Pathway Analysis - Treg cells
Input data: DEGs

Elena Eyre Sánchez, PhD

2025-01-04

Contents

1 Introducción y Objetivo

En este script realizo un pathways análisis con los DEGs obtenidos de los Venn Diagrams, y preparo el resultado en archivos gem para extraer un gráfico posteriormente en Cytoscape (https://cran.r-project.org/web/packages/gprofiler2/vignettes/gprofiler2.html).

2 Paquetes y datos

3 Datos

Cargo inicialmente los listados de células Treg obtenidas en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_Tregcells_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
head(jVenn_Tregcells_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509 GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>    <chr>       <chr>
## 1 H3C4     RPS16       LINC01091
## 2 H2BC6    POLR2I      CTLA4
## 3 OR10K2   RNPS1       CYP27B1
## 4 CEMIP    RPS29       CLDND1
## 5 JSRP1    RPL10       VMA21
## 6 PIM2     HCFC1       OGT
```

```
jVenn_Tregcells_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_T
jVenn_Tregcells_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Treg
jVenn_Tregcells_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Treg
jVenn_Tregcells_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Treg
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_Tregcells <- read_delim("~/Desktop/ELENA_UOC/TFM/DEGs/Venn_diagrams_cell_type_treatments_comparis
    delim = ",", escape_double = FALSE, trim_ws = TRUE)
```

4 Uncovered

En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

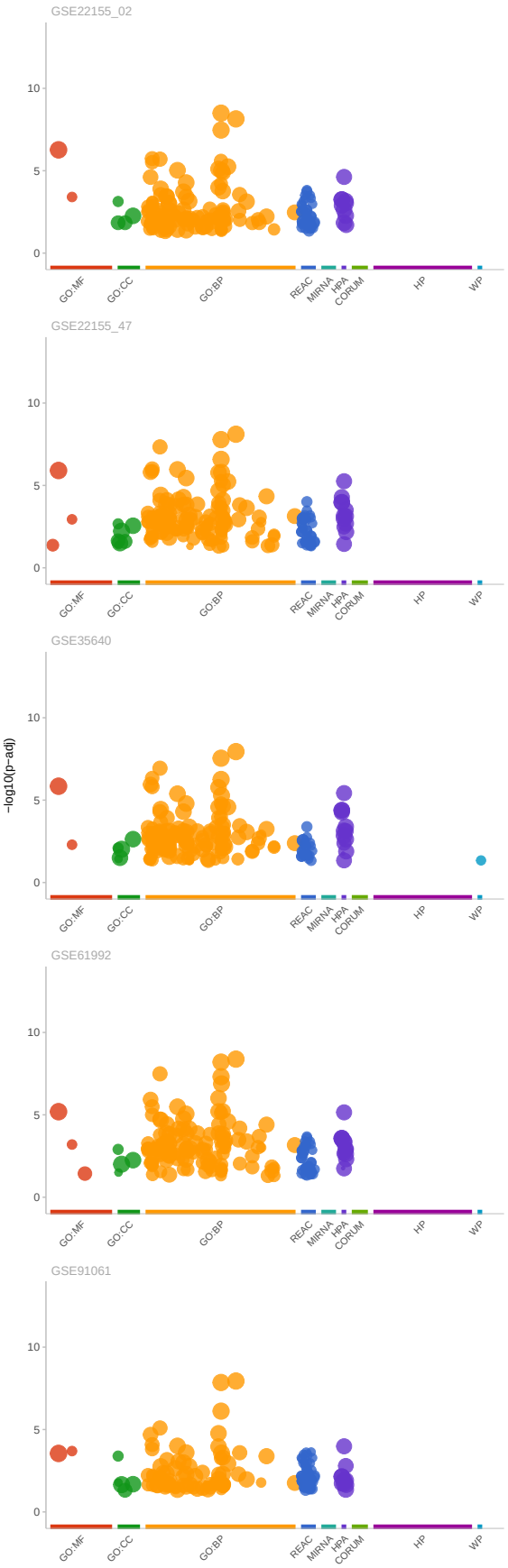
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_Tregcells_vs_GSE22155_02$GSE50509[which
                                   "GSE22155_47" = jVenn_Tregcells_vs_GSE22155_47$GSE50509[which(is.na(jVenn_T
                                   "GSE35640" = jVenn_Tregcells_vs_GSE35640$GSE50509[which(is.na(jVenn_Tregce
                                   "GSE91061" = jVenn_Tregcells_vs_GSE91061$GSE50509[which(is.na(jVenn_Tregce
                                   "GSE61992" = jVenn_Tregcells_vs_GSE61992$GSE50509[which(is.na(jVenn_Tregce
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##         query significant      p_value term_size query_size intersection_size
## 1 GSE22155_02        TRUE 3.202117e-09     12287        479               353
## 2 GSE22155_02        TRUE 7.158934e-09     12680        479               360
## 3 GSE22155_02        TRUE 3.399091e-08     11738        479               338
## 4 GSE22155_02        TRUE 1.828431e-06      1512        479                73
## 5 GSE22155_02        TRUE 1.985401e-06      1988        479                88
## 6 GSE22155_02        TRUE 2.498789e-06       881        479                51
##   precision
## 1 0.7369520
## 2 0.7515658
## 3 0.7056367
## 4 0.1524008
## 5 0.1837161
## 6 0.1064718
```

```
p <- gostplot(multi_gostres, capped = FALSE, interactive = F)
p
```


