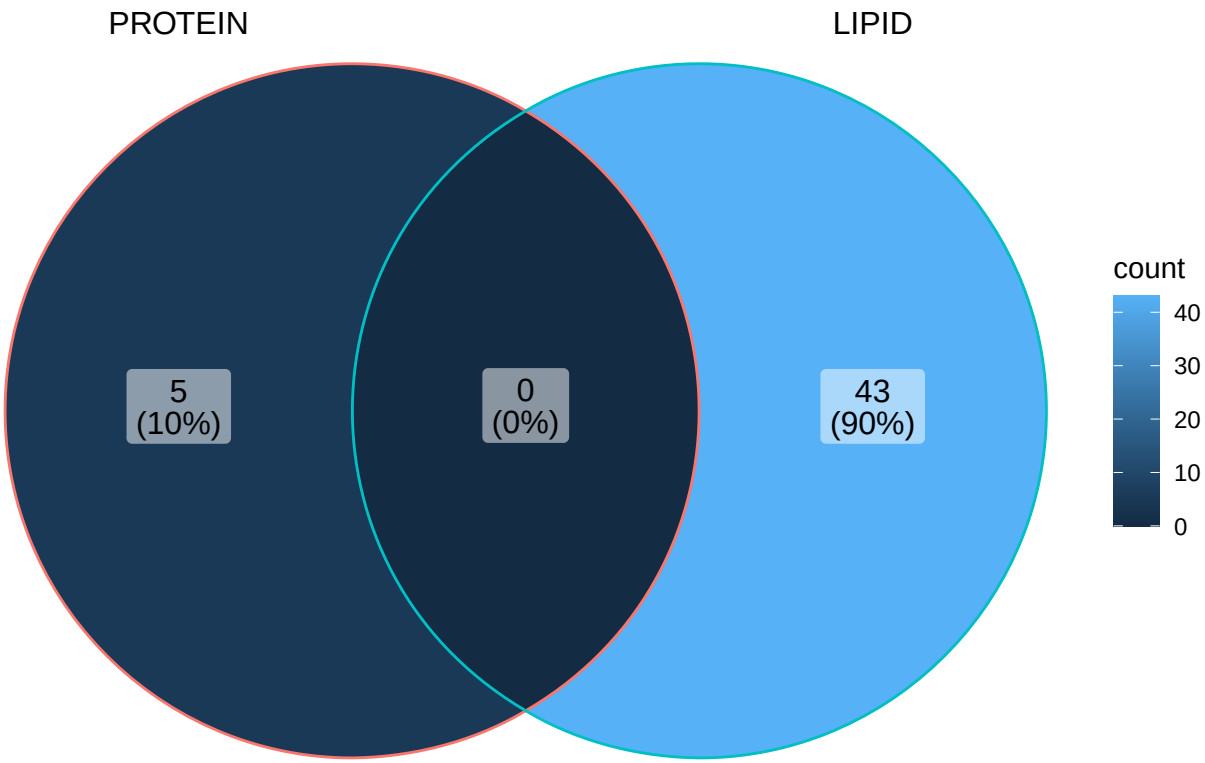

## Get genes all genes with a given GO term

```
# modify for easier formatting
ID2GO_up <- GOdata_up %>%
  genesInTerm() %>%
  enframe()

