Pathway Analysis - T CD4 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 T CD4 obtenidas en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_TCD4cells_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
head(jVenn_TCD4cells_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509  GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>     <chr>       <chr>
## 1 TNFRSF13B VAC14       FCMR
## 2 CSTF3     KTI12       CD79A
## 3 RPS17     DSTYK       CD19
## 4 SRSF10    PARS2       TMEM14C
## 5 POU2AF1   ADIPOR1     DKC1
## 6 JSRP1     MS4A1       RPS16
```

```
jVenn_TCD4cells_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
jVenn_TCD4cells_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD
jVenn_TCD4cells_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD
jVenn_TCD4cells_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_TCD4cells <- 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:

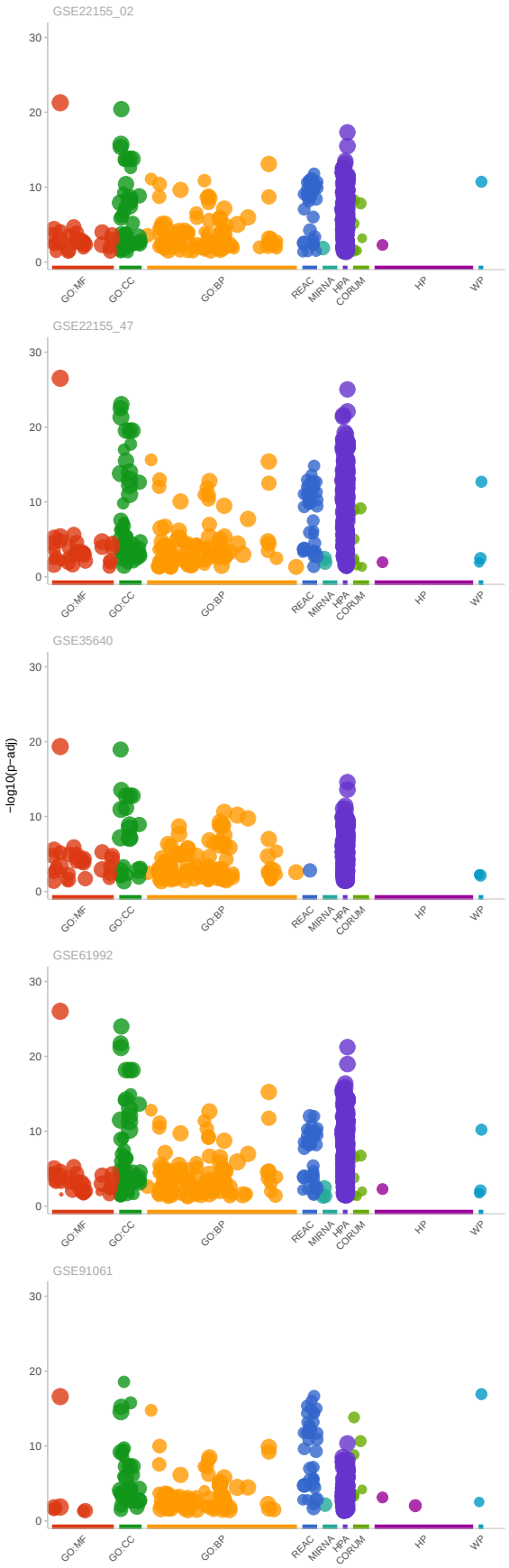
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_TCD4cells_vs_GSE22155_02$GSE50509[which
                                   "GSE22155_47" = jVenn_TCD4cells_vs_GSE22155_47$GSE50509[which(is.na(jVenn_
                                   "GSE35640" = jVenn_TCD4cells_vs_GSE35640$GSE50509[which(is.na(jVenn_TCD4ce
                                   "GSE91061" = jVenn_TCD4cells_vs_GSE91061$GSE50509[which(is.na(jVenn_TCD4ce
                                   "GSE61992" = jVenn_TCD4cells_vs_GSE61992$GSE50509[which(is.na(jVenn_TCD4ce
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##         query significant      p_value term_size query_size intersection_size
## 1 GSE22155_02        TRUE 5.334126e-09        74        299               27
## 2 GSE22155_02        TRUE 1.459715e-08        93        299               30
## 3 GSE22155_02        TRUE 7.105894e-06        43        299               17
## 4 GSE22155_02        TRUE 6.572587e-04        21        299               10
## 5 GSE22155_02        TRUE 1.960581e-02        29        299               10
## 6 GSE22155_02        TRUE 2.975596e-02        20        299                8
##    precision
## 1 0.09030100
## 2 0.10033445
## 3 0.05685619
## 4 0.03344482
## 5 0.03344482
## 6 0.02675585
```

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


GSE22155_02
GSE22155_47
GSE35640
GSE61992
GSE91061
−log10(p−adj)
30
20
10
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_TCD4cells_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
##         429         489         371         489         326
```