GSE22155_02
GSE22155_47
GSE35640
GSE61992
GSE91061
−log10(p−adj)
10
5
0
GO:MF
GO:CC
GO:BP
REAC
MIRNA
HPA
CORUM
HP
WP

Estos resultados son los que utilizo para generar los archivos gem que archivo para usar más tarde en Cytoscape.

```
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

gem %>% group_by(query) %>%
  group_walk(~
    write.table(data.frame(.x[,c("GO.ID", "Description", "p.Val", "FDR", "Phenotype", "Genes")]),
                file = paste0("./Venn_diagrams_cell_type_treatments_vs_untreated/gProfiler_Tregcells_On
                sep = "\t", quote = F, row.names = F))
```

Proporciones de los pathways para cada estudio:

```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061
##         169         178         169         181         131
```

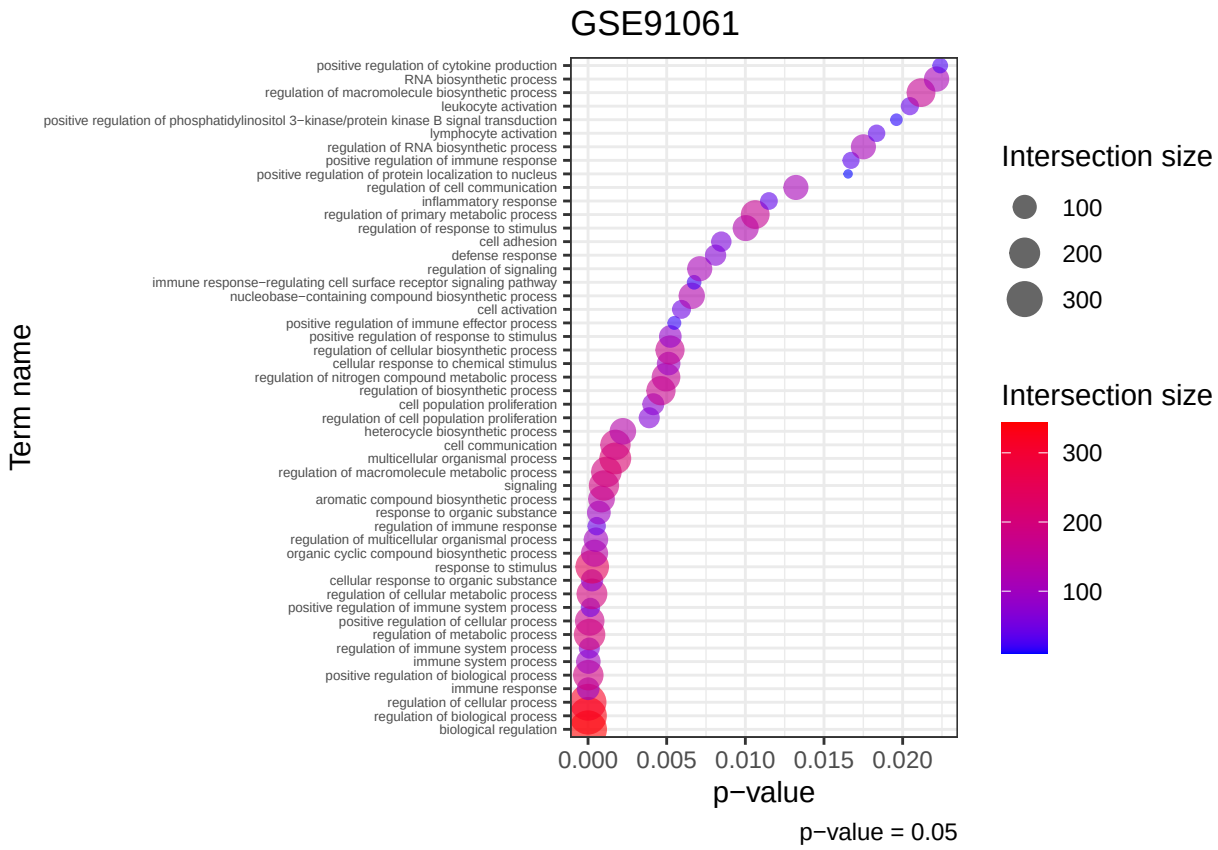
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061
##    20.41063    21.49758    20.41063    21.85990    15.82126
```

