Comparative functional enrichment analysis between proteins annotated as displaying intensity variation in one of the four ciliary locations and annotated as stable

Konstantin Kahnert

2025-05-13

Load libraries

```
library(tidyverse)
library(clusterProfiler)
library(enrichplot)
library(org.Hs.eg.db)
library(Cairo)
library(svglite)
library(VennDiagram)
library(grid)
library(simplifyEnrichment)
library(readxl)
```

Define functions

```
# map_cluster_number:
#   - x: enrichResult or compareClusterResult (from clusterProfiler)
#   - df: simplifyGO result data.frame with columns ID (term IDs) and Cluster_num
#   - comp: if TRUE, use x@compareClusterResult, else x@result
# Returns a data.frame of all enrich columns + term_size + Cluster_num
map_cluster_number <- function(x,
                               df,
                               comp = FALSE) {
  ## 1. Standardize the input DF's column names
  colnames(df) <- c("ID", "Cluster_num")

  ## 2. Select the correct slot from the clusterProfiler object
  if (comp) {
    # from compareClusterResult
    results <- x@compareClusterResult
  } else {
    # from a single enrichResult
    results <- x@result
```

```
  }

  ## 3. Get term_size (the numerator of the "BgRatio" string)
  ##    e.g. "12/21273" → 12
  results$term_size <- as.numeric(
    sapply(strsplit(results$BgRatio, "/"),
           function(parts) as.numeric(parts[1]))
  )

  ## 4. Merge the enrichment results with the cluster assignments
  merged_results <- merge(
    results,
    df,
    by    = "ID",
    all.x = TRUE
  )

  ## 5. Return the processed data frame
  return(merged_results)
}


# Function to map Ensembl IDs to gene symbols for a single row
map_geneID_to_symbol <- function(geneID_str) {
  geneIDs <- unlist(strsplit(geneID_str, "/"))
  mapped_ids <- bitr(geneIDs,
                     fromType = "ENSEMBL",
                     toType = "SYMBOL",
                     OrgDb = org.Hs.eg.db)

  # drop duplicates in ensembl column and keep first occurence
  mapped_ids <- mapped_ids[!duplicated(mapped_ids$ENSEMBL),]

  gene_symbols <- mapped_ids$SYMBOL
  return(paste(gene_symbols, collapse = "/"))
}


# Function to map Ensembl IDs to gene symbols for a single row
map_geneID_to_name <- function(geneIDs) {
  mapped_ids <- bitr(geneIDs,
                     fromType = "ENSEMBL",
                     toType = c("GENENAME", "SYMBOL"),
                     OrgDb = org.Hs.eg.db)

  # drop duplicates in ensembl column and keep first occurence
  mapped_ids <- mapped_ids[!duplicated(mapped_ids$ENSEMBL),]

  return(mapped_ids)
}
```

Data loading and preprocessing

Set input and output paths

```
# set input and output paths
in_path <- "/mnt/Data/Projects/Cilia/revision/NonRestricted/data/"
out_path <- "/mnt/Data/Projects/Cilia/revision/NonRestricted/analysis/GO_BP/"
```

Load data

```
# Load the data
df_all <- read.delim(paste0(in_path, "All_files_combined_as_cytoscape_input.csv"), sep = "\t", header =
head(df_all)
```

```
##                 BasalBody_ BasalBody_ASC52telo
## ENSG00000001497       True               False
## ENSG00000002330      False               False
## ENSG00000002549      False               False
## ENSG00000003249      False               False
## ENSG00000003756      False               False
## ENSG00000004766       True               False
##                 BasalBody_hTERT_RPE1_serum_starved BasalBody_RPTEC_TERT1
## ENSG00000001497                               True                False
## ENSG00000002330                              False                False
## ENSG00000002549                              False                False
## ENSG00000003249                              False                False
## ENSG00000003756                              False                False
## ENSG00000004766                               True                 True
##                 PrimaryCilia_ PrimaryCilia_ASC52telo
## ENSG00000001497          True                 False
## ENSG00000002330         False                 False
## ENSG00000002549         False                 False
## ENSG00000003249         False                 False
## ENSG00000003756         False                 False
## ENSG00000004766          True                  True
##                 PrimaryCilia_hTERT_RPE1_serum_starved PrimaryCilia_RPTEC_TERT1
## ENSG00000001497                                 True                     False
## ENSG00000002330                                False                     False
## ENSG00000002549                                False                     False
## ENSG00000003249                                False                     False
## ENSG00000003756                                False                     False
## ENSG00000004766                                 True                      True
##                 PrimaryCiliaTip_ PrimaryCiliaTip_ASC52telo
## ENSG00000001497            False                     False
## ENSG00000002330            False                     False
## ENSG00000002549            False                     False
## ENSG00000003249            False                     False
## ENSG00000003756            False                     False
## ENSG00000004766            False                     False
##                 PrimaryCiliaTip_hTERT_RPE1_serum_starved
```

```
## ENSG00000001497                                    False
## ENSG00000002330                                    False
## ENSG00000002549                                    False
## ENSG00000003249                                    False
## ENSG00000003756                                    False
## ENSG00000004766                                    False
##                 PrimaryCiliaTip_RPTEC_TERT1 PrimaryCiliaTZ_
## ENSG00000001497                       False            True
## ENSG00000002330                       False           False
## ENSG00000002549                       False           False
## ENSG00000003249                       False           False
## ENSG00000003756                       False           False
## ENSG00000004766                       False           False
##                 PrimaryCiliaTZ_ASC52telo
## ENSG00000001497                    False
## ENSG00000002330                    False
## ENSG00000002549                    False
## ENSG00000003249                    False
## ENSG00000003756                    False
## ENSG00000004766                    False
##                 PrimaryCiliaTZ_hTERT_RPE1_serum_starved
## ENSG00000001497                                    True
## ENSG00000002330                                   False
## ENSG00000002549                                   False
## ENSG00000003249                                   False
## ENSG00000003756                                   False
## ENSG00000004766                                   False
##                 PrimaryCiliaTZ_RPTEC_TERT1 Nucleus Mitotic Membrane Cytoplasm
## ENSG00000001497                      False    True   False    False      True
## ENSG00000002330                      False   False   False    False      True
## ENSG00000002549                      False   False   False    False      True
## ENSG00000003249                      False    True   False    False     False
## ENSG00000003756                      False    True   False     True      True
## ENSG00000004766                      False   False   False     True      True
##                 BasalBody_num PrimaryCilia_num PrimaryCiliaTip_num
## ENSG00000001497             1                1                   0
## ENSG00000002330             0                0                   0
## ENSG00000002549             0                0                   0
## ENSG00000003249             0                0                   0
## ENSG00000003756             0                0                   0
## ENSG00000004766             2                3                   0
##                 PrimaryCiliaTZ_num ASC52telo hTERT_RPE1_serum_starved
## ENSG00000001497                  1     False                      True
## ENSG00000002330                  0     False                     False
## ENSG00000002549                  0     False                     False
## ENSG00000003249                  0     False                     False
## ENSG00000003756                  0     False                     False
## ENSG00000004766                  0      True                      True
##                 RPTEC_TERT1
## ENSG00000001497       False
## ENSG00000002330       False
## ENSG00000002549       False
## ENSG00000003249       False
## ENSG00000003756       False
```

```
## ENSG00000004766        True
```