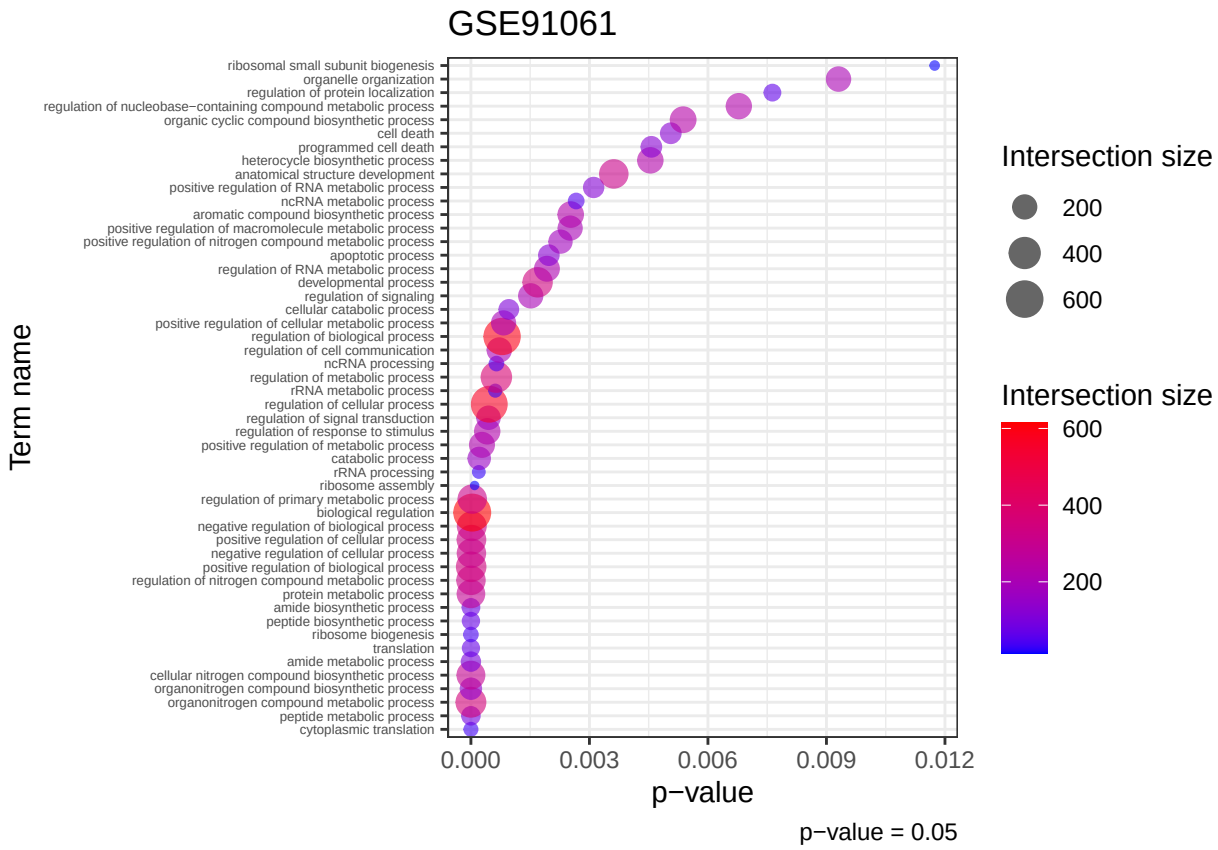
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061
##    20.38973    23.24144    17.63308    23.24144    15.49430
```