```
#table(df$term_name)
#prop.table(table(df$term_name))*100
df2 <- as.data.frame(table(df$term_name, df$query))
#length(unique(df2$Var1))
write.table(df, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/Tregcells_GOBPs.txt", sep = "
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/Tregcells_GOBPs_freq.txt", s
#rm(jVenn_Tregcells_vs_GSE22155_02, jVenn_Tregcells_vs_GSE22155_47, jVenn_Tregcells_vs_GSE35640, jVenn_
```

Barplot of the top GO-BPs:

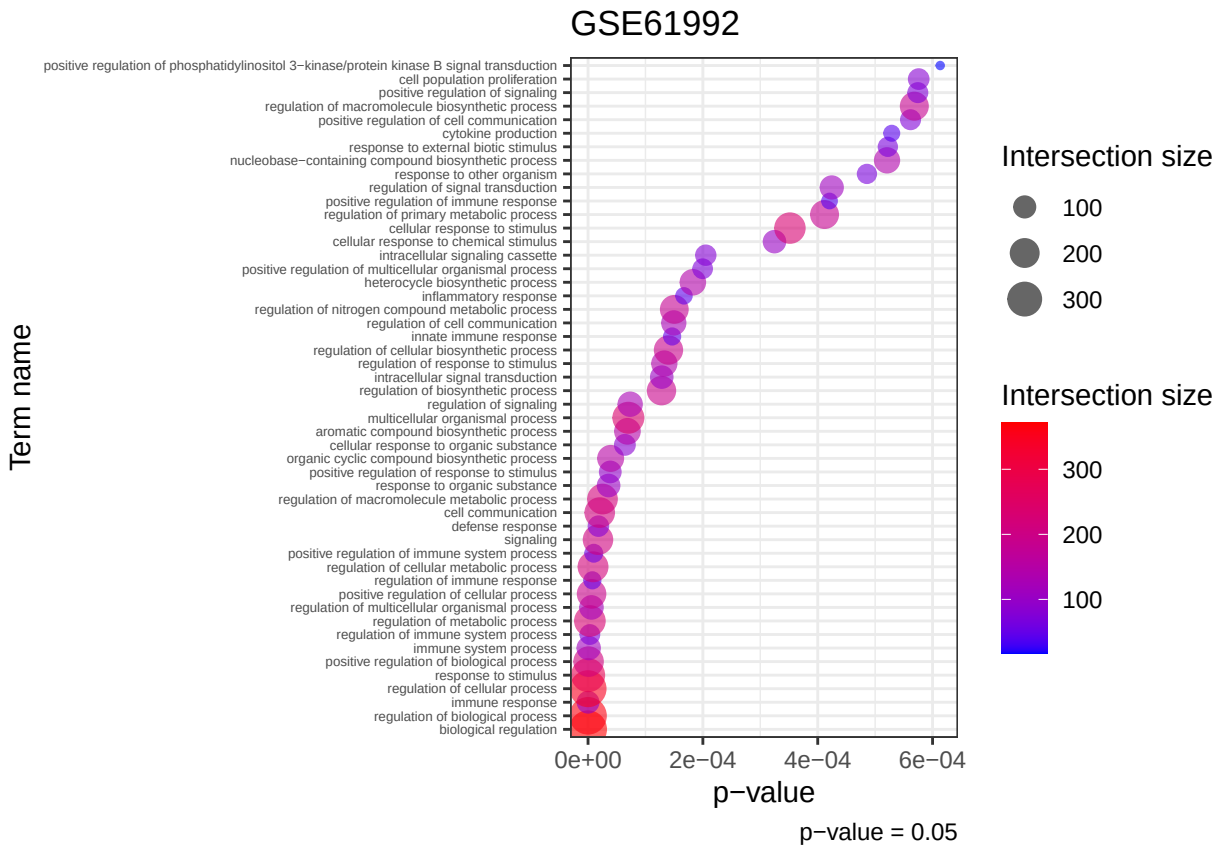
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size, size = intersection_s
  theme_bw() + theme(axis.text.y.left = element_text(size = 5)) +
  labs(x= "p-value", y="Term name", colour = "Intersection size", size = "Intersection size", caption =
  scale_color_gradient(low="blue", high="red")
}

plot_gobps(study = "GSE91061")
```




```
print("\n")
```

```
## [1] "\n"
```