# find out how many contain GOTERM.
GOTERM <- "GO:0007186" #G Protein Coupled Receptor signaling pathway
GOTERM <- "GO:0002250" # adaptive immune response
GOTERM_ <- "GO:0030890" # positive regulation of B cell proliferation

# # select rows with GOTERM
ID2GO_up %>%
  filter(str_detect(name, GOTERM))
```

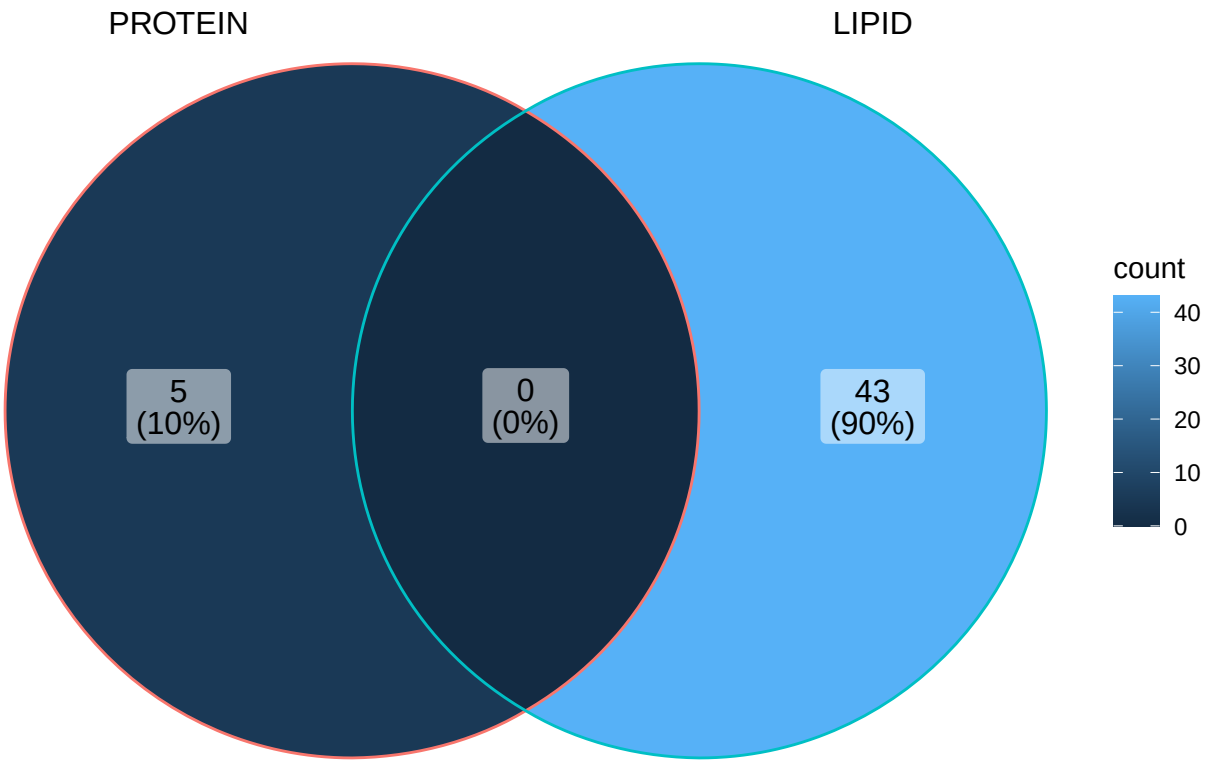
```
## # A tibble: 1 x 2
##   name       value
##   <chr>      <list>
## 1 GO:0002250 <chr [116]>
```

```
GcoupledUP <- genesInTerm(GOdata_up)[GOTERM][[1]][genesInTerm(GOdata_up)[GOTERM][[1]] %in% DEgenes_up]
GcoupledDN <- genesInTerm(GOdata_down)[GOTERM][[1]][genesInTerm(GOdata_down)[GOTERM][[1]] %in% DEgenes_

AdImUP <- genesInTerm(GOdata_up)[GOTERM][[1]][genesInTerm(GOdata_up)[GOTERM][[1]] %in% DEgenes_up]
AdImDN <- genesInTerm(GOdata_down)[GOTERM][[1]][genesInTerm(GOdata_down)[GOTERM][[1]] %in% DEgenes_down

# b cell pos reg (Diet post0)
PosBUP <- genesInTerm(GOdata_up)[GOTERM_][[1]][genesInTerm(GOdata_up)[GOTERM_][[1]] %in% DEgenes_up]


# let's run the same thing but not for a given go term
# reformat genesInTerm as vector
all_up_genesGOTERM <- names(unlist(genesInTerm(GOdata_up)))
