Comparative functional enrichment analysis between subciliary locations for all proteins and cell types combined

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)
```

Define function

```
# 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)
df_nucleus <- df_all %>% filter(Nucleus == TRUE)
df_mitotic <- df_all %>% filter(Mitotic == TRUE)
df_membrane <- df_all %>% filter(Membrane == TRUE)
df_cytoplasm <- df_all %>% filter(Cytoplasm == TRUE)

# save input data as csv file
write.csv(df_bb, file = paste0(out_path, "NonRestricted_comparison_locations_input_bb.csv"), row.names =
write.csv(df_pc, file = paste0(out_path, "NonRestricted_comparison_locations_input_pc.csv"), row.names =
write.csv(df_tip, file = paste0(out_path, "NonRestricted_comparison_locations_input_tip.csv"), row.name
write.csv(df_tz, file = paste0(out_path, "NonRestricted_comparison_locations_input_tz.csv"), row.names =
write.csv(df_nucleus, file = paste0(out_path, "NonRestricted_comparison_locations_input_nucleus.csv"), r
write.csv(df_mitotic, file = paste0(out_path, "NonRestricted_comparison_locations_input_mitotic.csv"), r
write.csv(df_membrane, file = paste0(out_path, "NonRestricted_comparison_locations_input_membrane.csv")
write.csv(df_cytoplasm, file = paste0(out_path, "NonRestricted_comparison_locations_input_cytoplasm.csv
```

```
# 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$Ensembl_ID
gene_id_nucleus <- df_nucleus$Ensembl_ID
gene_id_mitotic <- df_mitotic$Ensembl_ID
gene_id_membrane <- df_membrane$Ensembl_ID
gene_id_cytoplasm <- df_cytoplasm$Ensembl_ID
```

Combine bb & tz and pc & tip

```
# combine bb and tz
df_bb_tz <- rbind(df_bb, df_tz)

# drop duplicates based on Ensembl_ID column
df_bb_tz <- df_bb_tz[!duplicated(df_bb_tz$Ensembl_ID), ]
gene_id_bb_tz <- df_bb_tz$Ensembl_ID


# combine pc and tip
df_pc_tip <- rbind(df_pc, df_tip)

# drop duplicates based on Ensembl_ID column
df_pc_tip <- df_pc_tip[!duplicated(df_pc_tip$Ensembl_ID), ]
gene_id_pc_tip <- df_pc_tip$Ensembl_ID


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

# 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

# get intersection of all genes
gene_id_all_cilia <- Reduce(union, list(gene_id_bb_tz, gene_id_pc_tip))
print(paste("Number of genes in all cilia locations:", length(gene_id_all_cilia)))
```

```
## [1] "Number of genes in all cilia locations: 715"
```

All proteins

GO BP enrichment analysis

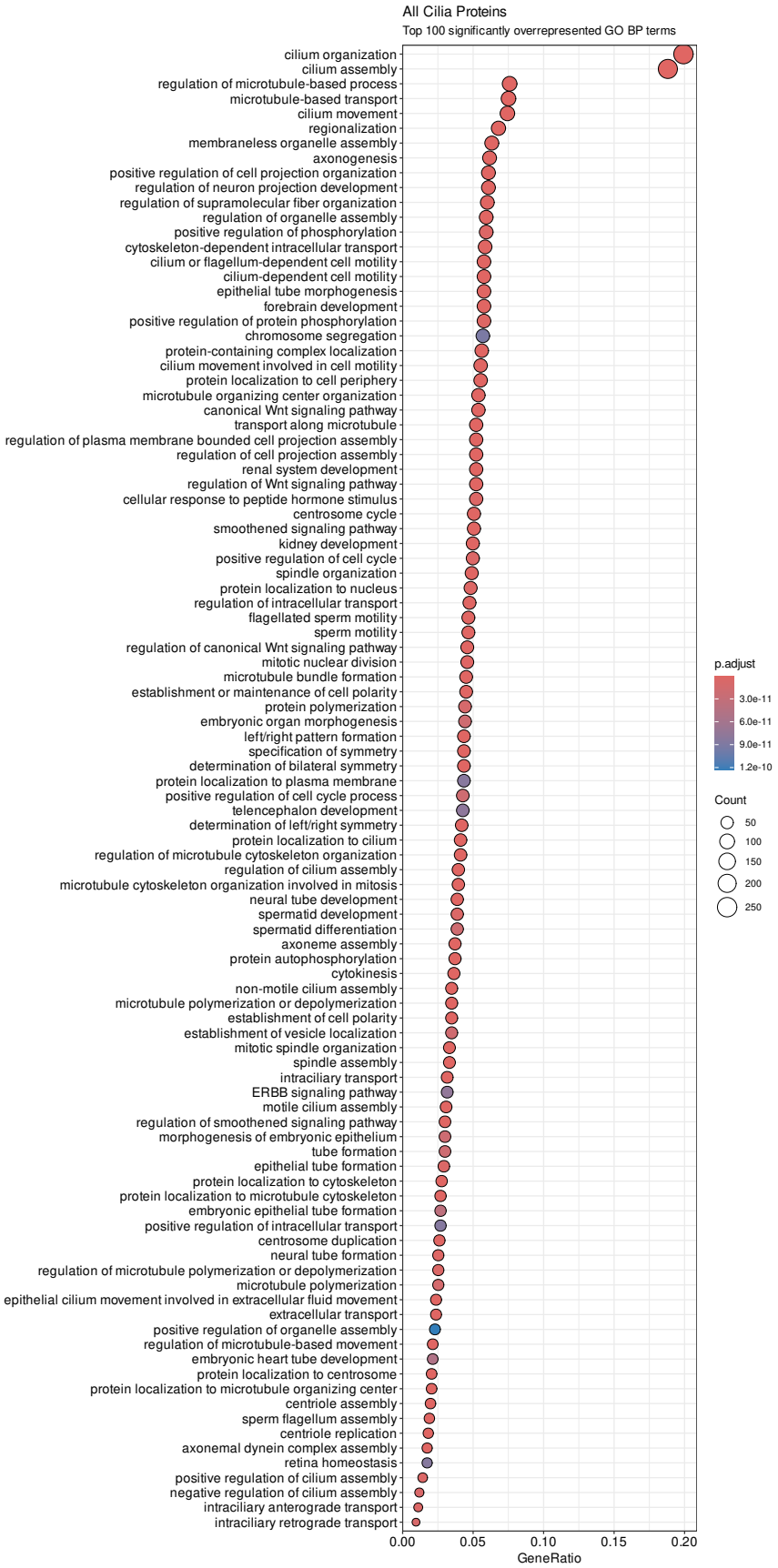
```
ego <- enrichGO(gene          = gene_id_all,
                OrgDb         = org.Hs.eg.db,
```

```
                keyType       = 'ENSEMBL',
                ont           = "BP",
                pAdjustMethod = "BH",
                pvalueCutoff  = 0.01,
                qvalueCutoff  = 0.01,
                readable      = TRUE)

ego@result <- ego@result[ego@result$p.adjust < 0.01,]
ego
```

```
## #
## # over-representation test
## #
## #...@organism     Homo sapiens
## #...@ontology     BP
## #...@keytype      ENSEMBL
## #...@gene     chr [1:1360] "ENSG00000001497" "ENSG00000002330" "ENSG00000002549" ...
## #...pvalues adjusted by 'BH' with cutoff <0.01
## #...1019 enriched terms found
## 'data.frame':    1019 obs. of  12 variables:
##  $ ID            : chr  "GO:0044782" "GO:0060271" "GO:0099111" "GO:0003341" ...
##  $ Description   : chr  "cilium organization" "cilium assembly" "microtubule-based transport" "ciliu
##  $ GeneRatio     : chr  "252/1265" "238/1265" "95/1265" "94/1265" ...
##  $ BgRatio       : chr  "441/21273" "411/21273" "235/21273" "272/21273" ...
##  $ RichFactor    : num  0.571 0.579 0.404 0.346 0.684 ...
##  $ FoldEnrichment: num  9.61 9.74 6.8 5.81 11.51 ...
##  $ zScore        : num  45.9 45 22.5 20.1 23.1 ...
##  $ pvalue        : num  2.66e-195 6.63e-186 1.69e-54 6.69e-47 6.33e-46 ...
##  $ p.adjust      : num  1.50e-191 1.87e-182 3.19e-51 9.45e-44 7.15e-43 ...
##  $ qvalue        : num  1.01e-191 1.26e-182 2.14e-51 6.34e-44 4.80e-43 ...
##  $ geneID        : chr  "ZMYND10/MARK4/FUZ/MKS1/EHD3/NUDCD3/EHD2/IFT88/CLXN/OFD1/MAP4/LIMA1/MPHOSPH9/
##  $ Count         : int  252 238 95 94 52 96 50 40 68 64 ...
## #...Citation
## S Xu, E Hu, Y Cai, Z Xie, X Luo, L Zhan, W Tang, Q Wang, B Liu, R Wang, W Xie, T Wu, L Xie, G Yu. Usi
```

Dot plot of top 100 enriched proteins