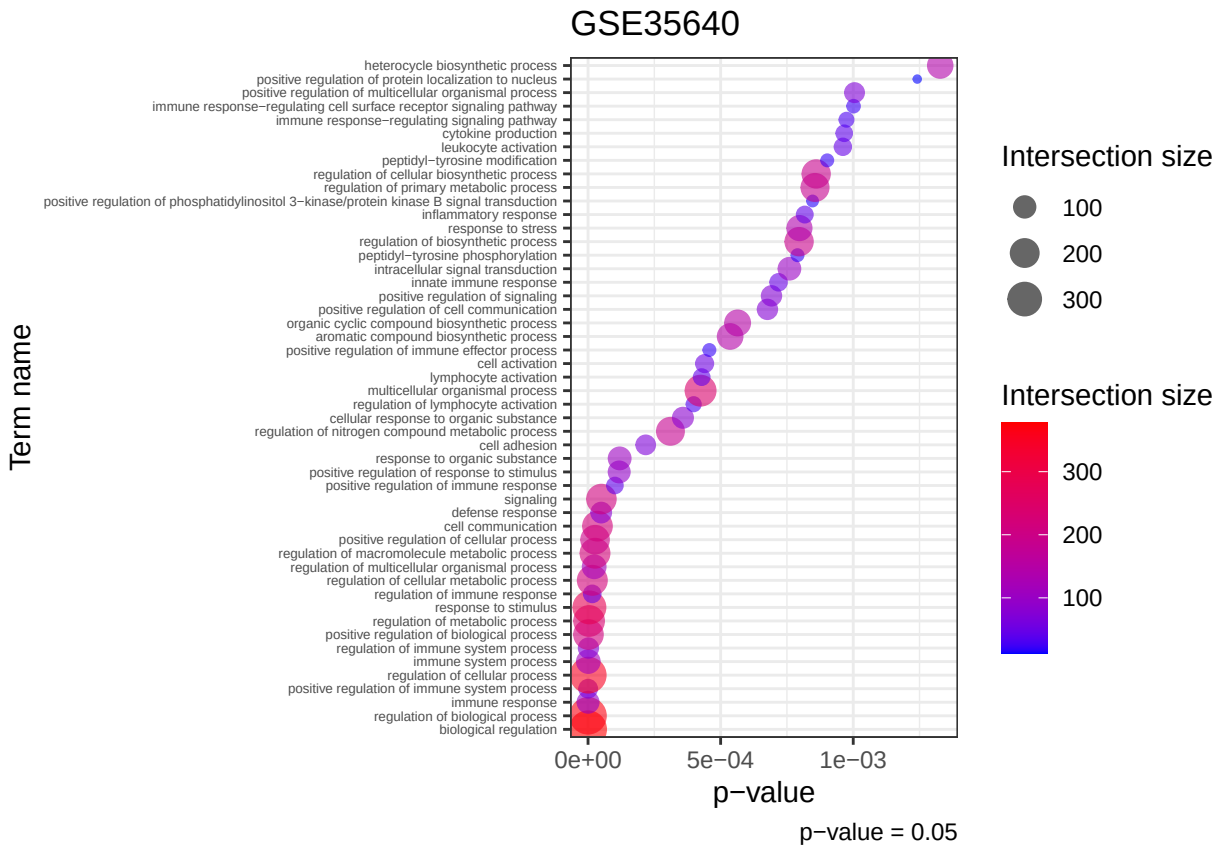
```
plot_gobps(study = "GSE61992")
```




```
print("\n")
```

```
## [1] "\n"
```

