GTEx data analysis

ZHAO Kai

Here is a brief introduction of analyzing the results of our proposed approach on GTEx. In this analysis, we considered to model the variance contributed by gender, brain regions, and genes.

Preparations

```
suppressPackageStartupMessages({
  require(ggplot2)
  require(formatR)
  require(knitr)
  require(cluster)
  require(factoextra)
  require(dplyr)
  require(RColorBrewer)
  require(clusterProfiler)
  require(org.Hs.eg.db)
  require(enrichplot)
  require(stringr)
  require(forcats)
  require(DOSE)
  require(extrafont)
})

truncated_var <- function(x){
    remove_idx <- c(which.max(x), which.min(x))
    var(x[-remove_idx])
}
wrap_labal <- function(x, width = 60){
    str_wrap(x, width=60)
}
```

```
loadfonts(quiet = T)
setwd("~/data/Results/gtex")
# load("GTEX_l1l2_penalty_13.RData")
# str(fitted_obj)
# attach(fitted_obj)

load("insider_gtex_fitted_object.RData")
str(object)
```

```
## List of 7
##  $ data           : num [1:988, 1:49999] 0.0229 0 0.0216 0.0204 0.0149 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:988] "GTEX-WZTO-0011-R4A-SM-3NMC7" "GTEX-1HFI6-0011-R4b-SM-A96RT" "GTEX-12WSA-0011
```

```
##   .. ..$ : chr [1:49999] "DDX11L1" "WASH7P" "MIR6859-1" "MIR1302-2HG" ...
##  $ confounder     : num [1:988, 1:3] 1 1 1 1 1 1 1 1 1 1 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:988] "GTEX-WZTO-0011-R4A-SM-3NMC7" "GTEX-1HFI6-0011-R4b-SM-A96RT" "GTEX-12WSA-0011
##   .. ..$ : chr [1:3] "gender" "structure" "interaction_indicator"
##  $ train_indicator: int [1:988, 1:49999] 1 0 1 1 1 1 1 1 1 1 ...
##  $ params         :List of 4
##   ..$ global_tol : num 1e-10
##   ..$ sub_tol    : num 1e-05
##   ..$ tuning_iter: num 30
##   ..$ max_iter   : num 50000
##  $ cfd_matrices   :List of 3
##   ..$ factor0: num [1:2, 1:12] -0.39 -0.25 -0.786 -0.872 0.214 ...
##   ..$ factor1: num [1:13, 1:12] -0.6083 -0.5394 4.4236 1.269 -0.0841 ...
##   ..$ factor2: num [1:26, 1:12] -0.3526 -0.2556 -0.5653 0.0259 2.4555 ...
##  $ column_factor  : num [1:12, 1:49999] 0 0 0 0 0 ...
##  $ test_rmse      : num 4.65e-310
##  - attr(*, "class")= chr "insider"
```

```
attach(object)

gender_factor <- cfd_matrices[[1]]
tissue_factor <- cfd_matrices[[2]]
interaction <- cfd_matrices[[3]]

donor_info <- read.csv("meta.csv", stringsAsFactors = F)

load("included_gene_mapping.RData")
ensg_id <- gsub(included_gene_mapping[[1]], pattern="\\.[0-9]+$", replacement="")
meta <- data.frame(ensg_id, included_gene_mapping, stringsAsFactors = F)

cat("Overall age distribution:\n")
```

```
## Overall age distribution:
```

```
print(table(donor_info[,c(3,4)]))
```