```
dotplot(ego, showCategory=100, label_format = 50) + ggtitle("All Cilia Proteins", subtitle = "Top 100 si
  scale_y_discrete(labels = function(x) str_wrap(x, width = 70))
```


All Cilia Proteins
Top 100 significantly overrepresented GO BP terms
cilium organization
cilium assembly
regulation of microtubule-based process
microtubule-based transport
cilium movement
regionalization
membraneless organelle assembly
axonogenesis
positive regulation of cell projection organization
regulation of neuron projection development
regulation of supramolecular fiber organization
regulation of organelle assembly
positive regulation of phosphorylation
cytoskeleton-dependent intracellular transport
cilium or flagellum-dependent cell motility
cilium-dependent cell motility
epithelial tube morphogenesis
forebrain development
positive regulation of protein phosphorylation
chromosome segregation
protein-containing complex localization
cilium movement involved in cell motility
protein localization to cell periphery
microtubule organizing center organization
canonical Wnt signaling pathway
transport along microtubule
regulation of plasma membrane bounded cell projection assembly
regulation of cell projection assembly
renal system development
regulation of Wnt signaling pathway
cellular response to peptide hormone stimulus
centrosome cycle
smoothened signaling pathway
kidney development
positive regulation of cell cycle
spindle organization
protein localization to nucleus
regulation of intracellular transport
flagellated sperm motility
sperm motility
regulation of canonical Wnt signaling pathway
mitotic nuclear division
microtubule bundle formation
establishment or maintenance of cell polarity
protein polymerization
embryonic organ morphogenesis
left/right pattern formation
specification of symmetry
determination of bilateral symmetry
protein localization to plasma membrane
positive regulation of cell cycle process
telencephalon development
determination of left/right symmetry
protein localization to cilium
regulation of microtubule cytoskeleton organization
regulation of cilium assembly
microtubule cytoskeleton organization involved in mitosis
neural tube development
spermatid development
spermatid differentiation
axoneme assembly
protein autophosphorylation
cytokinesis
non-motile cilium assembly
microtubule polymerization or depolymerization
establishment of cell polarity
establishment of vesicle localization
mitotic spindle organization
spindle assembly
intraciliary transport
ERBB signaling pathway
motile cilium assembly
regulation of smoothened signaling pathway
morphogenesis of embryonic epithelium
tube formation
epithelial tube formation
protein localization to cytoskeleton
protein localization to microtubule cytoskeleton
embryonic epithelial tube formation
positive regulation of intracellular transport
centrosome duplication
neural tube formation
regulation of microtubule polymerization or depolymerization
microtubule polymerization
epithelial cilium movement involved in extracellular fluid movement
extracellular transport
positive regulation of organelle assembly
regulation of microtubule-based movement
embryonic heart tube development
protein localization to centrosome
protein localization to microtubule organizing center
centriole assembly
sperm flagellum assembly
centriole replication
axonemal dynein complex assembly
retina homeostasis
positive regulation of cilium assembly
negative regulation of cilium assembly
intraciliary anterograde transport
intraciliary retrograde transport
0.00
0.05
0.10
0.15
0.20
GeneRatio
p.adjust
3.0e-11
6.0e-11
9.0e-11
1.2e-10
Count
50
100
150
200
250

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

Compare biological themes for different locations (cilia only, PC+tip vs TZ+BB)

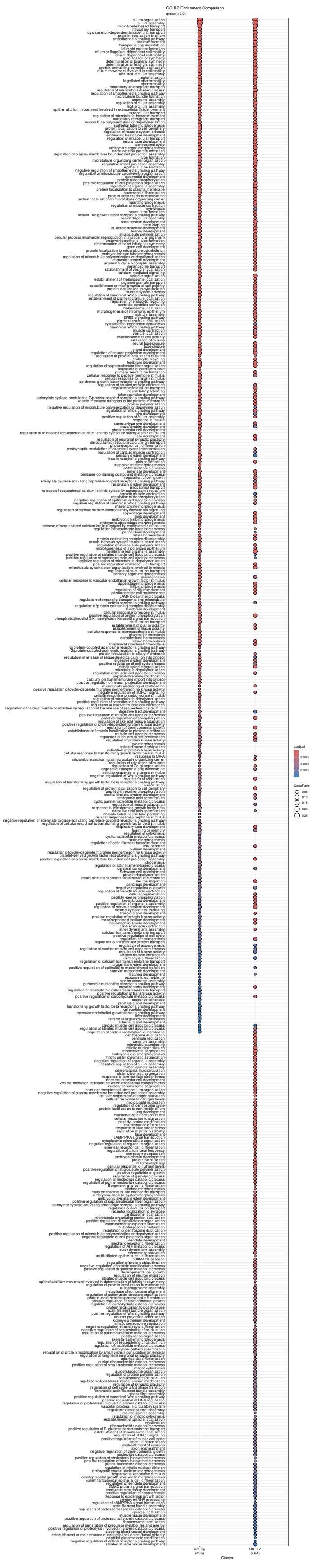
```r
# prepare input list
input_genes <- list(
    PC_tip = gene_id_pc_tip,
    BB_TZ = gene_id_bb_tz
)
```

GO BP enrichment analysis

```r
# Perform the compareCluster analysis
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

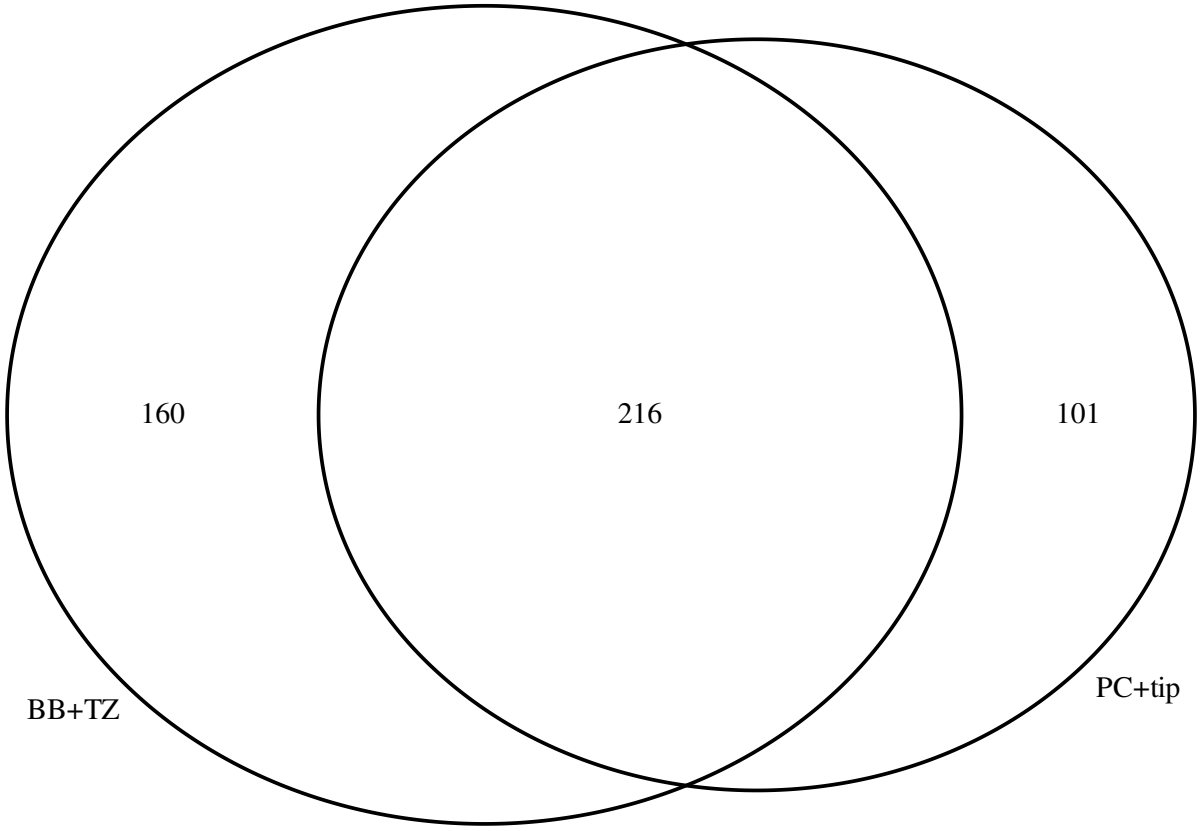
```r
dotplot(comp, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue < 0.01")
  scale_y_discrete(labels = function(x) str_wrap(x, width = 100))
```


GO BP Enrichment Comparison
qvalue < 0.01
PC_tip
(455)
BB_TZ
(464)
Cluster

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_locations_cilia_only_combined_GO_BP_dotplot.svg"), pl
```

Visualize overlap of enriched terms as Venn diagram

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

# split by location
pc_tip <- results[results$Cluster == "PC_tip",]
bb_tz <- results[results$Cluster == "BB_TZ",]

# Create a list of the four sets
go_lists <- list(
  PC_tip = pc_tip$ID,
  BB_TZ = bb_tz$ID
)

# Plot the Venn diagram
venn.plot <- venn.diagram(
  x = go_lists,
  category.names = c("PC+tip", "BB+TZ"),
  filename = NULL,
  output = TRUE
)

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




```
# Save the captured plot as an SVG file
svglite(paste0(out_path, "NonRestricted_comparison_locations_cilia_only_combined_GO_BP_venn.svg"), widt
grid.draw(venn.plot)
dev.off()
```

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

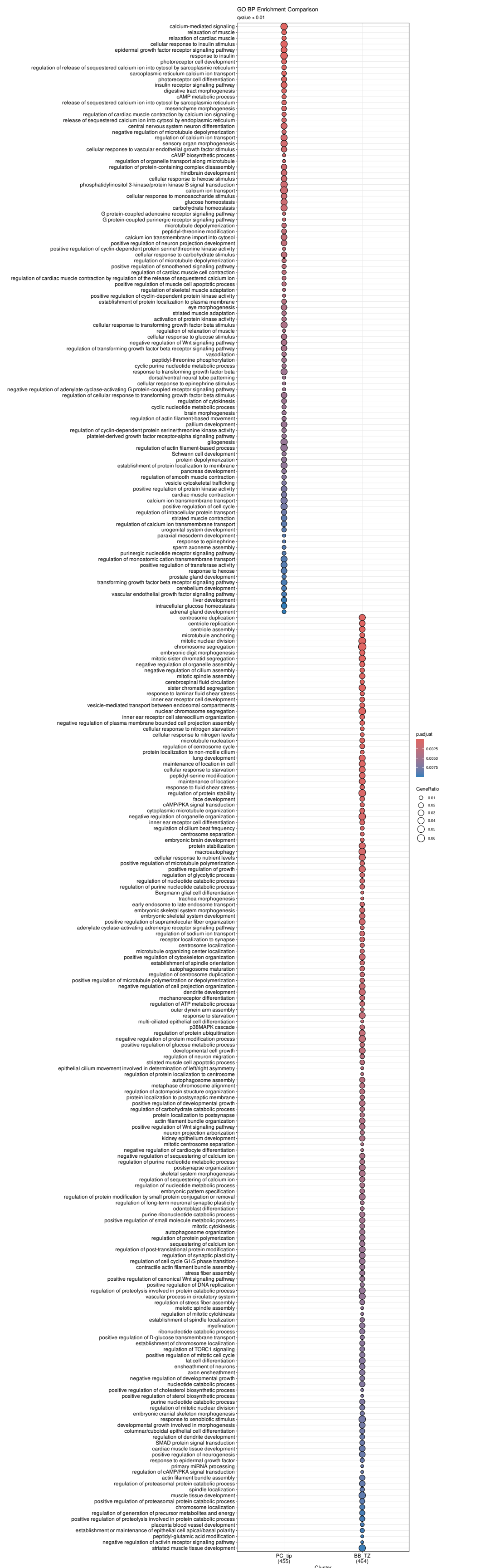
Determine uniquely enriched and unspecifically enriched terms

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

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

# create a copy of comp
```

```r
comp_filtered <- comp

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

Dot plot of uniquely enriched terms