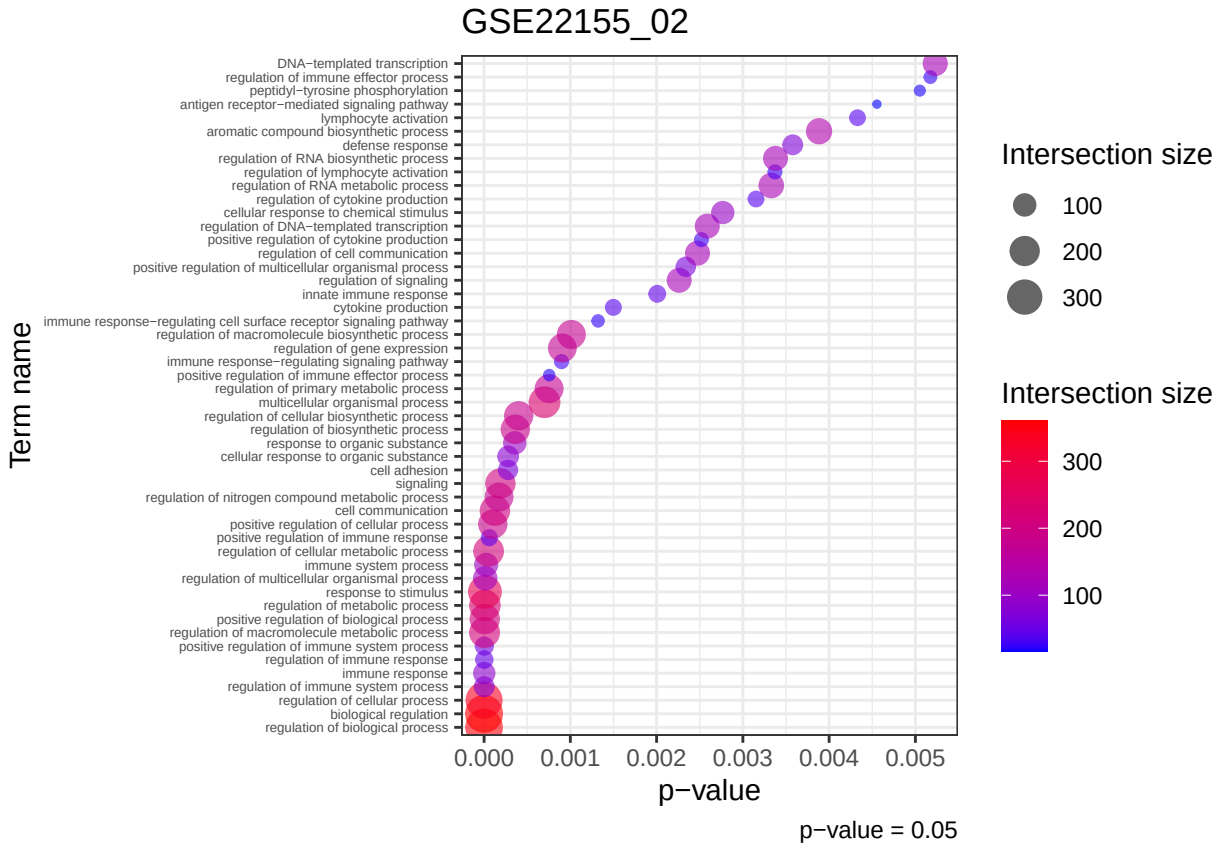
```
plot_gobps(study = "GSE35640")
```




```
print("\n")
```

```
## [1] "\n"
```

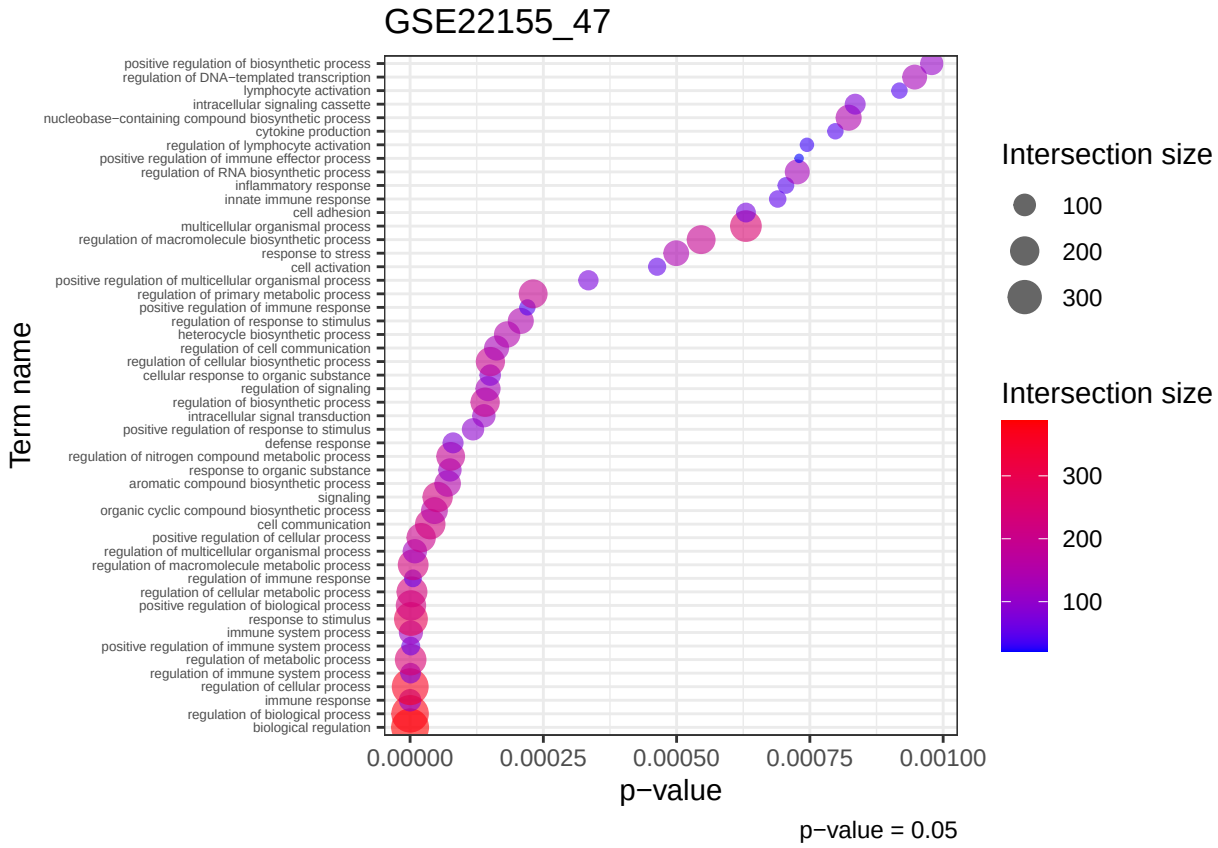
```
plot_gobps(study = "GSE22155_02")
```




```
print("\n")
```

```
## [1] "\n"
```

```
plot_gobps(study = "GSE22155_47")
```