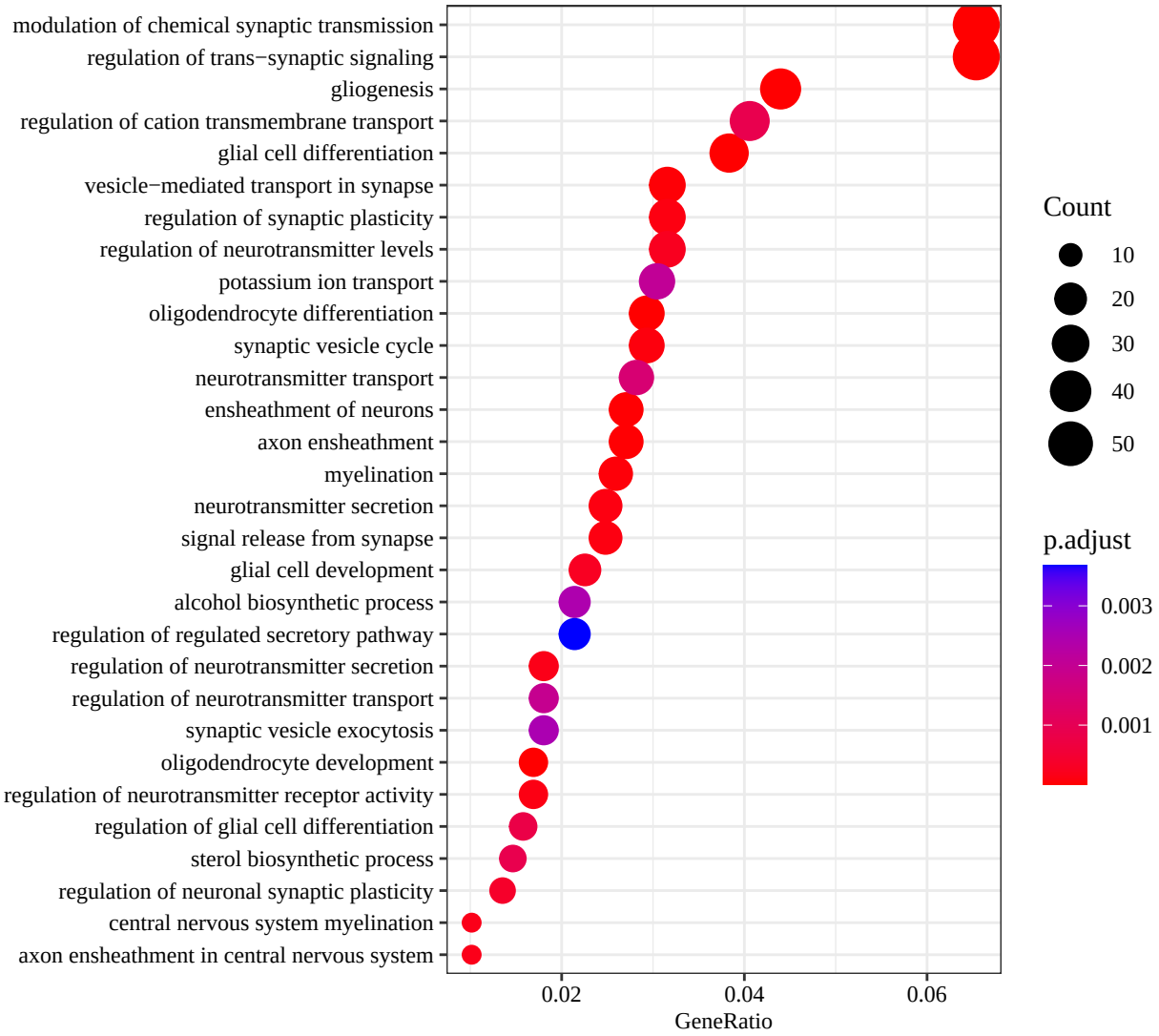
```
##    AGE
## SEX 20-29 30-39 40-49 50-59 60-69 70-79
##   1    15     3    41   149   248    38
##   2    16    17    62   121   256    22
```

Investigate the difference in BPs between both genders across different brain regions

We considered all metagenes in this analysis. The figure below shows the up-regulated BPs of genes with the greatest positive difference between male and female.