Map gene names and symbols

```
# Identify columns that contain "num" in their names
num_columns <- grep("num", names(df_all), value = TRUE)

# Convert all other columns from string to logical
df_all <- df_all %>%
  mutate(across(-all_of(num_columns), ~ as.logical(.)))

# map gene IDs to gene names
mapped_ids <- map_geneID_to_name(rownames(df_all))

# index by Ensembl
rownames(mapped_ids) <- mapped_ids$ENSEMBL

# pull out exactly one entry per row of df_all
df_all$GeneSymbol <- mapped_ids[ rownames(df_all), "SYMBOL"   ]
df_all$GeneName   <- mapped_ids[ rownames(df_all), "GENENAME" ]

# add rownames as column and reset index
df_all$Ensembl_ID <- rownames(df_all)
rownames(df_all) <- NULL

# reorder columns to have Ensembl_ID first, Symbol and Gene name first and then all other columns
df_all <- df_all %>%  dplyr::select(Ensembl_ID, GeneSymbol, GeneName, everything())
```

Split data by location

```
# Perform filtering again
df_bb <- df_all %>% filter(BasalBody_ == TRUE)
df_pc <- df_all %>% filter(PrimaryCilia_ == TRUE)
df_tip <- df_all %>% filter(PrimaryCiliaTip_ == TRUE)
df_tz <- df_all %>% filter(PrimaryCiliaTZ_ == TRUE)

# filter gene_id by location
gene_id_all <- df_all$Ensembl_ID
gene_id_bb <- df_bb$Ensembl_ID
gene_id_pc <- df_pc$Ensembl_ID
gene_id_tip <- df_tip$Ensembl_ID
gene_id_tz <- df_tz$Ensembl_ID
```

Combine bb & tz and pc & tip

```
# combine pc and tip and tz
df_pc_tip_tz <- rbind(df_pc, df_tip, df_tz)
```

```r
# drop duplicates based on Ensembl_ID column
df_pc_tip_tz <- df_pc_tip_tz[!duplicated(df_pc_tip_tz$Ensembl_ID), ]
gene_id_pc_tip_tz <- df_pc_tip_tz$Ensembl_ID
```

Compare biological themes for different locations with variable proteins

Define variable genes

```r
# Load the data from the Excel file
df <- read_excel("/mnt/Data/Projects/Cilia/revision/Filtered_Staining_List_combined_exploded.xlsx")

# Identify Ensembl IDs for each category by searching substrings in the "Annotation (Intensity)" column
tip_variable <- unique(df$`Ensembl id`[grepl("Primary cilium tip", df$`Annotation (Intensity)`)])
pc_variable  <- unique(df$`Ensembl id`[grepl("Primary cilium",     df$`Annotation (Intensity)`)])
tz_variable  <- unique(df$`Ensembl id`[grepl("Primary cilium transition zone", df$`Annotation (Intensity`

# Combine all variable genes and remove duplicates
variable_genes <- unique(c(pc_variable, tip_variable, tz_variable))

# Identify any non-variable genes
non_variable <- setdiff(gene_id_pc_tip_tz, variable_genes)
```

Map as columns to df_all and save as csv file

```r
# add "Intensity variation" columns to df_all with true if in variable genes, false if in non_variable
df_all <- df_all %>%
  mutate(
    Intensity_variation = case_when(
      Ensembl_ID %in% variable_genes ~ TRUE,
      Ensembl_ID %in% non_variable   ~ FALSE,
      TRUE                           ~ NA   # catch-all case for any other Ensembl IDs
    )
  )

# write to file
write.csv(df_all, file = paste0(out_path, "NonRestricted_all_proteins_enrichment_input_mapping.csv"))
```

GO BP enrichment analysis

```r
# Split data by location
input_genes <- list(
  non_variable = non_variable,
  variable = variable_genes
)
```

```r
# check length of each list
lapply(input_genes, length)
```

```
## $non_variable
## [1] 109
##
## $variable
## [1] 409
```

```r
comp <- compareCluster(geneCluster = input_genes,
                       fun = "enrichGO",
                       OrgDb = org.Hs.eg.db,
                       keyType       = 'ENSEMBL',
                       ont           = "BP",
                       pAdjustMethod = "BH",
                       pvalueCutoff  = 0.01,
                       qvalueCutoff  = 0.01)
```