```r
# plot(xx, type="dot", caption="GO Enrichment Comparison")
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
PC_tip
(455)
BB_TZ
(464)
Cluster
p.adjust
GeneRatio

```r
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_locations_cilia_only_combined_GO_BP_dotplot_specific_
```

Prepare data for clustering of enriched terms

```r
# Split by location
pc_tip <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "PC_tip", ]
bb_tz <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "BB_TZ", ]

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

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

# Create new compareClusterResult objects for each subset
comp_filtered_pc_tip <- comp_filtered
comp_filtered_bb_tz <- comp_filtered

comp_filtered_pc_tip@compareClusterResult <- pc_tip
comp_filtered_bb_tz@compareClusterResult <- bb_tz
```

Cluster results - PC&tip

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

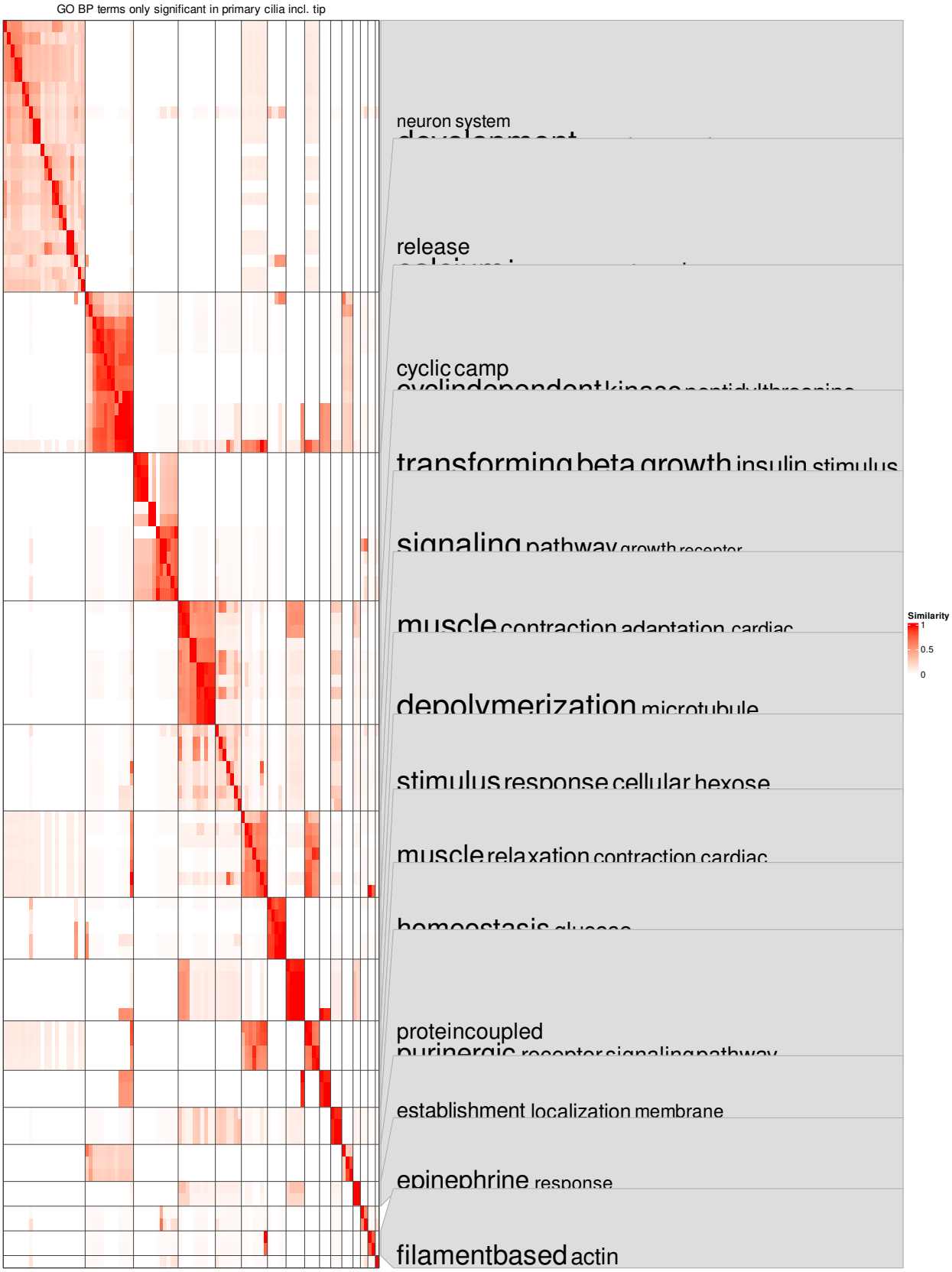
Plot cluster heatmap

```r
# Capture the plot
heatmap_plot <- grid.grabExpr({
 df <- simplifyGO(mat,
             method = 'binary_cut',
             plot = TRUE,
             column_title = "GO BP terms only significant in primary cilia incl. tip",
             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_locations_cilia_only_combined_GO_BP_dotplot_specific_
grid.draw(heatmap_plot)
dev.off()
```

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

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


GO BP terms only significant in primary cilia incl. tip
neuron system
release
cyclic camp
transforming beta growth insulin stimulus
signaling pathway growth receptor
muscle contraction adaptation cardiac
depolymerization microtubule
stimulus response cellular hexose
muscle relaxation contraction cardiac
homeostasis glucose
proteincoupled
purinergic receptor signaling pathway
establishment localization membrane
epinephrine response
filamentbased actin
Similarity
1
0.5
0

Process and save results

```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_pc_tip,
                              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_locations_cilia_only_combined_GO_BI
```

Cluster results - BB&TZ

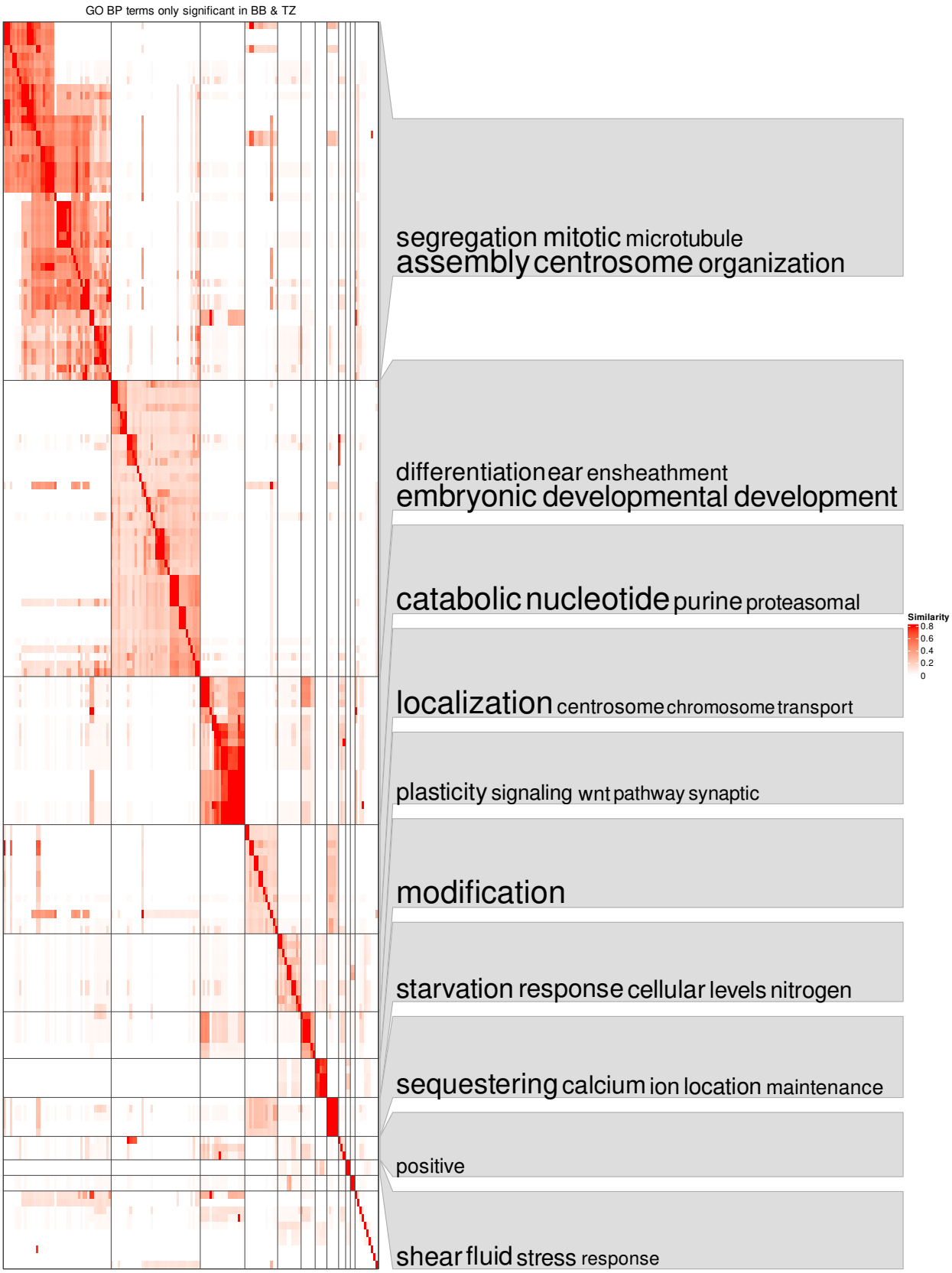
```
go_id = bb_tz$ID
mat = GO_similarity(go_id,
                    ont = 'BP',
                    db = 'org.Hs.eg.db',
                    measure = "Sim_Relevance_2006"
                    )
```