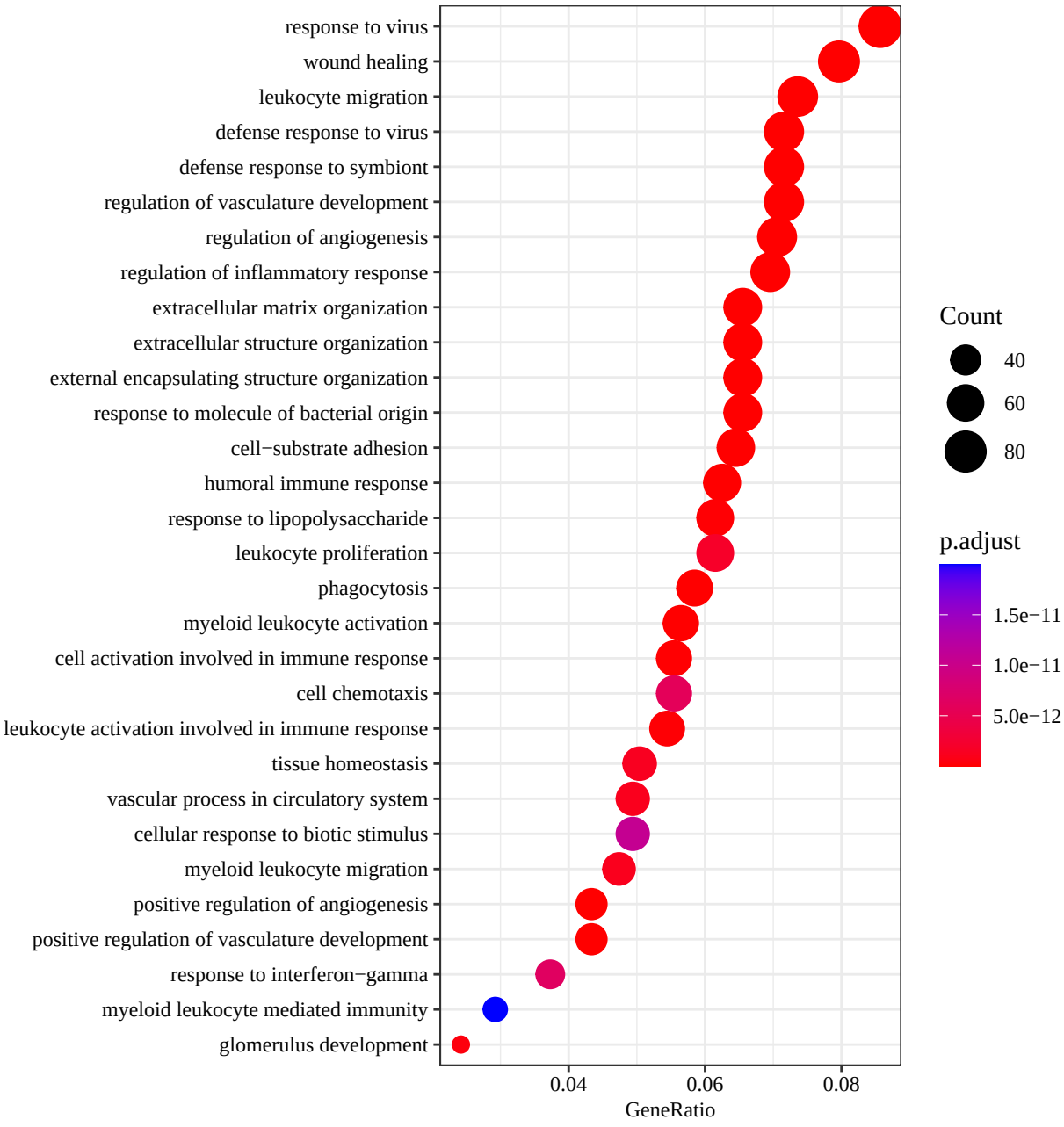
```
metagene_var <- apply(gender_factor, 2, function(x) abs(x[1] - x[2]))
ord <- order(metagene_var, decreasing = T)

k <- 12
gender_expr_proflies <- gender_factor[, ord[1:k]] %*% column_factor[ord[1:k],]
colnames(gender_expr_proflies) <- colnames(data)

# The first row stands for the male and the second for the female.
```

```r
diff <- gender_expr_proflies[1, ] - gender_expr_proflies[2, ]

cutoffs <- quantile(diff, probs = seq(0, 1, 0.025))
# selected <- (diff <= cutoffs[2]) # greatest negative difference
selected <- (diff >= cutoffs[length(cutoffs)-1])

upreg <- enrichGO(gene      = unique(meta[selected,1]),
                  OrgDb     = 'org.Hs.eg.db',
                  keyType   = 'ENSEMBL',
                  ont       = "BP",
                  readable  = TRUE)

# summary(upreg)
dotplot(upreg, font.size = 9, showCategory=30, label_format = 60) +
  theme(text=element_text(family="Times New Roman"))
```