5 Exclusive

En esta sección analizo los DEGs exclusivos para cada estudio, y realizo un multi-análisis para cada estudio. Los resultados los guardo como archivos gem específicos para usar en Cytoscape.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE61992" = jVenn_Tregcells$GSE61992[which(is.na(jVenn_Tregcells$GSE
                                   "GSE91061" = jVenn_Tregcells$GSE91061[which(is.na(jVenn_Tregcells$GSE91061
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

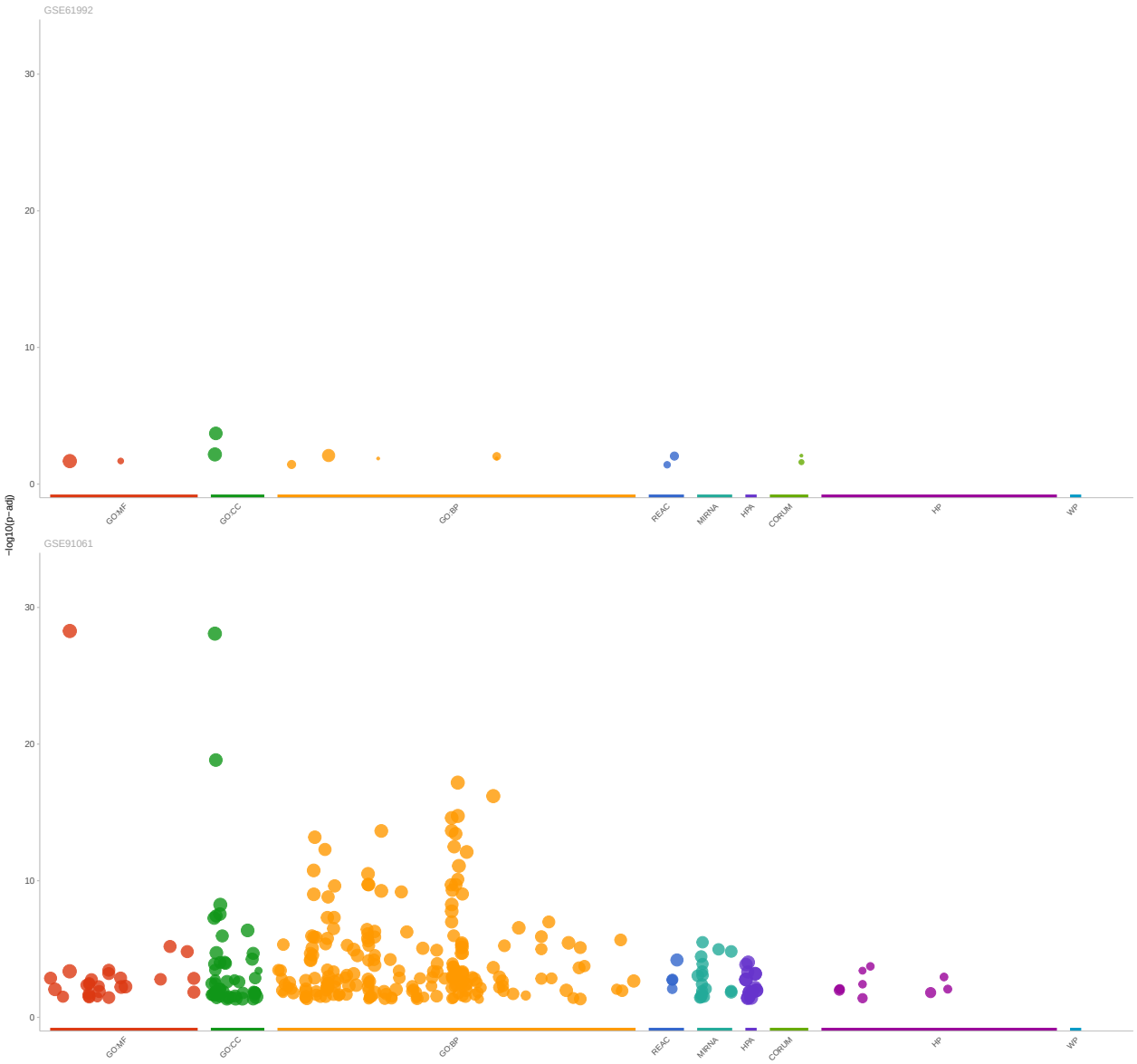
```
##      query significant     p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 0.008324058         2         16                2
## 2 GSE61992        TRUE 0.024903238         3         16                2
## 3 GSE61992        TRUE 0.008105184      2637         54               20
## 4 GSE61992        TRUE 0.009362157        13         54                3
## 5 GSE61992        TRUE 0.013481220         2         54                2
## 6 GSE61992        TRUE 0.013481220         2         54                2
##    precision
## 1 0.12500000
## 2 0.12500000
## 3 0.37037037
## 4 0.05555556
## 5 0.03703704
```

```
## 6 0.03703704
```