Dot plot of all enriched terms

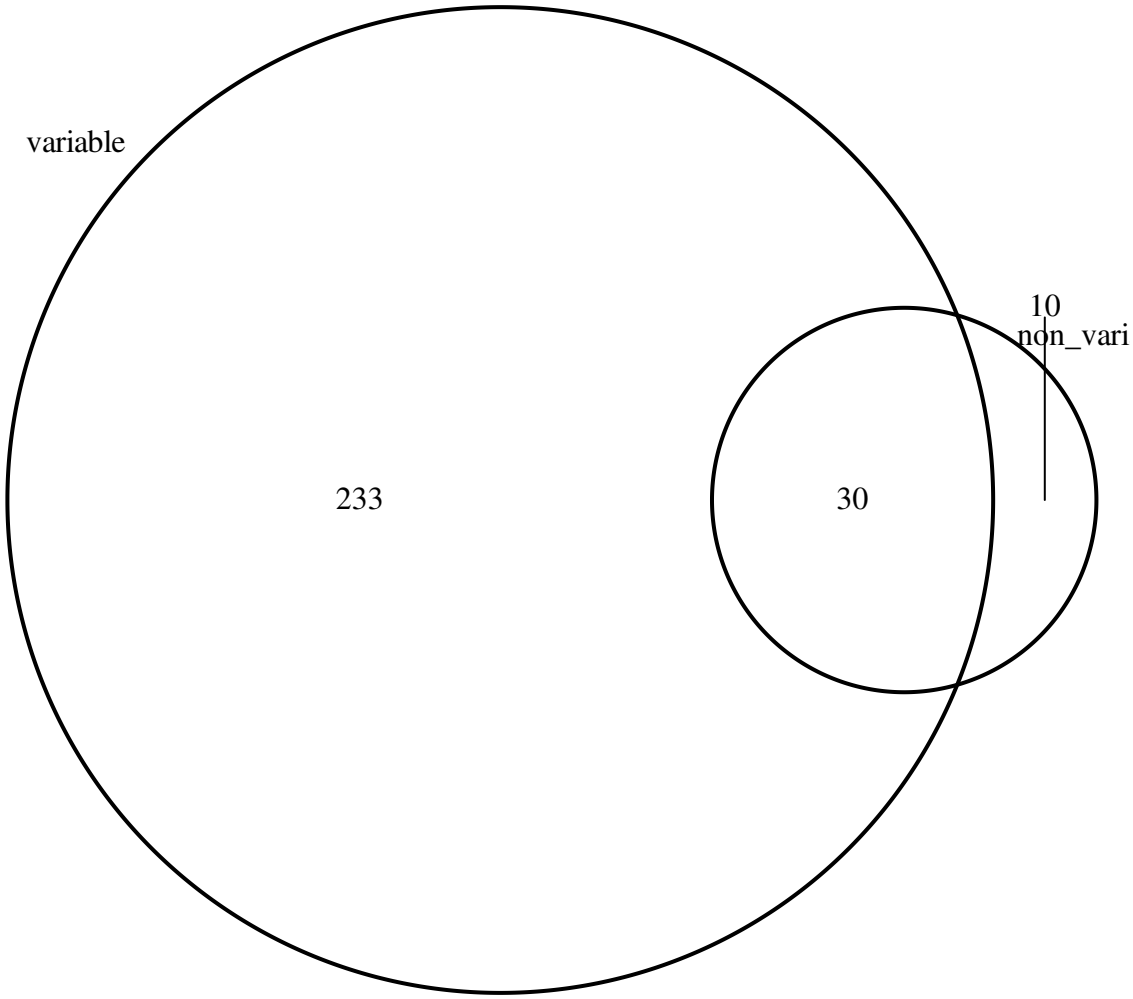
```r
# plot dotplot
dotplot(comp, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue < 0.01") +
  scale_y_discrete(labels = function(x) str_wrap(x, width = 100))
```


GO BP Enrichment Comparison
qvalue < 0.01
cilium assembly
cilium organization
microtubule-based transport
cilium movement
epithelial cilium movement involved in extracellular fluid movement
cilium or flagellum-dependent cell motility
cilium-dependent cell motility
axoneme assembly
regulation of cilium assembly
cilium movement involved in cell motility
axonemal dynein complex assembly
smoothened signaling pathway
intraciliary transport
intraciliary retrograde transport
regulation of protein-containing complex disassembly
specification of symmetry
determination of left/right symmetry
flagellated sperm motility
sperm motility
axoneme assembly
regulation of plasma membrane bounded cell projection assembly
regulation of cell projection assembly
microtubule bundle formation
determination of left/right symmetry
left/right pattern formation
regulation of organelle assembly
cytoskeleton-dependent intracellular transport
regulation of microtubule-based process
regionalization
regulation of protein localization to cilium
protein localization to cilium
transport along microtubule
cilium assembly
protein-containing complex disassembly
inner dynein arm assembly
positive regulation of protein-containing complex disassembly
regulation of smoothened signaling pathway
intraciliary anterograde transport
JNK cascade
protein-containing complex localization
non-motile cilium assembly
epithelial tube morphogenesis
negative regulation of smoothened signaling pathway
epithelial tube formation
protein localization to cell periphery
regulation of intracellular transport
microtubule polymerization
kidney development
protein localization to plasma membrane
microtubule polymerization or depolymerization
cilium organization
embryonic heart tube development
tube formation
neural tube development
neural tube formation
regulation of muscle system process
centrosome cycle
motile cilium assembly
embryonic epithelial tube formation
protein localization to centrosome
protein localization to microtubule organizing center
heart morphogenesis
microtubule organizing center organization
melanosome transport
embryonic organ morphogenesis
establishment of melanosome localization
microtubule cytoskeleton organization
in utero embryonic development
establishment of pigment granule localization
melanosome localization
regulation of microtubule-based movement
pigment granule localization
protein polymerization
regulation of microtubule cytoskeleton organization
spindle organization
morphogenesis of an epithelial tube
establishment of vesicle localization
heart looping
regulation of neuron projection development
positive regulation of cell projection organization
cytokinesis
muscle system process
determination of heart left/right asymmetry
embryonic heart tube morphogenesis
regulation of microtubule polymerization or depolymerization
regulation of muscle contraction
dorsal/ventral pattern formation
vesicle localization
regulation of microtubule polymerization
neural tube closure
protein localization to microtubule cytoskeleton
cytokinetic process
protein autophosphorylation
spindle assembly
regulation of neuronal synaptic plasticity
primary neural tube formation
spermatid development
regulation of endocytic recycling
vesicle-mediated transport to the plasma membrane
camera-type eye development
cytokinetic process
cytoskeleton-dependent cytokinesis
calcium-mediated signaling
spermatid differentiation
cellular process involved in reproduction in multicellular organism
sensory organ morphogenesis
protein localization to cytoskeleton
cytokinesis
insulin-like growth factor receptor signaling pathway
sperm flagellum assembly
endosomal transport
forebrain development
regulation of supramolecular fiber organization
establishment or maintenance of cell polarity
eye development
ERBB2 signaling pathway
benzene-containing compound metabolic process
visual system development
germ cell development
establishment of protein localization to membrane
regulation of dephosphorylation
sensory system development
telencephalon development
regulation of cardiac muscle contraction by calcium ion signaling
endocytic recycling
cytokinesis
photoreceptor cell development
endocrine system development
central nervous system neuron differentiation
photoreceptor cell differentiation
regulation of canonical Wnt signaling pathway
relaxation of cardiac muscle
cytokinesis
canonical Wnt signaling pathway
regulation of cardiac muscle contraction
axonogenesis
ureteric bud development
regulation of striated muscle contraction
mesonephric epithelium development
mesonephric tubule development
regulation of cell growth