Plot cluster heatmap

```
# Capture the plot
heatmap_plot <- grid.grabExpr({
 df <- simplifyGO(mat,
             method = 'binary_cut',
             plot = TRUE,
             column_title = "GO BP terms only significant in BB & TZ",
             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_locations_cilia_only_combined_GO_BP_dotplot_specific_
grid.draw(heatmap_plot)
dev.off()
```

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

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


GO BP terms only significant in BB & TZ
segregation mitotic microtubule
assembly centrosome organization
differentiation ear ensheathment
embryonic developmental development
catabolic nucleotide purine proteasomal
localization centrosome chromosome transport
plasticity signaling wnt pathway synaptic
modification
starvation response cellular levels nitrogen
sequestering calcium ion location maintenance
positive
shear fluid stress response
Similarity
0.8
0.6
0.4
0.2
0

Process and save results

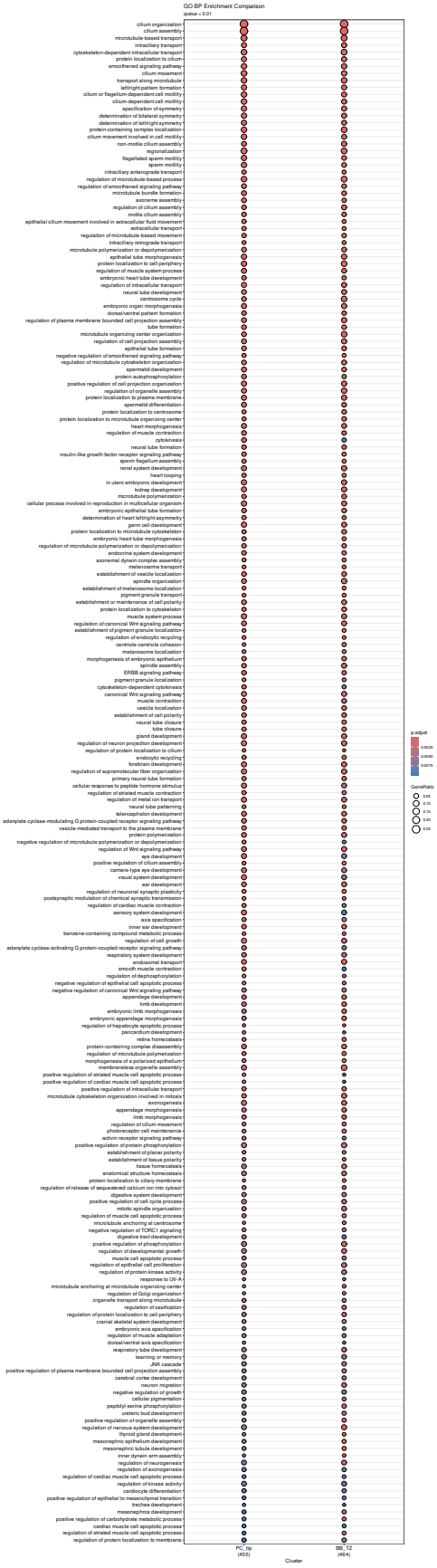
```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_bb_tz,
                              df = df,
                              comp = TRUE
)

# 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_locations_cilia_only_combined_GO_B
```

Filter for terms enriched in both locations

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

# get all terms significant in each location
pc_tip <- comp_results %>% filter(Cluster == "PC_tip")
bb_tz <- comp_results %>% filter(Cluster == "BB_TZ")

# calculate intersection of the two
unspecific_terms <- intersect(pc_tip$ID, bb_tz$ID)

# remove all unspecific terms
unspecific_terms <- comp_results %>% filter(ID %in% unspecific_terms)

# create a copy of comp
comp_filtered <- comp

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

Dot plot 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))
```


GO BP Enrichment Comparison
qvalue < 0.01
cilium organization
cilium assembly
microtubule-based transport
intraciliary transport
cytoskeleton-dependent intracellular transport
protein localization to cilium
smoothened signaling pathway
cilium movement
transport along microtubule
left/right pattern formation
cilium or flagellum-dependent cell motility
cilium-dependent cell motility
specification of symmetry
determination of bilateral symmetry
determination of left/right symmetry
protein-containing complex localization
cilium movement involved in cell motility
non-motile cilium assembly
regionalization
flagellated sperm motility
sperm motility
intraciliary anterograde transport
regulation of microtubule-based process
regulation of smoothened signaling pathway
microtubule bundle formation
axoneme assembly
regulation of cilium assembly
motile cilium assembly
epithelial cilium movement involved in extracellular fluid movement
extracellular transport
regulation of microtubule-based movement
intraciliary retrograde transport
microtubule polymerization or depolymerization
epithelial tube morphogenesis
protein localization to cell periphery
regulation of muscle system process
embryonic heart tube development
regulation of intracellular transport
neural tube development
centrosome cycle
embryonic organ morphogenesis
dorsal/ventral pattern formation
regulation of plasma membrane bounded cell projection assembly
tube formation
microtubule organizing center organization
regulation of cell projection assembly
epithelial tube formation
negative regulation of smoothened signaling pathway
regulation of microtubule cytoskeleton organization
spermatid development
protein autophosphorylation
positive regulation of cell projection organization
regulation of organelle assembly
protein localization to plasma membrane
spermatid differentiation
protein localization to centrosome
protein localization to microtubule organizing center
heart morphogenesis
regulation of muscle contraction
cytokinesis
neural tube formation
insulin-like growth factor receptor signaling pathway
sperm flagellum assembly
renal system development
heart looping
in utero embryonic development
kidney development
microtubule polymerization
cellular process involved in reproduction in multicellular organism
embryonic epithelial tube formation
determination of heart left/right asymmetry
germ cell development
protein localization to microtubule cytoskeleton
embryonic heart tube morphogenesis
regulation of microtubule polymerization or depolymerization
endocrine system development
axonemal dynein complex assembly
melanosome transport
establishment of vesicle localization
spindle organization
establishment of melanosome localization
pigment granule transport
establishment or maintenance of cell polarity
protein localization to cytoskeleton
muscle system process
regulation of canonical Wnt signaling pathway
establishment of pigment granule localization
regulation of endocytic recycling
centriole-centriole cohesion
melanosome localization
morphogenesis of embryonic epithelium
spindle assembly
ERBB signaling pathway
pigment granule localization
cytoskeleton-dependent cytokinesis
canonical Wnt signaling pathway
muscle contraction
vesicle localization
establishment of cell polarity
neural tube closure
tube closure
gland development
regulation of neuron projection development
regulation of protein localization to cilium
endocytic recycling
forebrain development
regulation of supramolecular fiber organization
primary neural tube formation
cellular response to peptide hormone stimulus
regulation of striated muscle contraction
regulation of metal ion transport
neural tube patterning
telencephalon development
adenylate cyclase-modulating G protein-coupled receptor signaling pathway
vesicle-mediated transport to the plasma membrane
protein polymerization
negative regulation of microtubule polymerization or depolymerization
regulation of Wnt signaling pathway
eye development
positive regulation of cilium assembly
camera-type eye development
visual system development
ear development
regulation of neuronal synaptic plasticity
postsynaptic modulation of chemical synaptic transmission
regulation of cardiac muscle contraction
sensory system development
axis specification
inner ear development
benzene-containing compound metabolic process
regulation of cell growth
adenylate cyclase-activating G protein-coupled receptor signaling pathway
respiratory system development
endosomal transport
smooth muscle contraction
regulation of dephosphorylation
negative regulation of epithelial cell apoptotic process
negative regulation of canonical Wnt signaling pathway
appendage development
limb development
embryonic limb morphogenesis
embryonic appendage morphogenesis
regulation of hepatocyte apoptotic process
pericardium development
retina homeostasis
protein-containing complex disassembly
regulation of microtubule polymerization
morphogenesis of a polarized epithelium
membraneless organelle assembly
positive regulation of striated muscle cell apoptotic process
positive regulation of cardiac muscle cell apoptotic process
positive regulation of intracellular transport
microtubule cytoskeleton organization involved in mitosis
axonogenesis
appendage morphogenesis
limb morphogenesis
regulation of cilium movement
photoreceptor cell maintenance
activin receptor signaling pathway
positive regulation of protein phosphorylation
establishment of planar polarity
establishment of tissue polarity
tissue homeostasis
anatomical structure homeostasis
protein localization to ciliary membrane
regulation of release of sequestered calcium ion into cytosol
digestive system development
positive regulation of cell cycle process
mitotic spindle organization
regulation of muscle cell apoptotic process
microtubule anchoring at centrosome
negative regulation of TORC1 signaling
digestive tract development
positive regulation of phosphorylation
regulation of developmental growth
muscle cell apoptotic process
regulation of epithelial cell proliferation
regulation of protein kinase activity
response to UV-A
microtubule anchoring at microtubule organizing center
regulation of Golgi organization
organelle transport along microtubule
regulation of ossification
regulation of protein localization to cell periphery
cranial skeletal system development
embryonic axis specification
regulation of muscle adaptation
dorsal/ventral axis specification
respiratory tube development
learning or memory
JNK cascade
positive regulation of plasma membrane bounded cell projection assembly
cerebral cortex development
neuron migration
negative regulation of growth
cellular pigmentation
peptidyl-serine phosphorylation
ureteric bud development
positive regulation of organelle assembly
regulation of nervous system development
thyroid gland development
mesonephric epithelium development
mesonephric tubule development
inner dynein arm assembly
regulation of neurogenesis
regulation of exocytosis
regulation of cardiac muscle cell apoptotic process
regulation of kinase activity
cardiocyte differentiation
positive regulation of epithelial to mesenchymal transition
trachea development
mesonephros development
positive regulation of carbohydrate metabolic process
cardiac muscle cell apoptotic process
regulation of striated muscle cell apoptotic process
regulation of protein localization to membrane
p.adjust
0.0035
0.0030
0.0025
GeneRatio
0.05
0.10
0.15
0.20
0.25
PC_3p
(455)
BB_T2
(464)
Cluster

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_locations_cilia_only_combined_GO_BP_dotplot_shared_te
```