all_down_genesGOTERM <- as.vector(unlist(genesInTerm(GOdata_down)))

allUP <- all_up_genesGOTERM[all_up_genesGOTERM %in% DEgenes_up]
allDN <- all_down_genesGOTERM[all_down_genesGOTERM %in% DEgenes_down]

Gcoupled <-
  list(
    PROTEIN = GcoupledUP,
    LIPID = GcoupledDN
  )

AdaptImm <-
  list(
    PROTEIN = AdImUP,
    LIPID = AdImDN
  )

all <-
  list(
    PROTEIN = allUP,
    LIPID = allDN
  )
```

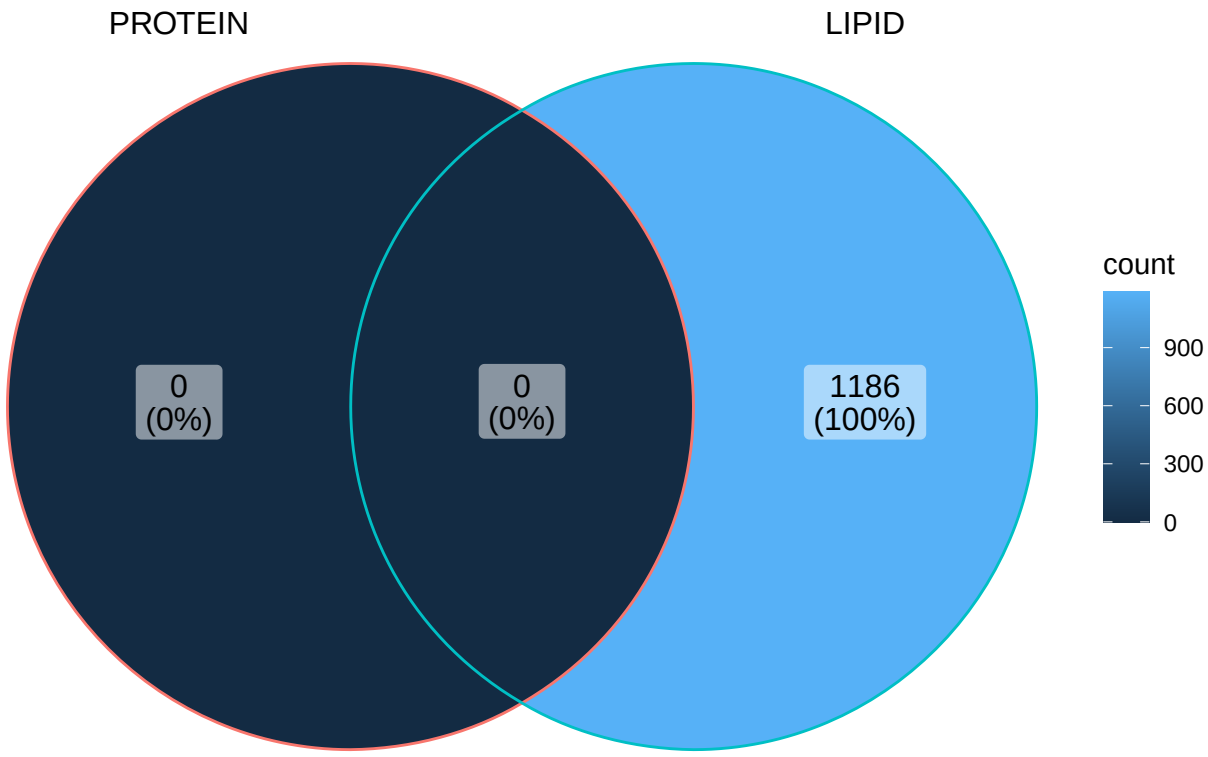
```
# if (!require(devtools)) install.packages("devtools")
# devtools::install_github("gaospecial/ggVennDiagram")
library(ggVennDiagram)

ggVennDiagram(Gcoupled)
```




```
ggVennDiagram(AdaptImm)
```




```
ggVennDiagram(all)
```




```
# ADA
# AGTRAP
# AKAP13
# C3
# CCR10
# CCR6
# CCR7
# DGKH
# DGKZ
# GNG2
# GNG7
# GPR132
# GPR162
# GPR171
# GPR174
# GPR18
# GPR82
# LOC103812291
# LOC103812980
# LOC103812985
# LOC103813169
# LOC103813540
# LOC103815708
# LOC103819762
# LOC103824208
# LOC108964377
```

```
# LOC108964492
# LOC108964818
# LOC115485368
# LOC115485604
# LOC115485605
# P2RY10
# P2RY12
# P2RY6
# PIK3CG
# PIK3R6
# PLCL1
# PLCL2
# PREX1
# PRMT5
# PTK2B
# RAMP2
# RGS1
# RGS4
# RRH
# S1PR2
# S1PR4
# SLC24A4
# SORL1
# TYRO3
# ZDHHC21
```