Estos resultados son los que utilizo para generar los archivos gem que archivo para usar más tarde en Cytoscape.

```
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

gem %>% group_by(query) %>%
  group_walk(~
    write.table(data.frame(.x[,c("GO.ID", "Description", "p.Val", "FDR", "Phenotype", "Genes")]),
                file = paste0("./Treatment_comparisons/gProfiler_Tregcells_Only_geneset_", unique(.y$qu
                sep = "\t", quote = F, row.names = F))
```

```
p <- gostplot(multi_gostres, capped = FALSE, interactive = F)
p
```



Proporciones de los pathways para cada estudio:

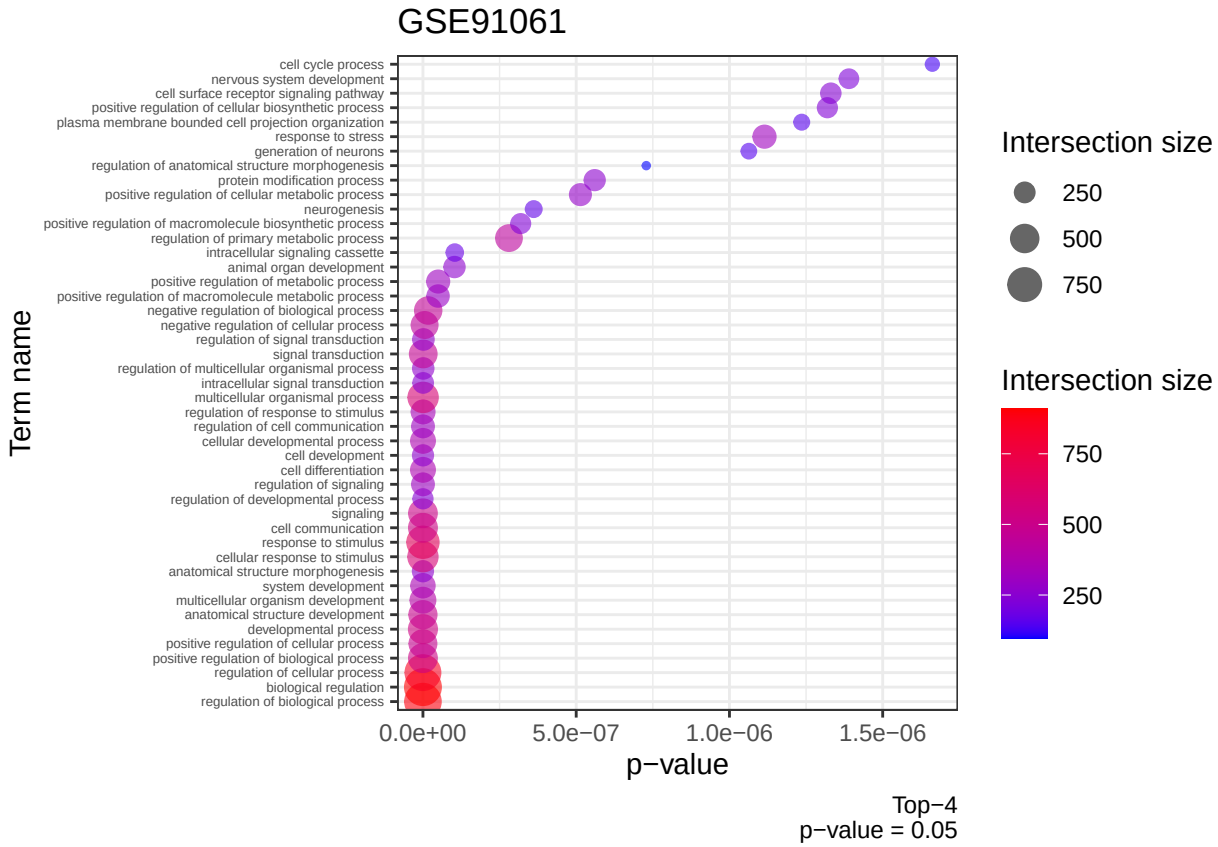
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061
##       13      325
```