Prepare data for clustering of shared enriched terms

```
# Split by location
pc_tip <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "PC_tip", ]
bb_tz <- comp_filtered@compareClusterResult[comp_filtered@compareClusterResult$Cluster == "BB_TZ", ]

# Create new compareClusterResult objects for each subset
comp_filtered_pc_tip <- comp_filtered
comp_filtered_bb_tz <- comp_filtered

comp_filtered_pc_tip@compareClusterResult <- pc_tip
comp_filtered_bb_tz@compareClusterResult <- bb_tz
```

Cluster results - Both locations

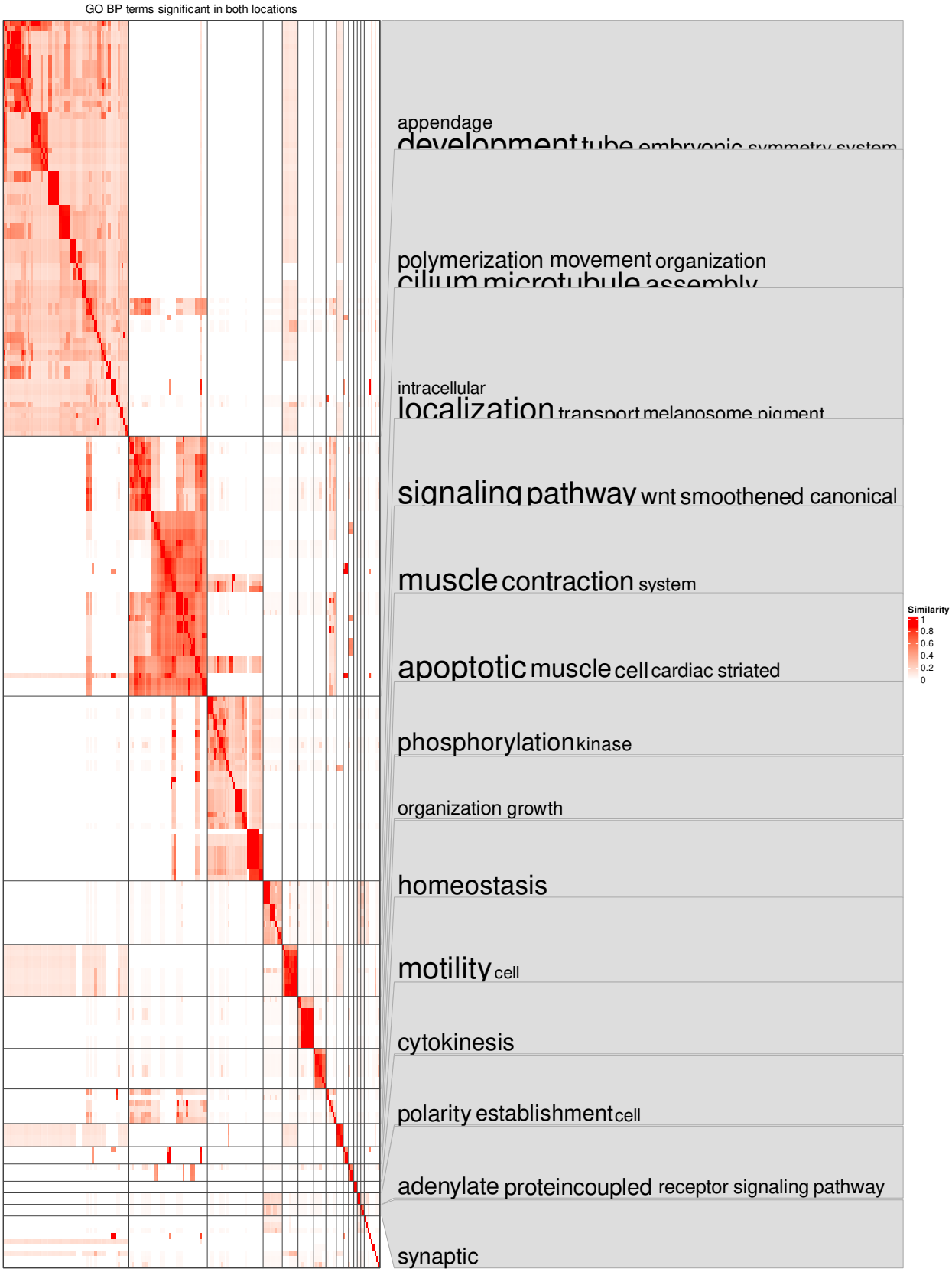
```
go_id = pc_tip$ID
mat = GO_similarity(go_id,
                    ont = 'BP',
                    db = 'org.Hs.eg.db',
                    measure = "Sim_Relevance_2006"
                    )
```

Plot cluster heatmap

```
# 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_locations_cilia_only_combined_GO_BP_dotplot_specific_
grid.draw(heatmap_plot)
dev.off()
```

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

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


GO BP terms significant in both locations
appendage
development tube embryonic symmetry system
polymerization movement organization
cilium microtubule assembly
intracellular
localization transport melanosome pigment
signaling pathway wnt smoothened canonical
muscle contraction system
apoptotic muscle cell cardiac striated
phosphorylation kinase
organization growth
homeostasis
motility cell
cytokinesis
polarity establishment cell
adenylate proteincoupled receptor signaling pathway
synaptic
Similarity
1
0.8
0.6
0.4
0.2
0

Process and save results

```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_pc_tip,
                              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_locations_cilia_only_combined_GO_B
```

```
# add cluster number from GO term clustering
results <- map_cluster_number(comp_filtered_bb_tz,
                              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_locations_cilia_only_combined_GO_B
```

Compare biological themes for different locations (cilia vs other parts of cell)

```
# Split data by location
input_genes <- list(
  PrimaryCiliaTip = gene_id_tip,
  PrimaryCilia = gene_id_pc,
  PrimaryCiliaTZ = gene_id_tz,
  BasalBody = gene_id_bb,
  Cytoplasm = gene_id_cytoplasm,
  Membrane = gene_id_membrane,
  Nucleus = gene_id_nucleus,
  Mitotic = gene_id_mitotic
)
```

GO BP enrichment analysis

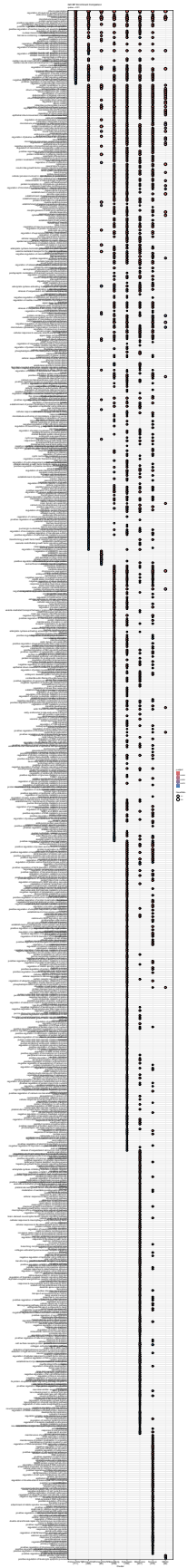
```
# Perform the compareCluster analysis
comp <- compareCluster(geneCluster = input_genes,
                       fun = "enrichGO",
                       OrgDb = org.Hs.eg.db,
```

```
                      keyType       = 'ENSEMBL',
                      ont           = "BP",
                      pAdjustMethod = "BH",
                      pvalueCutoff  = 0.01,
                      qvalueCutoff  = 0.01)
```

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

# 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_all_locations_GO_BP_result.csv"))
```

Dot plot of all enriched terms

```
dotplot(comp, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue < 0.01") +
  scale_y_discrete(labels = function(x) str_wrap(x, width = 60))
```



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

```r
# extract results
results <- comp@compareClusterResult$ID

# get unique terms
results <- unique(results)

# filter for p.adjust < 0.01 and qvalue < 0.01
results_signif <- comp@compareClusterResult %>% filter(p.adjust < 0.01 & qvalue < 0.01)
results_signif <- unique(results_signif$ID)

# print number of unique terms
print(paste("Number of unique terms in comp:", length(results)))
```

```
## [1] "Number of unique terms in comp: 941"
```

```r
print(paste("Number of unique terms in comp:", length(results_signif)))
```

```
## [1] "Number of unique terms in comp: 941"
```

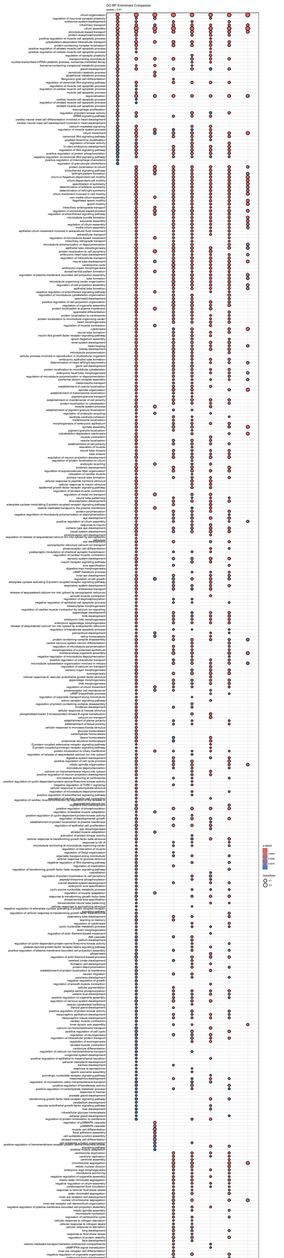
Filter for cilia terms

Here we filter for terms that are enriched in any of the four cilia locations, so basically cutting of the long dotplot above to only show the terms of the cilia locations.

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

# get all terms that are significant in any of the 4 non cilia locations
non_cilia_terms <- comp_results %>% filter(Cluster == "Cytoplasm" | Cluster == "Membrane" | Cluster ==

# get all the terms that are significant in any of the cilia locations
cilia_terms <- comp_results %>% filter(Cluster == "PrimaryCilia" | Cluster == "PrimaryCiliaTip" | Cluste

# remove all cilia terms from non_cilia_terms
non_cilia_only_terms <- non_cilia_terms %>% filter(!ID %in% cilia_terms$ID)

# remove all non cilia only terms from cilia terms
cilia_only_terms <- cilia_terms %>% filter(!ID %in% non_cilia_only_terms$ID)

# filter comp results for terms in cilia_only_terms
comp_results_filtered <- comp_results %>% filter(ID %in% cilia_only_terms$ID)

# create a copy of comp
comp_filtered <- comp

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

Dot plot of terms enriched in ciliary locations