positive regulation of striated muscle contraction
positive regulation of cardiac muscle cell apoptotic process
sarcoplasmic reticulum calcium ion transport
mesonephros development
respiratory system development
embryonic limb morphogenesis
embryonic appendage morphogenesis
eye morphogenesis
regulation of organelle transport along microtubule
regulation of release of sequestered calcium ion into cytosol
kidney epithelium development
digestive tract morphogenesis
appendage morphogenesis
limb morphogenesis
ear development
epidermal growth factor receptor signaling pathway
postsynaptic modulation of chemical synaptic transmission
establishment of protein localization to plasma membrane
establishment of cell polarity
animal organ formation
organelle transport along microtubule
positive regulation of muscle cell apoptotic process
microtubule anchoring at centrosome
positive regulation of cyclin-dependent protein serine/threonine kinase activity
cerebral cortex development
regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion
appendage development
limb development
gliogenesis
positive regulation of neuron projection development
positive regulation of protein phosphorylation
mesenchyme morphogenesis
learning or memory
positive regulation of cyclin-dependent protein kinase activity
release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
morphogenesis of a polarized epithelium
embryonic axis specification
regulation of Wnt signaling pathway
regulation of cyclin-dependent protein serine/threonine kinase activity
regulation of neurogenesis
centriole-centriole cohesion
microtubule anchoring at microtubule organizing center
regulation of Golgi organization
release of sequestered calcium ion into cytosol by endoplasmic reticulum
inner ear development
microtubule cytoskeleton organization involved in mitosis
cellular response to insulin stimulus
activin receptor signaling pathway
cellular pigmentation
Schwann cell development
positive regulation of microtubule polymerization
dephosphorylation
response to insulin
regulation of protein polymerization
regulation of transforming growth factor beta receptor signaling pathway
dorsal/ventral axis specification
smooth muscle contraction
pallium development
positive regulation of phosphorylation
protein dephosphorylation
negative regulation of growth
negative regulation of cell growth
regulation of cellular response to transforming growth factor beta stimulus
regulation of nervous system development
calcium ion transport
cellular response to hexose stimulus
regulation of protein localization to cell periphery
thyroid gland development
positive regulation of cell cycle
cellular response to vascular endothelial growth factor stimulus
cellular response to monosaccharide stimulus
peripheral nervous system development
regulation of synapse structure or activity
neural tube patterning
cognition
positive regulation of microtubule polymerization or depolymerization
glucose homeostasis
carbohydrate homeostasis
sperm axoneme assembly
cellular response to peptide hormone stimulus
Schwann cell differentiation
paraxial mesoderm development
trachea development
establishment of planar polarity
cellular response to hormone stimulus
regulation of protein kinase activity
post-Golgi vesicle-mediated transport
negative regulation of TORC1 signaling
regulation of metal ion transport
hormone transport
cellular response to carbohydrate stimulus
regulation of animal organ formation
membraneless organelle assembly
mitotic cell cycle phase transition
cochlea development
peptidyl-glutamic acid modification
regulation of muscle adaptation
regulation of cellular response to growth factor stimulus
cardiac muscle contraction
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
cochlea development
regulation of protein localization to membrane
regulation of intracellular protein transport
regulation of axonogenesis
embryonic pattern specification
regulation of synaptic plasticity
anoikis
relaxation of muscle
peptidyl-threonine phosphorylation
cranial skeletal system development
regulation of protein localization to plasma membrane
digestive tract development
regulation of cardiac muscle cell contraction
striated muscle adaptation
regulation of synapse organization
respiratory tube development
kidney morphogenesis
cell cycle G1/S phase transition
regulation of calcium ion transport
G1/S transition of mitotic cell cycle
neuronal action potential propagation
paraxial mesoderm morphogenesis
embryonic foregut morphogenesis
manchette assembly
positive regulation of protein polymerization
cerebellum development
phosphatidylinositol-3-phosphate biosynthetic process
heart formation
cell cycle G1/S phase transition
cellular response to glucose stimulus
hormone secretion
regulation of epithelial cell proliferation
p.adjust
0.0035
0.0030
0.0075
GeneRatio
0.1
0.2
non_variable
(104)
variable
(369)
Cluster