```
#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/TCD4cells_GOBPs.txt", sep = "
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/TCD4cells_GOBPs_freq.txt", s
#rm(jVenn_TCD4cells_vs_GSE22155_02, jVenn_TCD4cells_vs_GSE22155_47, jVenn_TCD4cells_vs_GSE35640, jVenn_
```

Barplot of the top GO-BPs:

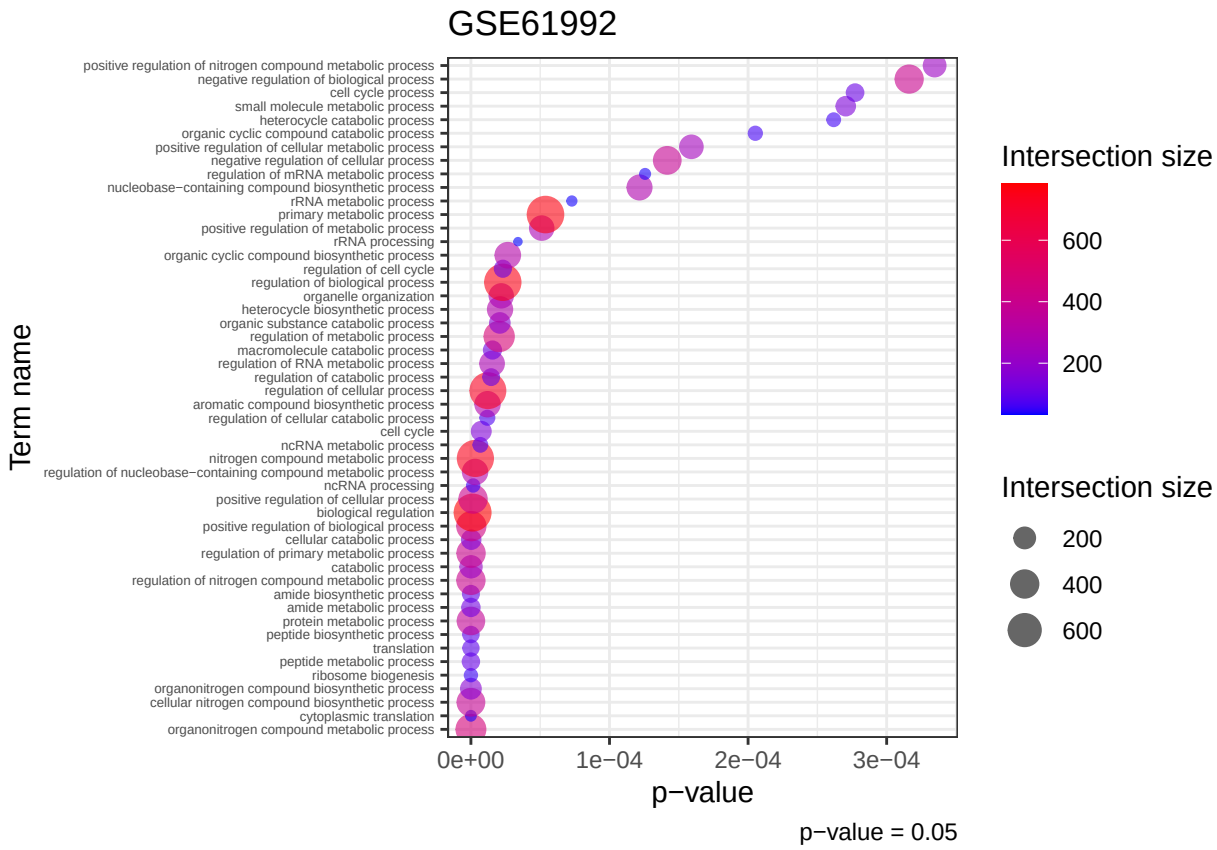
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size, 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"
```