```
# plot dotplot
dotplot(comp_filtered, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalue
  scale_y_discrete(labels = function(x) str_wrap(x, width = 80))
```


GO BP Enrichment Comparison
qvalue < 0.01

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_all_comparison_locations_GO_BP_dotplot_filtered_cilia
```

Filter for cilia only terms

Here we filter for terms that are only enriched in any of the four cilia locations and not in any of the other four non cilia locations. Note: This removes the non cilia locations from the plot as there is no enrichment of any of the terms in any non cilia location.

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

# get all terms that are significant in any of the 4 non cilia locations
non_cilia_terms <- comp_results %>% filter(Cluster == "Cytoplasm" | Cluster == "Membrane" | Cluster ==

# remove all non cilia terms from results
cilia_only_terms <- comp_results %>% filter(!ID %in% non_cilia_terms$ID)

# create a copy of comp
comp_filtered <- comp

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

Dot plot of terms only enriched in ciliary locations

```
# 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
xenobiotic catabolic process
glutathione metabolic process
Bergmann glial cell differentiation
regulation of muscle cell apoptotic process
regulation of cardiac muscle cell apoptotic process
muscle cell apoptotic process
cardiac muscle cell apoptotic process
regulation of striated muscle cell apoptotic process
striated muscle cell apoptotic process
macrophage proliferation
cardiac neural crest cell differentiation involved in heart development
cardiac neural crest cell development involved in heart development
positive regulation of macrophage chemotaxis
regulation of granulocyte chemotaxis
regulation of striated muscle contraction
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
sarcoplasmic reticulum calcium ion transport
regulation of cardiac muscle contraction
digestive tract morphogenesis
release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
mesenchyme morphogenesis
regulation of cardiac muscle contraction by calcium ion signaling
release of sequestered calcium ion into cytosol by endoplasmic reticulum
cellular response to hexose stimulus
cellular response to monosaccharide stimulus
glucose homeostasis
carbohydrate homeostasis
cellular response to carbohydrate stimulus
regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion
regulation of skeletal muscle adaptation
striated muscle adaptation
regulation of relaxation of muscle
regulation of Golgi organization
cellular response to glucose stimulus
regulation of ossification
embryonic axis specification
regulation of muscle adaptation
dorsal/ventral axis specification
negative regulation of adenylate cyclase-activating G protein-coupled receptor signaling pathway
learning or memory
brain morphogenesis
negative regulation of growth
cellular pigmentation
thyroid gland development
cardiac muscle contraction
striated muscle contraction
cardiocyte differentiation
urogenital system development
paraxial mesoderm development
trachea development
response to hexose
prostate gland development
cerebellum development
intracellular glucose homeostasis
regulation of p38MAPK cascade
p38MAPK cascade
muscle cell differentiation
focal adhesion assembly
cell-substrate junction assembly
striated muscle cell differentiation
skeletal muscle adaptation
inner ear receptor cell development
inner ear receptor cell stereocilium organization
microtubule nucleation
inner ear receptor cell differentiation
autophagosome maturation
establishment of spindle orientation
trachea morphogenesis
mechanoreceptor differentiation
regulation of protein localization to centrosome
negative regulation of cardiocyte differentiation
establishment of spindle localization
cognition
early endosome to late endosome transport
regulation of macroautophagy
spindle localization
positive regulation of cholesterol biosynthetic process
positive regulation of sterol biosynthetic process
placenta blood vessel development
positive regulation of miRNA metabolic process
cholesterol biosynthetic process
secondary alcohol biosynthetic process
peptidyl-glutamic acid modification
negative regulation of activin receptor signaling pathway
adrenergic receptor signaling pathway
regulation of cardiocyte differentiation
regulation of purine nucleotide metabolic process
protein localization to postsynaptic specialization membrane
neurotransmitter receptor localization to postsynaptic specialization membrane
regulation of carbohydrate catabolic process
animal organ formation
positive regulation of epithelial cell proliferation
regulation of nucleotide metabolic process
regulation of early endosome to late endosome transport
cerebellar cortex development
establishment of mitotic spindle orientation
positive regulation of ATP metabolic process
ERBB3 signaling pathway
PrimaryCiliaTip (111)
PrimaryCilia (455)
PrimaryCiliaTZ (93)
BasalBody (408)
Cluster
p.adjust
0.0025
0.0050
0.0075
GeneRatio
0.025
0.050
0.075
0.100

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_all_comparison_locations_GO_BP_dotplot_filtered_cilia
```

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

# get unique terms
results <- unique(results)

# print number of unique terms
print(paste("Number of unique terms in comp:", length(results)))
```

```
## [1] "Number of unique terms in comp: 941"
```

Simplify results

```
comp_simplified <- simplify(comp, cutoff=0.6, by="p.adjust", select_fun=min)

# print number of unique terms
print(paste("Number of signficant terms:", nrow(comp)))
```

```
## [1] "Number of signficant terms: 2172"
```

```
print(paste("Number of significant terms after simplifying:", nrow(comp_simplified)))
```

```
## [1] "Number of significant terms after simplifying: 462"
```

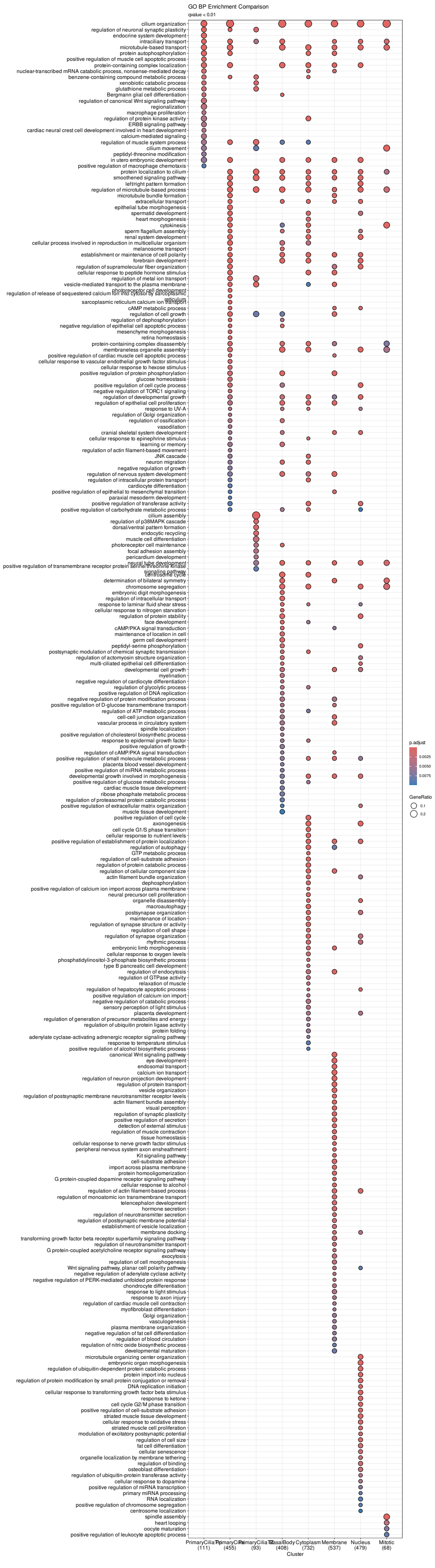
```
# extract results
results <- comp_simplified@compareClusterResult$ID

# get unique terms
results <- unique(results)

# print number of unique terms
print(paste("Number of unique terms in comp_simplified:", length(results)))
```

```
## [1] "Number of unique terms in comp_simplified: 256"
```

```
# get results
results <- comp_simplified@compareClusterResult

# 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_all_locations_GO_BP_result_simplif
```

Dot plot of all enriched terms (simplified)

```r
# plot dotplot
dotplot(comp_simplified, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle = "qvalu
  scale_y_discrete(labels = function(x) str_wrap(x, width = 80))
```


GO BP Enrichment Comparison
qvalue < 0.01
PrimaryCilia
(111)
PrimaryCilia
(455)
PrimaryCilia
(93)
BasalBody
(408)
Cytoplasm
(732)
Membrane
(537)
Nucleus
(479)
Mitotic
(68)
Cluster
p.adjust
0.0025
0.0050
0.0075
GeneRatio
0.1
0.2

Filter for cilia terms

Here we filter for terms that are enriched in any of the four cilia locations, so basically cutting of the long dotplot above to only show the terms of the cilia locations. But this time using the simplified results.