The figure below shows the down-regulated BPs of genes with the greatest negative difference between male and female.

```r
k <- 12
gender_expr_proflies <- gender_factor[, ord[1:k]] %*% column_factor[ord[1:k],]
diff <- gender_expr_proflies[1, ] - gender_expr_proflies[2, ]

cutoffs <- quantile(diff, probs = seq(0, 1, 0.025))
selected <- (diff <= cutoffs[2]) # greatest negative difference

downreg <- enrichGO(gene          = unique(meta[selected,1]),
                    OrgDb         = 'org.Hs.eg.db',
                    keyType       = 'ENSEMBL',
                    ont           = "BP",
                    readable      = TRUE)
dotplot(downreg, font.size = 9, showCategory=30, label_format = 60) +
  theme(text=element_text(family="Times New Roman"))
```



Investigate BPs invloving in the interaction between genders and different brain regions

In the analysis, we explored the BPs of genes showing the difference in the brain region in which there is the greatest difference in correlation between both gender.

```
combinations <- unique(object$confounder)

meta_info <- inner_join(data.frame(SAMPID = row.names(data)), donor_info)
```

```
## Joining, by = "SAMPID"
```

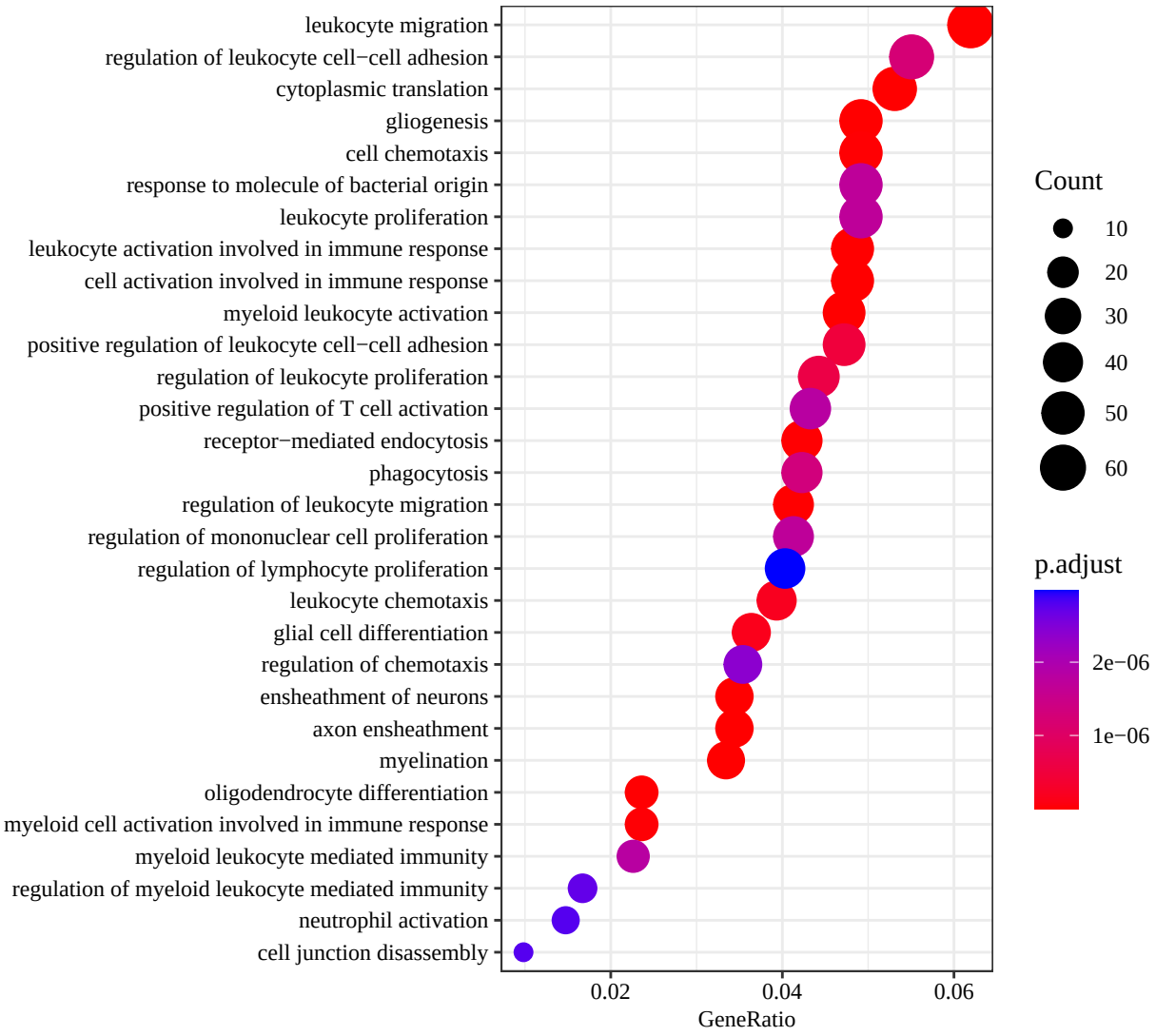
```
mapping <- cbind(meta_info[,c(3,7)], object$confounder)
```

```r
distantces <- sapply(seq(13), function(i){
  interaction_idx <- which(combinations[,2] == i)
  selected <- interaction[interaction_idx, ]
  distance <- mean((selected[1, ] - selected[2, ])^2)
})

idx <- which.max(distantces)
cat("brain region name:", unique(mapping[,c(2,4)])[idx, 1], "\n")
```