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot.svg"), plot = last_plot(), devi
```

Visualize overlap of enriched terms

```
# extract results
results <- comp@compareClusterResult

# split by location
variable <- results[results$Cluster == "variable",]
non_variable <- results[results$Cluster == "non_variable",]

# Create a list of sets
go_lists <- list(
  variable = variable$ID,
  non_variable = non_variable$ID
)

# Plot the Venn diagram
venn.plot <- venn.diagram(
  x = go_lists,
  category.names = c("variable", "non_variable"),
  filename = NULL,
  output = TRUE
)

grid.newpage()
grid.draw(venn.plot)
```




```
# Save the captured plot as an SVG file
svglite(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_venn.svg"), width = 6, height = 6)
grid.draw(venn.plot)
dev.off()
```

```
## cairo_pdf
##         2
```

Filter terms only enriched in one condition

```
# calculate intersection of the two
unspecific_terms <- intersect(variable$ID, non_variable$ID)

# remove all unspecific terms
specific_terms <- results %>% filter(!ID %in% unspecific_terms)
```

```r
# create a copy of comp
comp_filtered <- comp

# update results in comp
comp_filtered@compareClusterResult <- specific_terms
```

Dot plot of uniquely enriched terms

```r
# plot dotplot
dotplot(comp_filtered, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue
  scale_y_discrete(labels = function(x) str_wrap(x, width = 100))
```


GO BP Enrichment Comparison
qvalue < 0.01
epithelial cilium movement involved in extracellular fluid movement
extracellular transport
axonemal dynein complex assembly
regulation of protein-containing complex disassembly
regulation of protein localization to cilium
cerebrospinal fluid circulation
protein-containing complex disassembly
inner dynein arm assembly
positive regulation of protein-containing complex disassembly
JNK cascade
non-motile cilium assembly
epithelial tube morphogenesis
negative regulation of smoothened signaling pathway
epithelial tube formation
protein localization to cell periphery
regulation of intracellular transport
microtubule polymerization
kidney development
protein localization to plasma membrane
microtubule polymerization or depolymerization
renal system development
embryonic heart tube development
tube formation
neural tube development
neural tube formation
regulation of muscle system process
centrosome cycle
motile cilium assembly
embryonic epithelial tube formation
protein localization to centrosome
protein localization to microtubule organizing center
heart morphogenesis
microtubule organizing center organization
melanosome transport
embryonic organ morphogenesis
establishment of melanosome localization
pigment granule transport
in utero embryonic development
establishment of pigment granule localization
melanosome localization
regulation of microtubule-based movement
pigment granule localization
protein polymerization
regulation of microtubule cytoskeleton organization
spindle organization
morphogenesis of embryonic epithelium
establishment of vesicle localization
heart looping
regulation of neuron projection development
positive regulation of cell projection organization
cytokinesis
muscle system process
determination of heart left/right asymmetry
embryonic heart tube morphogenesis
regulation of microtubule polymerization or depolymerization
regulation of muscle contraction
dorsal/ventral pattern formation
vesicle localization
regulation of microtubule polymerization
neural tube closure
protein localization to microtubule cytoskeleton
tube closure
protein autophosphorylation
spindle assembly
regulation of neuronal synaptic plasticity
primary neural tube formation
spermatid development
regulation of endocytic recycling
vesicle-mediated transport to the plasma membrane
camera-type eye development
gland development
cytoskeleton-dependent cytokinesis
calcium-mediated signaling
spermatid differentiation
cellular process involved in reproduction in multicellular organism
sensory organ morphogenesis
protein localization to cytoskeleton
muscle contraction
insulin-like growth factor receptor signaling pathway
sperm flagellum assembly
endosomal transport
forebrain development
regulation of supramolecular fiber organization
establishment or maintenance of cell polarity
eye development
ERBB signaling pathway
benzene-containing compound metabolic process
vitamin metabolic process
germ cell development
establishment of protein localization to membrane
regulation of dephosphorylation
sensory system development
telencephalon development
regulation of cardiac muscle contraction by calcium ion signaling
endocytic recycling
axis specification
photoreceptor cell development
endocrine system development
central nervous system neuron differentiation
photoreceptor cell differentiation
regulation of canonical Wnt signaling pathway
relaxation of cardiac muscle
pericardium development
canonical Wnt signaling pathway
regulation of cardiac muscle contraction
exocytosis
ureteric bud development
regulation of striated muscle contraction
mesonephric epithelium development
mesonephric tubule development
regulation of cell growth
positive regulation of striated muscle cell apoptotic process
positive regulation of cardiac muscle cell apoptotic process
sarcoplasmic reticulum calcium ion transport
mesonephros development
respiratory system development
embryonic limb morphogenesis
embryonic appendage morphogenesis
eye morphogenesis
regulation of organelle transport along microtubule
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
kidney epithelium development
digestive tract morphogenesis
appendage morphogenesis
limb morphogenesis
ear development
epidermal growth factor receptor signaling pathway
postsynaptic modulation of chemical synaptic transmission
establishment of protein localization to plasma membrane
establishment of cell polarity
animal organ formation
organelle transport along microtubule
positive regulation of muscle cell apoptotic process
microtubule anchoring at centrosome
positive regulation of cyclin-dependent protein serine/threonine kinase activity
cerebral cortex development
regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion
appendage development
limb development
gliogenesis
positive regulation of neuron projection development
positive regulation of protein phosphorylation
mesenchyme morphogenesis
learning or memory
positive regulation of cyclin-dependent protein kinase activity
release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
morphogenesis of a polarized epithelium
embryonic axis specification
regulation of Wnt signaling pathway
regulation of cyclin-dependent protein serine/threonine kinase activity
regulation of neurogenesis
centriole-centriole cohesion
microtubule anchoring at microtubule organizing center
regulation of Golgi organization
release of sequestered calcium ion into cytosol by endoplasmic reticulum
inner ear development
microtubule cytoskeleton organization involved in mitosis
cellular response to insulin stimulus
activin receptor signaling pathway
cellular pigmentation
Schwann cell development
positive regulation of microtubule polymerization
dephosphorylation
response to insulin
regulation of protein polymerization
regulation of transforming growth factor beta receptor signaling pathway
dorsal/ventral axis specification
smooth muscle contraction
pallium development
positive regulation of phosphorylation
peptidyl-threonine modification
negative regulation of growth
negative regulation of cell growth
regulation of cellular response to transforming growth factor beta stimulus
regulation of nervous system development
calcium ion transport
cellular response to hexose stimulus
regulation of protein localization to cell periphery
thyroid gland development
positive regulation of intracellular transport
cellular response to vascular endothelial growth factor stimulus
cellular response to monosaccharide stimulus
peripheral nervous system development
regulation of synapse structure or activity
neural tube patterning
cognition
positive regulation of microtubule polymerization or depolymerization
glucose homeostasis
carbohydrate homeostasis
sperm axoneme assembly
cellular response to peptide hormone stimulus
Schwann cell differentiation
paraxial mesoderm development
trachea development
establishment of planar polarity
establishment of tissue polarity
regulation of protein kinase activity
post-Golgi vesicle-mediated transport
negative regulation of TORC1 signaling
regulation of metal ion transport
hormone transport
cellular response to carbohydrate stimulus
regulation of animal organ formation
membraneless organelle assembly
mitotic spindle organization
cochlea development
peptidyl-glutamic acid modification
regulation of muscle adaptation
regulation of cellular response to growth factor stimulus
cardiac muscle contraction
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
regulation of protein transport
regulation of protein localization to membrane
regulation of intracellular protein transport
regulation of axonogenesis
embryonic pattern specification
regulation of synaptic plasticity
anoikis
relaxation of muscle
peptidyl-threonine phosphorylation
cranial skeletal system development
regulation of protein localization to plasma membrane
digestive tract development
regulation of cardiac muscle cell contraction
striated muscle adaptation
regulation of synapse organization
respiratory tube development
kidney morphogenesis
microtubule nucleation
regulation of calcium ion transport
G1/S transition of mitotic cell cycle
neuronal action potential propagation
paraxial mesoderm morphogenesis
embryonic foregut morphogenesis
manchette assembly
positive regulation of protein polymerization
centrosome duplication
phosphatidylinositol-3-phosphate biosynthetic process
heart formation
cell cycle G1/S phase transition
cellular response to glucose stimulus
hormone secretion
regulation of epithelial cell proliferation
non_variable
(104)
variable
(369)
Cluster
p.adjust
0.0035
0.0030
0.0025
0.0020
GeneRatio
0.02
0.04
0.06

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_specific_terms_only.svg"), plo
```

```
# Subset for pc_tip
variable <- comp_filtered@compareClusterResult[
  comp_filtered@compareClusterResult$Cluster == "variable",
]

# Subset for bb_tz
non_variable <- comp_filtered@compareClusterResult[
  comp_filtered@compareClusterResult$Cluster == "non_variable",
]

# Create new compareClusterResult objects for each subset
comp_filtered_variable <- comp_filtered
comp_filtered_non_variable <- comp_filtered

comp_filtered_variable@compareClusterResult <- variable
comp_filtered_non_variable@compareClusterResult <- non_variable
```