```
 # get results as data frame
comp_results <- comp_simplified@compareClusterResult

# get all terms that are significant in any of the 4 non cilia locations
non_cilia_terms <- comp_results %>% filter(Cluster == "Cytoplasm" | Cluster == "Membrane" | Cluster ==

# get all the terms that are significant in any of the cilia locations
cilia_terms <- comp_results %>% filter(Cluster == "PrimaryCilia" | Cluster == "PrimaryCiliaTip" | Cluste

# remove all cilia terms from non_cilia_terms
non_cilia_only_terms <- non_cilia_terms %>% filter(!ID %in% cilia_terms$ID)

# remove all non cilia only terms from cilia terms
cilia_only_terms <- cilia_terms %>% filter(!ID %in% non_cilia_only_terms$ID)

# filter comp results for terms in cilia_only_terms
comp_results_filtered <- comp_results %>% filter(ID %in% cilia_only_terms$ID)

# create a copy of comp
comp_simplified_filtered <- comp_simplified

# update results in comp
comp_simplified_filtered@compareClusterResult <- comp_results_filtered
```

Dot plot of terms enriched in ciliary locations (simplified)

```
# plot dotplot
dotplot(comp_simplified_filtered, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitle
  scale_y_discrete(labels = function(x) str_wrap(x, width = 80))
```


GO BP Enrichment Comparison
qvalue < 0.01
cilium organization
regulation of neuronal synaptic plasticity
endocrine system development
intraciliary transport
microtubule-based transport
protein autophosphorylation
positive regulation of muscle cell apoptotic process
protein-containing complex localization
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
benzene-containing compound metabolic process
xenobiotic catabolic process
glutathione metabolic process
Bergmann glial cell differentiation
regulation of canonical Wnt signaling pathway
regionalization
macrophage proliferation
regulation of protein kinase activity
ERBB signaling pathway
cardiac neural crest cell development involved in heart development
calcium-mediated signaling
regulation of muscle system process
cilium movement
peptidyl-threonine modification
in utero embryonic development
positive regulation of macrophage chemotaxis
protein localization to cilium
smoothened signaling pathway
left/right pattern formation
regulation of microtubule-based process
microtubule bundle formation
extracellular transport
epithelial tube morphogenesis
spermatid development
heart morphogenesis
cytokinesis
sperm flagellum assembly
renal system development
cellular process involved in reproduction in multicellular organism
melanosome transport
establishment or maintenance of cell polarity
forebrain development
regulation of supramolecular fiber organization
cellular response to peptide hormone stimulus
regulation of metal ion transport
vesicle-mediated transport to the plasma membrane
photoreceptor cell development
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
sarcoplasmic reticulum calcium ion transport
cAMP metabolic process
regulation of cell growth
regulation of dephosphorylation
negative regulation of epithelial cell apoptotic process
mesenchyme morphogenesis
retina homeostasis
protein-containing complex disassembly
membraneless organelle assembly
positive regulation of cardiac muscle cell apoptotic process
cellular response to vascular endothelial growth factor stimulus
cellular response to hexose stimulus
positive regulation of protein phosphorylation
glucose homeostasis
positive regulation of cell cycle process
negative regulation of TORC1 signaling
regulation of developmental growth
regulation of epithelial cell proliferation
response to UV-A
regulation of Golgi organization
regulation of ossification
vasodilation
cranial skeletal system development
cellular response to epinephrine stimulus
learning or memory
regulation of actin filament-based movement
JNK cascade
neuron migration
negative regulation of growth
regulation of nervous system development
regulation of intracellular protein transport
cardiocyte differentiation
positive regulation of epithelial to mesenchymal transition
paraxial mesoderm development
positive regulation of transferase activity
positive regulation of carbohydrate metabolic process
cilium assembly
regulation of p38MAPK cascade
dorsal/ventral pattern formation
endocytic recycling
muscle cell differentiation
photoreceptor cell maintenance
focal adhesion assembly
pericardium development
neural tube development
positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
centrosome cycle
determination of bilateral symmetry
chromosome segregation
embryonic digit morphogenesis
regulation of intracellular transport
response to laminar fluid shear stress
cellular response to nitrogen starvation
regulation of protein stability
face development
cAMP/PKA signal transduction
maintenance of location in cell
germ cell development
peptidyl-serine phosphorylation
postsynaptic modulation of chemical synaptic transmission
regulation of actomyosin structure organization
multi-ciliated epithelial cell differentiation
developmental cell growth
myelination
negative regulation of cardiocyte differentiation
regulation of glycolytic process
positive regulation of DNA replication
negative regulation of protein modification process
positive regulation of D-glucose transmembrane transport
regulation of ATP metabolic process
cell-cell junction organization
vascular process in circulatory system
spindle localization
positive regulation of cholesterol biosynthetic process
response to epidermal growth factor
positive regulation of growth
regulation of cAMP/PKA signal transduction
positive regulation of small molecule metabolic process
placenta blood vessel development
positive regulation of miRNA metabolic process
developmental growth involved in morphogenesis
positive regulation of glucose metabolic process
cardiac muscle tissue development
ribose phosphate metabolic process
regulation of proteasomal protein catabolic process
positive regulation of extracellular matrix organization
muscle tissue development
PrimaryCilia (111)
PrimaryCilia (455)
PrimaryCilia (93)
BasalBody (408)
Cytoplasm (732)
Membrane (537)
Nucleus (479)
Mitotic (68)
Cluster
p.adjust
0.0025
0.0050
0.0075
GeneRatio
0.1
0.2

```r
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_all_comparison_locations_GO_BP_dotplot_simplified_06_
```

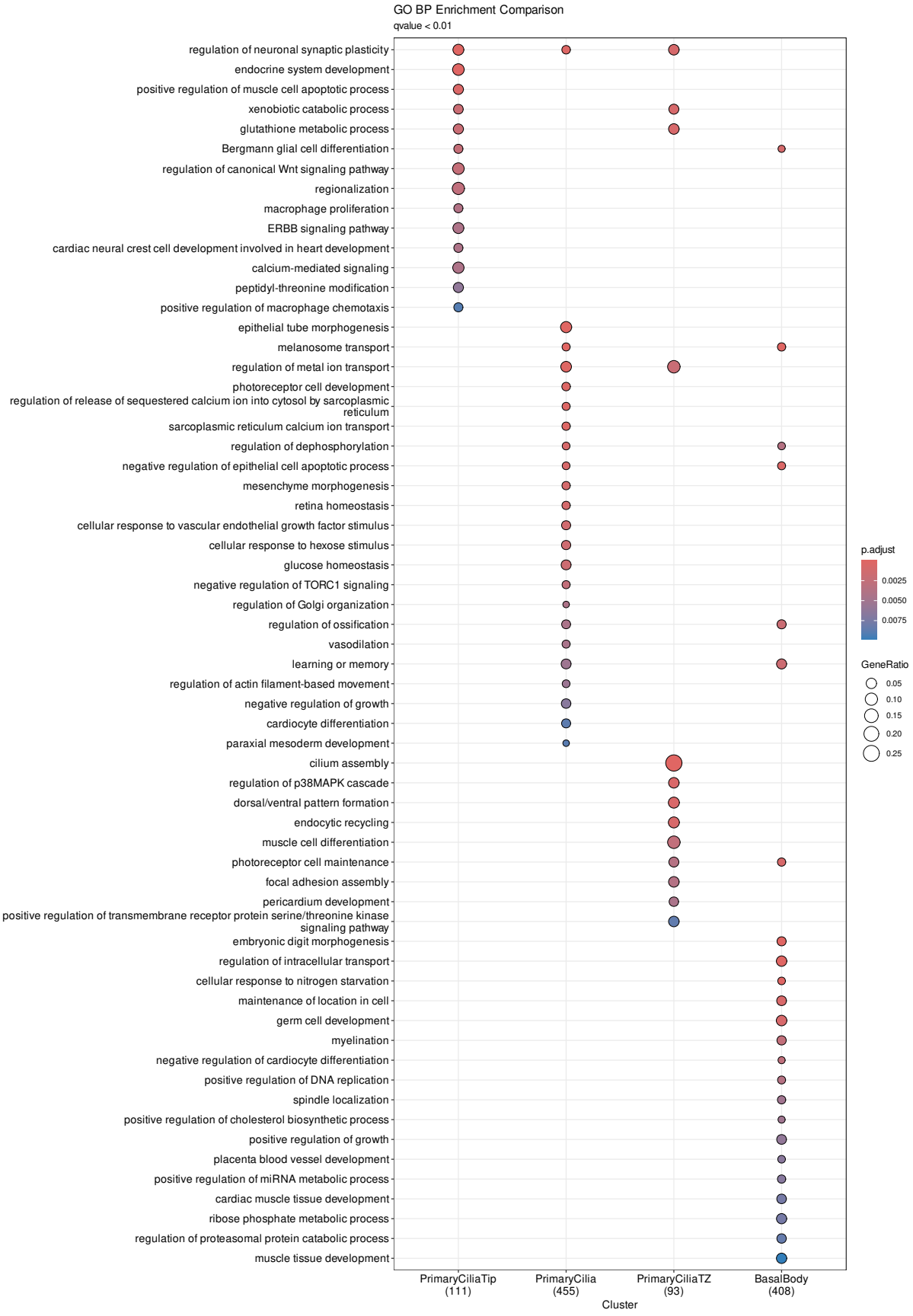
Filter for cilia only terms

Here we filter again for terms that are only enriched in any of the four cilia locations and not in any of the other four non cilia locations. But this time using the simplified results. Note: This removes the non cilia locations from the plot as there is no enrichment of any of the terms in any non cilia location.

