Functional Gene Enrichment Report

Contents

1. Load Required Libraries

```
library(readxl)
library(clusterProfiler)
library(org.Hs.eg.db)
library(org.Ce.eg.db)
library(enrichplot)
library(ReactomePA)
library(ggplot2)
library(dplyr)
library(ggplot2)
library(dplyr)
library(forcats)
```

2. Input Gene Lists

2.1 Worm Genes

```
# Table 1 is inside the file "List of verified genes.xlsx" as input of the enrichment analysis here
file_path_worm <- "./06-2024/List of verified genes.xlsx"
worm_df <- read_excel(file_path_worm)
wb_col <- grep("WormBase Gene", colnames(worm_df), ignore.case = TRUE, value = TRUE)
worm_df <- worm_df %>% rename(GeneID = all_of(wb_col))
worm_map <- bitr(worm_df$GeneID, fromType = "WORMBASE", toType = "ENTREZID", OrgDb = org.Ce.eg.db)
worm_entrez <- unique(worm_map$ENTREZID)
```

```
# all mapped Entrez IDs
mapped_entrez <- unique(worm_map$ENTREZID)

# Check GO annotations for those Entrez IDs
bp_annot <- AnnotationDbi::select(
  org.Ce.eg.db,
  keys    = mapped_entrez,
```

```
  keytype = "ENTREZID",
  columns = c("GO", "ONTOLOGY")
)

# Keep only BP ontology
bp_annot <- bp_annot %>%
  filter(ONTOLOGY == "BP" & !is.na(GO))

# Entrez IDs that have at least one BP GO term
genes_with_bp <- unique(bp_annot$ENTREZID)

length(genes_with_bp)
```

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

```
# should be 26

# Which 4 mapped genes have NO BP annotation?
no_bp_entrez <- setdiff(mapped_entrez, genes_with_bp)
no_bp_entrez
```

```
## [1] "171859" "188088" "187952" "185888"
```

3.4 C. elegans GO: Biological Process

```
ego_ce_bp <- enrichGO(gene = worm_entrez, OrgDb = org.Ce.eg.db, keyType = "ENTREZID",
                      ont = "BP", pAdjustMethod = "BH", pvalueCutoff = 0.05,
                      qvalueCutoff = 0.2, readable = TRUE)
# dotplot(ego_ce_bp, showCategory = 20) + ggtitle("C. elegans GO Enrichment (Biological Process)")
```

```
# Extract top 20 results ordered by adjusted p-value
# plot_data <- ego_ce_bp@result %>%
#   arrange(p.adjust) %>%
#   slice(1:6) %>%
#   mutate(Description = fct_reorder(Description, -p.adjust)) %>%
#   mutate(log10_padj = -log10(p.adjust))

# Create custom dot plot
# Step 1: Extract and modify the plot data
plot_data <- ego_ce_bp@result %>%
  arrange(p.adjust) %>%
  slice(1:6) %>%
  mutate(log10_padj = -log10(p.adjust))

# Step 2: Manually define your desired order for y-axis
custom_order <- c(
  "calcium ion-regulated exocytosis of neurotransmitter",
  "neuronal dense core vesicle exocytosis",
  "presynaptic dense core vesicle exocytosis",
  "vesicle-mediated transport to the plasma membrane",
```

```
  "dense core granule exocytosis",
  "regulation of protein localization"
)

# Step 3: Convert Description to factor with custom levels
plot_data <- plot_data %>%
  filter(Description %in% custom_order) %>%
  mutate(Description = factor(Description, levels = rev(custom_order)))
# reverse for top-to-bottom

# Step 4: Plot
dotplot_2nd <- ggplot(plot_data, aes(x = log10_padj, y = Description)) +
  geom_point(aes(size = Count, color = GeneRatio), alpha = 0.8) +
  scale_x_continuous(limits = c(1.42, 1.48)) +
  # scale_color_gradient(low = "lightblue", high = "darkblue", name = "GeneRatio") +
  scale_size(range = c(3, 8)) +
  labs(
    x = expression(-log[10]~adjusted~italic(p)),
    y = "GO Term",
    title = "C. elegans GO Enrichment (Biological Process)"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    axis.text.y = element_text(size = 9),
    plot.title = element_text(hjust = 0.5),
    panel.border = element_rect(color = "black", fill = NA, linewidth = 0.8)
  )
# dotplot_2nd #One can find the final output of dotplot
# in manscript figure 2 panel a
```

```
ggsave("./output_plots/dotplot_2nd_C.e.Biological_Process.pdf", dotplot_2nd, dpi = 600)
```

```
ego_ce_cc <- enrichGO(gene = worm_entrez, OrgDb = org.Ce.eg.db, keyType = "ENTREZID",
                      ont = "CC", pAdjustMethod = "BH", pvalueCutoff = 0.05,
                      qvalueCutoff = 0.2, readable = TRUE)
# dotplot(ego_ce_cc, showCategory = 20) + ggtitle("C. elegans
# GO Enrichment (Biological Process)")
```

```
plot_data <- ego_ce_cc@result %>%
  arrange(p.adjust) %>%
  slice(1:4) %>%
  mutate(log10_padj = -log10(p.adjust))

# Step 2: Manually define your desired order for y-axis
# not asked in this case

# Step 3: Convert Description to factor with custom levels
# plot_data <- plot_data %>%
#   filter(Description %in% custom_order) %>%
#   mutate(Description = factor(Description, levels = rev(custom_order)))  # reverse for top-to-bottom

# Step 1: Sort plot_data by p.adjust (increasing -log10 means sort by descending p.adjust)
plot_data <- plot_data %>%
```

```
  arrange(desc(p.adjust)) %>%
  mutate(log10_padj = -log10(p.adjust),
         Description = factor(Description, levels = Description))  # Set y-axis order

# Step 2: Custom plot with border and margins
dotplotCC_2nd <- ggplot(plot_data, aes(x = log10_padj, y = Description)) +
  geom_point(aes(size = Count, color = GeneRatio), alpha = 0.8) +
  scale_x_continuous(limits = c(1.59, 2.03)) +
  # scale_color_gradient(low = "lightblue", high = "darkblue", name = "GeneRatio") +
  scale_size(range = c(3, 8)) +
  labs(
    x = expression(-log[10]~adjusted~italic(p)),
    y = "GO Term",
    title = "C. elegans GO Enrichment (Cellular Compartment)"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    axis.text.y = element_text(size = 9),
    plot.title = element_text(hjust = 0.5),
    panel.border = element_rect(color = "black", fill = NA, linewidth = 0.8),  # Frame
    plot.margin = margin(15, 15, 15, 15)  # Top, Right, Bottom, Left (all in 'pt')
  )

# dotplotCC_2nd # dotplot_2nd #One can find the final
# output of dotplot in manscript figure 2 panel b
```

```
ggsave("./output_plots/dotplot_2nd_C.e.Cellular_Compartment.pdf", dotplotCC_2nd, dpi = 600)
```

4. Export Results (Optional)

```
# write.csv(as.data.frame(ego_bp), "Human_GO_BP.csv", row.names = FALSE)
# write.csv(as.data.frame(ego_mf), "Human_GO_MF.csv", row.names = FALSE)
# write.csv(as.data.frame(ereactome), "Human_Reactome.csv", row.names = FALSE)
write.csv(as.data.frame(ego_ce_bp), "Worm_GO_BP.csv", row.names = FALSE)
```