Cluster results - variable proteins

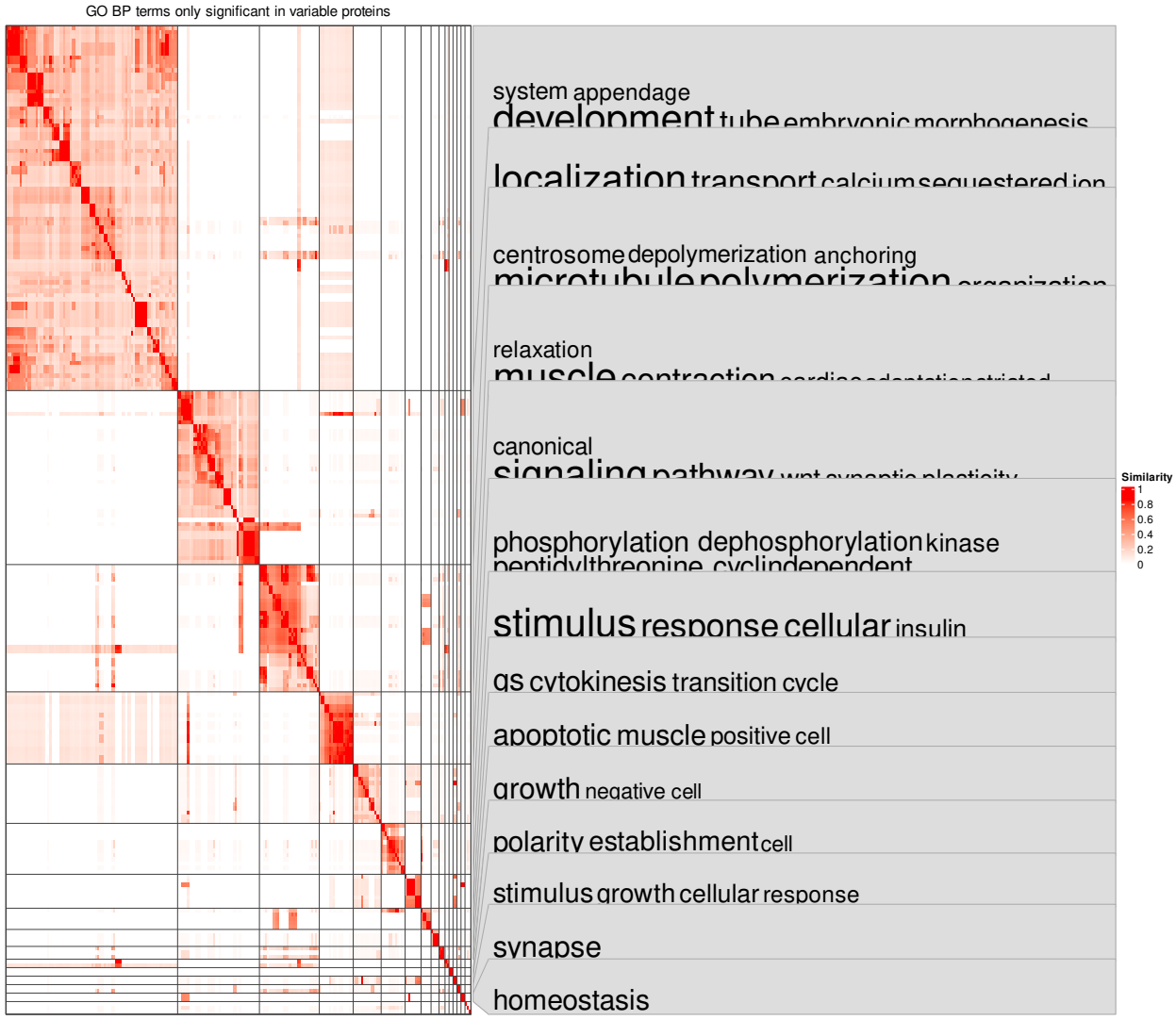
```
go_id = comp_filtered_variable@compareClusterResult$ID
mat = GO_similarity(go_id,
                    ont = 'BP',
                    db = 'org.Hs.eg.db',
                    measure = "Sim_Relevance_2006"
                    )
```

```
# Capture the plot
heatmap_plot <- grid.grabExpr({
 df <- simplifyGO(mat,
             method = 'binary_cut',
             plot = TRUE,
             column_title = "GO BP terms only significant in variable proteins",
             use_raster = FALSE,
             order_by_size = TRUE,
             fontsize_range = c(18,36),
             max_words = 6,
             word_cloud_grob_param = list(col = 'black',
                                          max_width = unit(200, "mm")))
})

# Save the captured plot as an SVG file
svglite(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_specific_terms_only_variable_
grid.draw(heatmap_plot)
dev.off()
```

Plot cluster heatmap

```
## cairo_pdf
##         2
```

```
grid.newpage()
grid.draw(heatmap_plot)
```




```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_variable,
                              df = df,
                              comp = TRUE
)

# Apply the function to each row of the DataFrame
results$GeneSymbol <- sapply(results$geneID, map_geneID_to_symbol)
```

```
# save results as csv file
write.csv(results, file = paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_specific_te
```

Process and save results

Cluster results - non-variable proteins

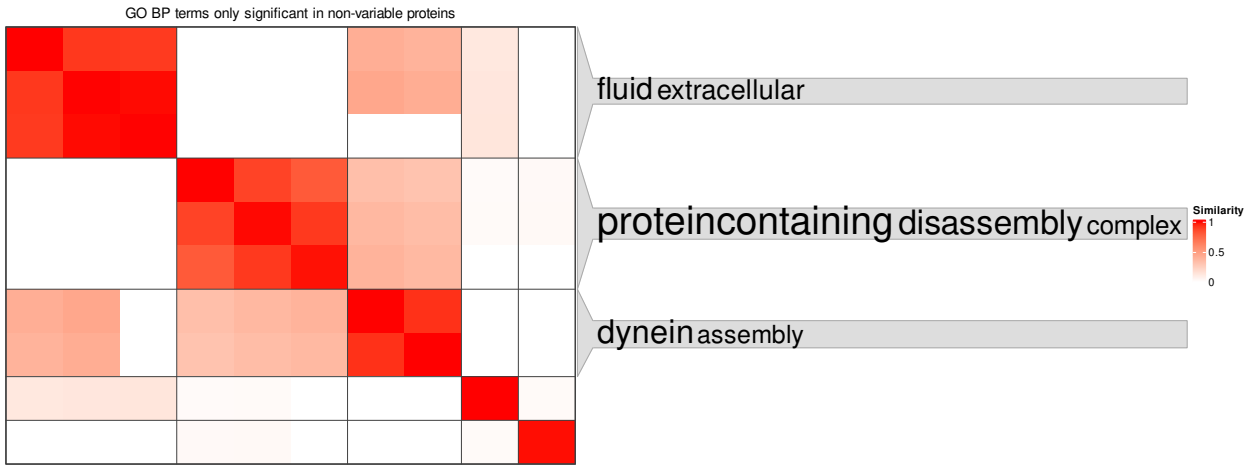
```
go_id = comp_filtered_non_variable@compareClusterResult$ID
mat = GO_similarity(go_id,
                    ont = 'BP',
                    db = 'org.Hs.eg.db',
                    measure = "Sim_Relevance_2006"
                    )
```

```
# Capture the plot
heatmap_plot <- grid.grabExpr({
 df <- simplifyGO(mat,
             method = 'binary_cut',
             plot = TRUE,
             column_title = "GO BP terms only significant in non-variable proteins",
             use_raster = FALSE,
             order_by_size = TRUE,
             fontsize_range = c(18, 36),
             max_words = 6,
             word_cloud_grob_param = list(col = 'black',
                                          max_width = unit(200, "mm")))
})

# Save the captured plot as an SVG file
svglite(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_specific_terms_only_nonvariab
grid.draw(heatmap_plot)
dev.off()
```