```r
 # get results as data frame
comp_results <- comp_simplified@compareClusterResult

# get all terms that are significant in any of the 4 non cilia locations
non_cilia_terms <- comp_results %>% filter(Cluster == "Cytoplasm" | Cluster == "Membrane" | Cluster == 

# remove all non cilia terms from results
cilia_only_terms <- comp_results %>% filter(!ID %in% non_cilia_terms$ID)

# create a copy of comp
comp_simplified_filtered <- comp_simplified

# update results in comp
comp_simplified_filtered@compareClusterResult <- cilia_only_terms
```

Dot plot of terms only enriched in ciliary locations(simplified)

```r
# plot dotplot
dotplot(comp_simplified_filtered, showCategory = NULL) + ggtitle("GO BP Enrichment Comparison", subtitl
  scale_y_discrete(labels = function(x) str_wrap(x, width = 80))
```


GO BP Enrichment Comparison
qvalue < 0.01
regulation of neuronal synaptic plasticity
endocrine system development
positive regulation of muscle cell apoptotic process
xenobiotic catabolic process
glutathione metabolic process
Bergmann glial cell differentiation
regulation of canonical Wnt signaling pathway
regionalization
macrophage proliferation
ERBB signaling pathway
cardiac neural crest cell development involved in heart development
calcium-mediated signaling
peptidyl-threonine modification
positive regulation of macrophage chemotaxis
epithelial tube morphogenesis
melanosome transport
regulation of metal ion transport
photoreceptor cell development
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum
sarcoplasmic reticulum calcium ion transport
regulation of dephosphorylation
negative regulation of epithelial cell apoptotic process
mesenchyme morphogenesis
retina homeostasis
cellular response to vascular endothelial growth factor stimulus
cellular response to hexose stimulus
glucose homeostasis
negative regulation of TORC1 signaling
regulation of Golgi organization
regulation of ossification
vasodilation
learning or memory
regulation of actin filament-based movement
negative regulation of growth
cardiocyte differentiation
paraxial mesoderm development
cilium assembly
regulation of p38MAPK cascade
dorsal/ventral pattern formation
endocytic recycling
muscle cell differentiation
photoreceptor cell maintenance
focal adhesion assembly
pericardium development
positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
embryonic digit morphogenesis
regulation of intracellular transport
cellular response to nitrogen starvation
maintenance of location in cell
germ cell development
myelination
negative regulation of cardiocyte differentiation
positive regulation of DNA replication
spindle localization
positive regulation of cholesterol biosynthetic process
positive regulation of growth
placenta blood vessel development
positive regulation of miRNA metabolic process
cardiac muscle tissue development
ribose phosphate metabolic process
regulation of proteasomal protein catabolic process
muscle tissue development
PrimaryCiliaTip (111)
PrimaryCilia (455)
PrimaryCiliaTZ (93)
BasalBody (408)
Cluster
p.adjust
0.0025
0.0050
0.0075
GeneRatio
0.05
0.10
0.15
0.20
0.25

```
# save dotplot as svg file
ggsave(paste0(out_path, "NonRestricted_comparison_all_comparison_locations_GO_BP_dotplot_simplified_06_
```

```
# calculate overlap between comp and comp_simplified results by plotting a venn diagramm
# extract results
results <- comp@compareClusterResult$ID
results_simplified <- comp_simplified@compareClusterResult$ID


# print number of unique terms
print(paste("Number of terms in comp:", length(results)))
```

Check overlap between significant terms and simplified terms

```
## [1] "Number of terms in comp: 2172"
```

```
print(paste("Number of terms in comp_simplified:", length(results_simplified)))
```

```
## [1] "Number of terms in comp_simplified: 462"
```

```
# get unique terms
results <- unique(results)
results_simplified <- unique(results_simplified)

# print number of unique terms
print(paste("Number of unique terms in comp:", length(results)))
```

```
## [1] "Number of unique terms in comp: 941"
```

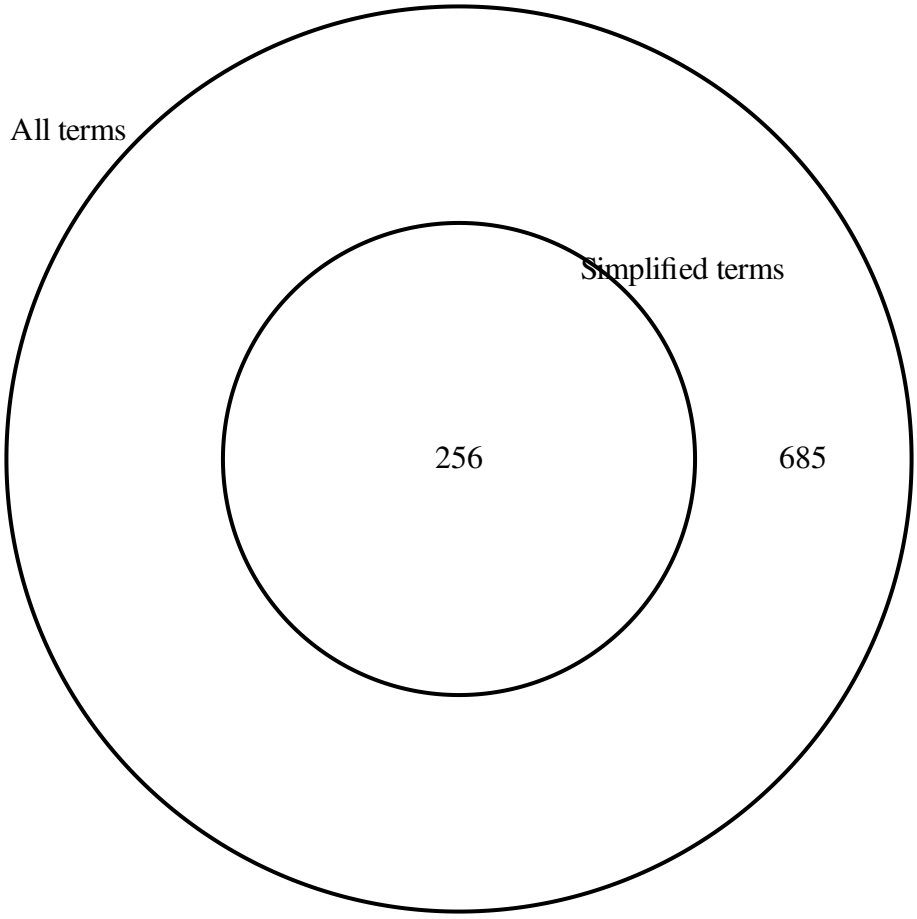
```
print(paste("Number of unique terms in comp_simplified:", length(results_simplified)))
```

```
## [1] "Number of unique terms in comp_simplified: 256"
```

```
intersection <- intersect(results, results_simplified)
print(paste("Number of terms in intersection:", length(intersection)))
```

```
## [1] "Number of terms in intersection: 256"
```

```
# plot venn diagram
venn.plot <- venn.diagram(
  x = list(results = results, results_simplified = results_simplified),
  category.names = c("All terms", "Simplified terms"),
  filename = NULL,
  output = FALSE
)

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



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] simplifyEnrichment_2.2.0 VennDiagram_1.7.3        futile.logger_1.4.3
##  [4] svglite_2.1.3            Cairo_1.6-2              org.Hs.eg.db_3.21.0
##  [7] AnnotationDbi_1.70.0     IRanges_2.42.0           S4Vectors_0.46.0
## [10] Biobase_2.68.0           BiocGenerics_0.54.0      generics_0.1.3
## [13] enrichplot_1.28.2        clusterProfiler_4.16.0   lubridate_1.9.4
## [16] forcats_1.0.0            stringr_1.5.1            dplyr_1.1.4
## [19] purrr_1.0.4              readr_2.1.5              tidyr_1.3.1
## [22] tibble_3.2.1             ggplot2_3.5.2            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          gridGraphics_0.5-1
##  [19] plyr_1.8.9              futile.options_1.0.1    cachem_1.1.0
##  [22] igraph_2.1.4            mime_0.13               lifecycle_1.0.4
##  [25] iterators_1.0.14        pkgconfig_2.0.3         Matrix_1.7-3
##  [28] R6_2.6.1                fastmap_1.2.0           gson_0.1.0
##  [31] GenomeInfoDbData_1.2.14 shiny_1.10.0            clue_0.3-66
##  [34] digest_0.6.37           aplot_0.2.5             colorspace_2.1-1
##  [37] patchwork_1.3.0         textshaping_1.0.1       RSQLite_2.3.11
##  [40] labeling_0.4.3          timechange_0.3.0        httr_1.4.7
##  [43] compiler_4.5.0          bit64_4.6.0-1           withr_3.0.2
##  [46] doParallel_1.0.17       BiocParallel_1.42.0     DBI_1.2.3
##  [49] R.utils_2.13.0          scatterplot3d_0.3-44    rjson_0.2.23
##  [52] tools_4.5.0             ape_5.8-1               flexclust_1.5.0
##  [55] httpuv_1.6.16           R.oo_1.27.1             glue_1.8.0
##  [58] promises_1.3.2          nlme_3.1-168            GOSemSim_2.34.0
##  [61] cluster_2.1.8.1         reshape2_1.4.4          fgsea_1.34.0
##  [64] gtable_0.3.6            tzdb_0.5.0              class_7.3-23
##  [67] R.methodsS3_1.8.2       data.table_1.17.0       hms_1.1.3
##  [70] xml2_1.3.8              XVector_0.48.0          ggrepel_0.9.6
##  [73] foreach_1.5.2           pillar_1.10.2           yulab.utils_0.2.0
##  [76] later_1.4.2             circlize_0.4.16         splines_4.5.0
##  [79] treeio_1.32.0           lattice_0.22-5          bit_4.6.0
##  [82] tidyselect_1.2.1        GO.db_3.21.0            ComplexHeatmap_2.24.0
##  [85] tm_0.7-16               Biostrings_2.76.0       knitr_1.50
##  [88] NLP_0.3-2               xfun_0.52               matrixStats_1.5.0
##  [91] stringi_1.8.7           UCSC.utils_1.4.0        lazyeval_0.2.2
##  [94] ggfun_0.1.8             yaml_2.3.10             evaluate_1.0.3
##  [97] codetools_0.2-20        qvalue_2.40.0           Polychrome_1.5.4
## [100] ggplotify_0.1.2         cli_3.6.5               xtable_1.8-4
## [103] systemfonts_1.2.3       Rcpp_1.0.14             GenomeInfoDb_1.44.0
## [106] png_0.1-8               parallel_4.5.0          simona_1.6.0
```

```
## [109] blob_1.2.4             DOSE_4.2.0             slam_0.1-55
## [112] tidytree_0.4.6         scales_1.4.0           crayon_1.5.3
## [115] GetoptLong_1.0.5       rlang_1.1.6            cowplot_1.1.3
## [118] fastmatch_1.1-6        KEGGREST_1.48.0        formatR_1.14
```