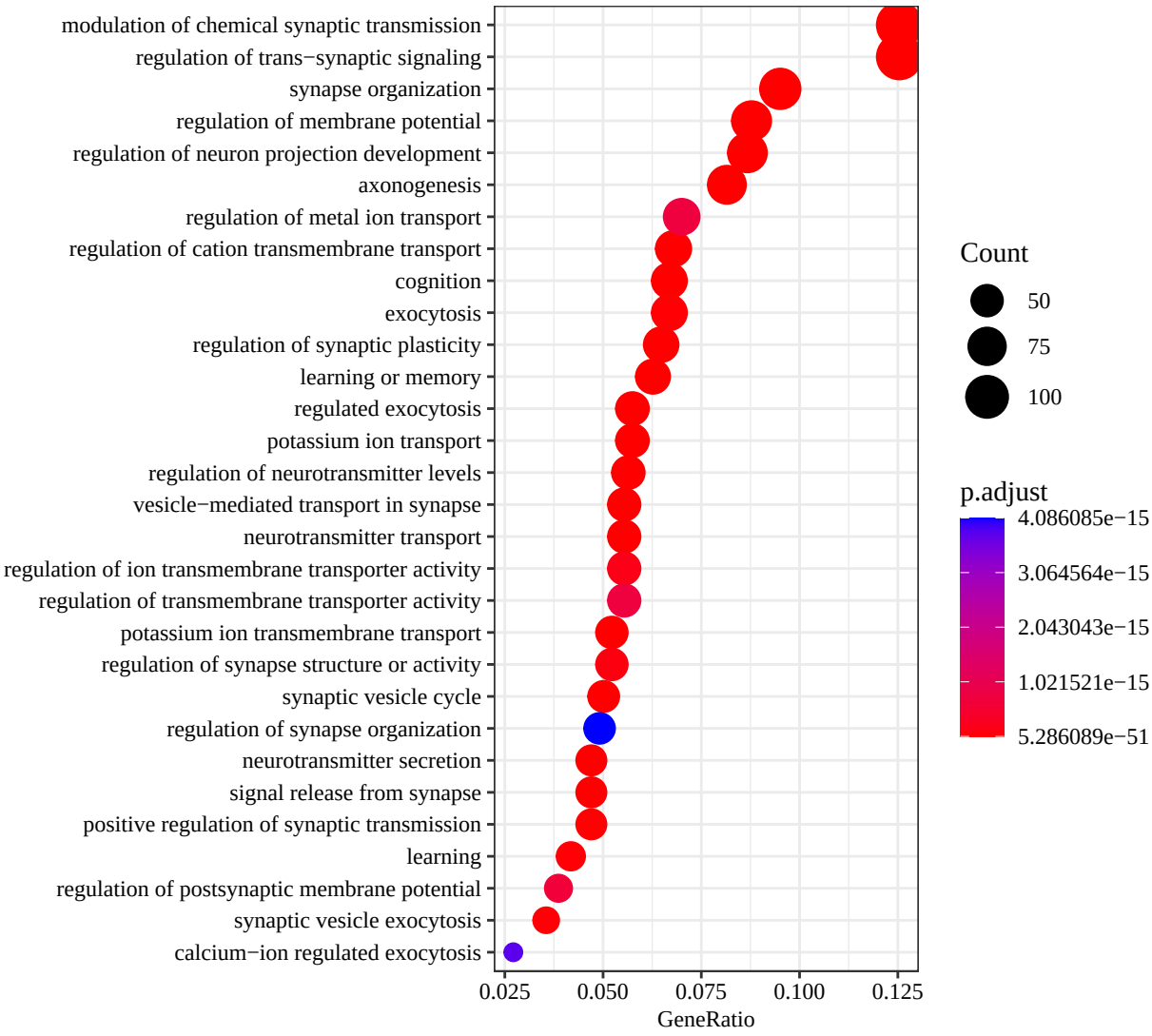
```
## brain region name: Brain - Frontal Cortex (BA9)
```

```r
selected <- combinations[which(combinations[,2] == idx),]
```

```r
indices <- order(apply(interaction[selected[,3], ], 2, var), decreasing = T)[1:2]
interaction_profiles <- interaction[selected[,3], indices] %*% column_factor[indices,]
diff <- interaction_profiles[1, ] - interaction_profiles[2, ]

cutoffs <- quantile(diff, probs = seq(0, 1, 0.025))
selected_idx <- (diff <= cutoffs[2]) # greatest negative difference

downreg <- enrichGO(gene      = unique(meta[selected_idx,1]),
                    OrgDb       = 'org.Hs.eg.db',
                    keyType     = 'ENSEMBL',
                    ont         = "BP",
                    readable    = TRUE)
dotplot(downreg, font.size = 9, showCategory=30, label_format = 60) +
  theme(text=element_text(family="Times New Roman"))
```




```r
# interaction_profiles <- interaction[, ord[1:k]] %*% column_factor[ord[1:k],]
# diff <- interaction_profiles[1, ] - interaction_profiles[2, ]
selected_idx <- (diff >= cutoffs[length(cutoffs)-1])

upreg <- enrichGO(gene       = unique(meta[selected_idx,1]),
                  OrgDb      = 'org.Hs.eg.db',
                  keyType    = 'ENSEMBL',
                  ont        = "BP",
                  readable   = TRUE)
dotplot(upreg, font.size = 9, showCategory=30, label_format = 60) +
  theme(text=element_text(family="Times New Roman"))
```


modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
synapse organization
regulation of membrane potential
regulation of neuron projection development
axonogenesis
regulation of metal ion transport
regulation of cation transmembrane transport
cognition
exocytosis
regulation of synaptic plasticity
learning or memory
regulated exocytosis
potassium ion transport
regulation of neurotransmitter levels
vesicle−mediated transport in synapse
neurotransmitter transport
regulation of ion transmembrane transporter activity
regulation of transmembrane transporter activity
potassium ion transmembrane transport
regulation of synapse structure or activity
synaptic vesicle cycle
regulation of synapse organization
neurotransmitter secretion
signal release from synapse
positive regulation of synaptic transmission
learning
regulation of postsynaptic membrane potential
synaptic vesicle exocytosis
calcium−ion regulated exocytosis
0.025
0.050
0.075
0.100
0.125
GeneRatio
Count
50
75
100
p.adjust
4.086085e−15
3.064564e−15
2.043043e−15
1.021521e−15
5.286089e−51