Plot cluster heatmap

```
## cairo_pdf
##         2
```

```
grid.newpage()
grid.draw(heatmap_plot)
```




```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_non_variable,
                              df = df,
                              comp = TRUE
)

# Apply the function to each row of the DataFrame
results$GeneSymbol <- sapply(results$geneID, map_geneID_to_symbol)

# save results as csv file
write.csv(results, file = paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_specific_te
```

Process and save results

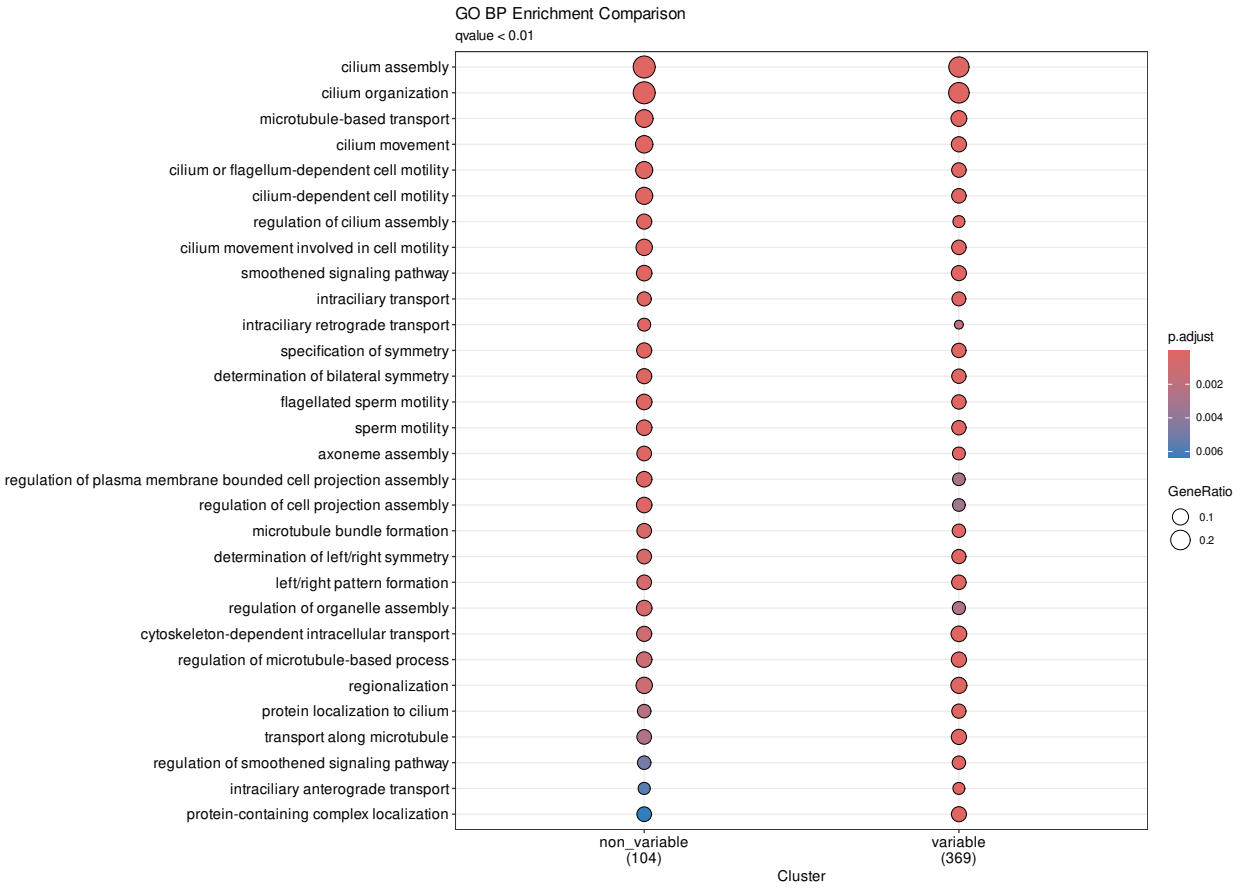
Filter for terms enriched in both variable and non-variable

```
 # get results as data frame
# extract results
results <- comp@compareClusterResult

# split by location
variable <- results[results$Cluster == "variable",]
non_variable <- results[results$Cluster == "non_variable",]

# calculate intersection of the two
unspecific_terms <- intersect(variable$ID, non_variable$ID)

# remove all unspecific terms
specific_terms <- results %>% filter(!ID %in% unspecific_terms)
unspecific_terms <- results %>% filter(ID %in% unspecific_terms)

# create a copy of comp
comp_filtered <- comp
```

```
# update results in comp
comp_filtered@compareClusterResult <- unspecific_terms
```

Plot dotplot of shared terms

```
# plot dotplot
dotplot(comp_filtered, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue
  scale_y_discrete(labels = function(x) str_wrap(x, width = 100))
```




```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_shared_terms_only.svg"), plot =
```

Cluster results - Both locations

```
# Split by location
variable <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "variable"
non_variable <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "non_va

# Create new compareClusterResult objects for each subset
comp_filtered_variable<- comp_filtered
```

```
comp_filtered_non_variable <- comp_filtered

comp_filtered_variable@compareClusterResult <- variable
comp_filtered_non_variable@compareClusterResult <- non_variable
```

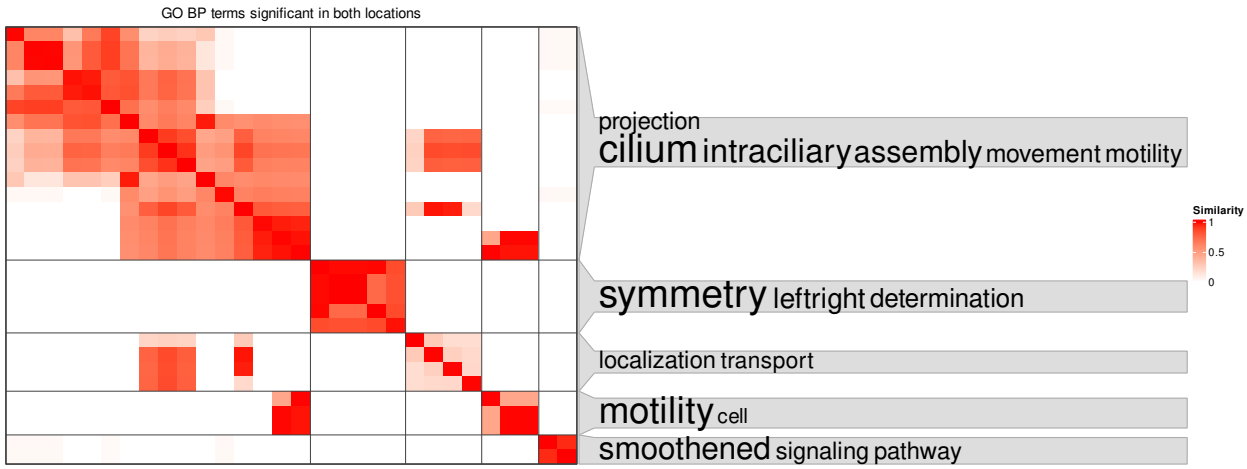
```
go_id = comp_filtered_variable@compareClusterResult$ID
mat = GO_similarity(go_id,
                    ont = 'BP',
                    db = 'org.Hs.eg.db',
                    measure = "Sim_Relevance_2006"
                    )
```

```
# Capture the plot
heatmap_plot <- grid.grabExpr({
 df <- simplifyGO(mat,
             method = 'binary_cut',
             plot = TRUE,
             column_title = "GO BP terms significant in both locations",
             use_raster = FALSE,
             order_by_size = TRUE,
             fontsize_range = c(18, 36),
             max_words = 6,
             word_cloud_grob_param = list(col = 'black',
                                          max_width = unit(200, "mm")))
})

# Save the captured plot as an SVG file
svglite(paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_shared_terms_heatmap.svg"), w
grid.draw(heatmap_plot)
dev.off()
```