```
write.table(gem, file = "./Treatment_comparisons/gProfiler_Tregcells_Only_genesets.txt", sep = "\t",quo
```

```
plot_gobps <- function(study){
df2_GSE91061 <- df2[df2$query == study,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size, size = intersection_si
  theme_bw() + theme(axis.text.y.left = element_text(size = 5)) +
  labs(x= "p-value", y="Term name", colour = "Intersection size", size = "Intersection size", caption =
  scale_color_gradient(low="blue", high="red")
```

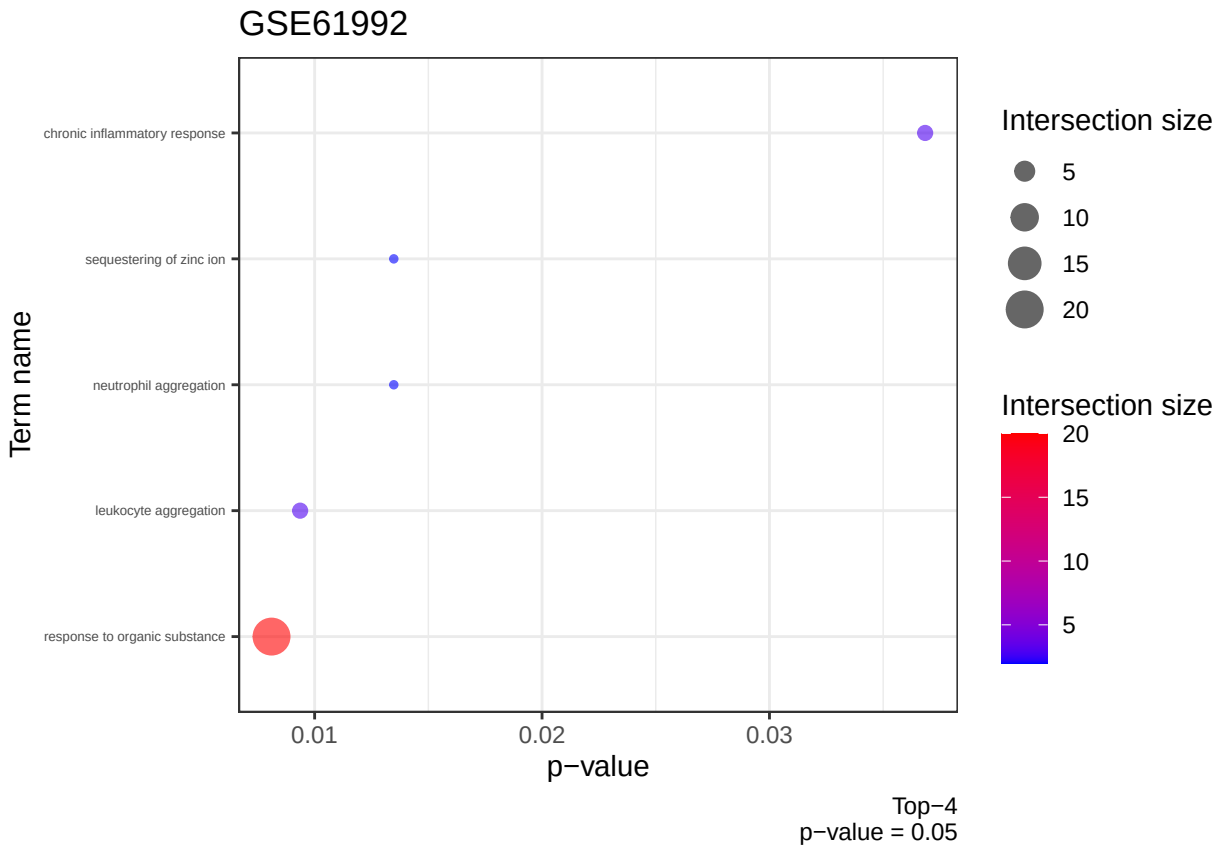
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




```
print("\n")
```

```
## [1] "\n"
```

```
df2 <- df[df$source == "GO:BP",]
plot_gobps(study = "GSE61992")
```



6 Máximo solapamiento

En esta sección analizo los DEGs que más se comparten entre los estudios para observar los mecanismos que pudieran haber en común entre los tratamientos.

Primero realizo el pathway multianálisis, pero no se obtiene resultado: "Empty query. Please double check your input." Sólo hay 5 genes, lo que debe impedir obtener un resultado.

```
#multi_gostres <- gost(query = list("TabrafenibTrametinib_antiCTLA4PD1" = jVenn_Tregcells$`GSE91061|GSE
#),
#                       evcodes = TRUE, multi_query = FALSE,
#                       sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
#head(multi_gostres$result[,1:7])
```