Plot cluster heatmap

```
## cairo_pdf
##         2
```

```
grid.newpage()
grid.draw(heatmap_plot)
```




```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_variable,
                              df = df,
                              comp = TRUE
)

# Apply the function to each row of the DataFrame
results$GeneSymbol <- sapply(results$geneID, map_geneID_to_symbol)

# save results as csv file
write.csv(results, file = paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_shared_terms
```

```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_non_variable,
                              df = df,
                              comp = TRUE
)

# Apply the function to each row of the DataFrame
results$GeneSymbol <- sapply(results$geneID, map_geneID_to_symbol)

# save results as csv file
write.csv(results, file = paste0(out_path, "NonRestricted_comparison_variable_GO_BP_dotplot_shared_terms
```

Process and save results

Session info

```
sessionInfo()
```

```
## R version 4.5.0 (2025-04-11)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 24.04.2 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.12.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0  LAPACK version 3.12.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/Los_Angeles
## tzcode source: system (glibc)
##
## attached base packages:
## [1] grid      stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] readxl_1.4.5             simplifyEnrichment_2.2.0 VennDiagram_1.7.3
##  [4] futile.logger_1.4.3      svglite_2.1.3            Cairo_1.6-2
##  [7] org.Hs.eg.db_3.21.0      AnnotationDbi_1.70.0     IRanges_2.42.0
## [10] S4Vectors_0.46.0         Biobase_2.68.0           BiocGenerics_0.54.0
## [13] generics_0.1.3           enrichplot_1.28.2        clusterProfiler_4.16.0
## [16] lubridate_1.9.4          forcats_1.0.0            stringr_1.5.1
## [19] dplyr_1.1.4              purrr_1.0.4              readr_2.1.5
## [22] tidyr_1.3.1              tibble_3.2.1             ggplot2_3.5.2
## [25] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##   [1] RColorBrewer_1.1-3      rstudioapi_0.17.1       jsonlite_2.0.0
##   [4] shape_1.4.6.1           magrittr_2.0.3          modeltools_0.2-24
##   [7] ggtangle_0.0.6          farver_2.1.2            rmarkdown_2.29
##  [10] ragg_1.4.0              GlobalOptions_0.1.2     fs_1.6.6
##  [13] vctrs_0.6.5             memoise_2.0.1           ggtree_3.16.0
##  [16] htmltools_0.5.8.1       lambda.r_1.2.4          cellranger_1.1.0
##  [19] gridGraphics_0.5-1      plyr_1.8.9              futile.options_1.0.1
##  [22] cachem_1.1.0            igraph_2.1.4            mime_0.13
##  [25] lifecycle_1.0.4         iterators_1.0.14        pkgconfig_2.0.3
##  [28] Matrix_1.7-3            R6_2.6.1                fastmap_1.2.0
##  [31] gson_0.1.0              GenomeInfoDbData_1.2.14 shiny_1.10.0
##  [34] clue_0.3-66             digest_0.6.37           aplot_0.2.5
##  [37] colorspace_2.1-1        patchwork_1.3.0         textshaping_1.0.1
##  [40] RSQLite_2.3.11          labeling_0.4.3          timechange_0.3.0
##  [43] httr_1.4.7              compiler_4.5.0          bit64_4.6.0-1
##  [46] withr_3.0.2             doParallel_1.0.17       BiocParallel_1.42.0
##  [49] DBI_1.2.3               R.utils_2.13.0          scatterplot3d_0.3-44
##  [52] rjson_0.2.23            tools_4.5.0             ape_5.8-1
##  [55] flexclust_1.5.0         httpuv_1.6.16           R.oo_1.27.1
```

```
##  [58] glue_1.8.0              promises_1.3.2          nlme_3.1-168
##  [61] GOSemSim_2.34.0         cluster_2.1.8.1         reshape2_1.4.4
##  [64] fgsea_1.34.0            gtable_0.3.6            tzdb_0.5.0
##  [67] class_7.3-23            R.methodsS3_1.8.2       data.table_1.17.0
##  [70] hms_1.1.3               xml2_1.3.8              XVector_0.48.0
##  [73] ggrepel_0.9.6           foreach_1.5.2           pillar_1.10.2
##  [76] yulab.utils_0.2.0       later_1.4.2             circlize_0.4.16
##  [79] splines_4.5.0           treeio_1.32.0           lattice_0.22-5
##  [82] bit_4.6.0               tidyselect_1.2.1        GO.db_3.21.0
##  [85] ComplexHeatmap_2.24.0   tm_0.7-16               Biostrings_2.76.0
##  [88] knitr_1.50              NLP_0.3-2               xfun_0.52
##  [91] matrixStats_1.5.0       stringi_1.8.7           UCSC.utils_1.4.0
##  [94] lazyeval_0.2.2          ggfun_0.1.8             yaml_2.3.10
##  [97] evaluate_1.0.3          codetools_0.2-20        qvalue_2.40.0
## [100] Polychrome_1.5.4        ggplotify_0.1.2         cli_3.6.5
## [103] xtable_1.8-4            systemfonts_1.2.3       Rcpp_1.0.14
## [106] GenomeInfoDb_1.44.0     png_0.1-8               parallel_4.5.0
## [109] simona_1.6.0            blob_1.2.4              DOSE_4.2.0
## [112] slam_0.1-55             tidytree_0.4.6          scales_1.4.0
## [115] crayon_1.5.3            GetoptLong_1.0.5        rlang_1.1.6
## [118] cowplot_1.1.3           fastmatch_1.1-6         KEGGREST_1.48.0
## [121] formatR_1.14
```