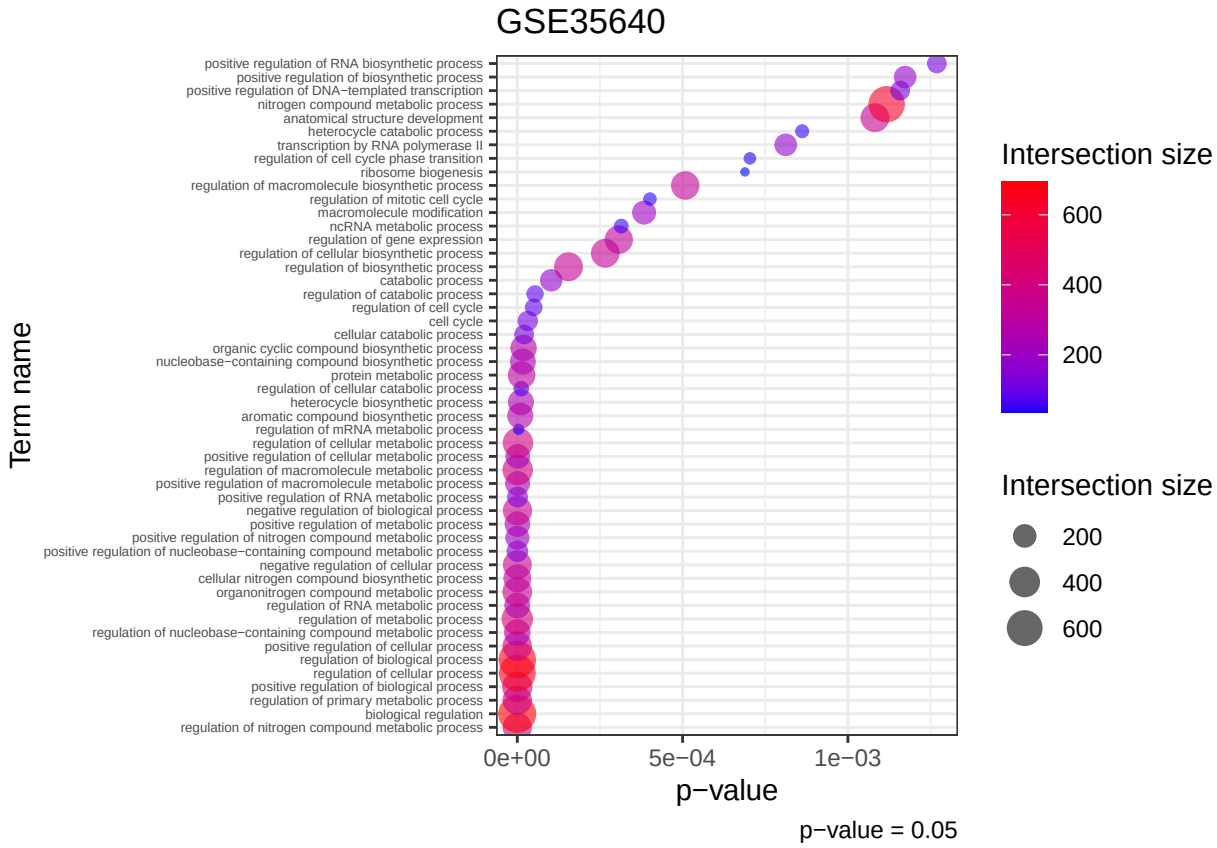
```
plot_gobps(study = "GSE61992")
```




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

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

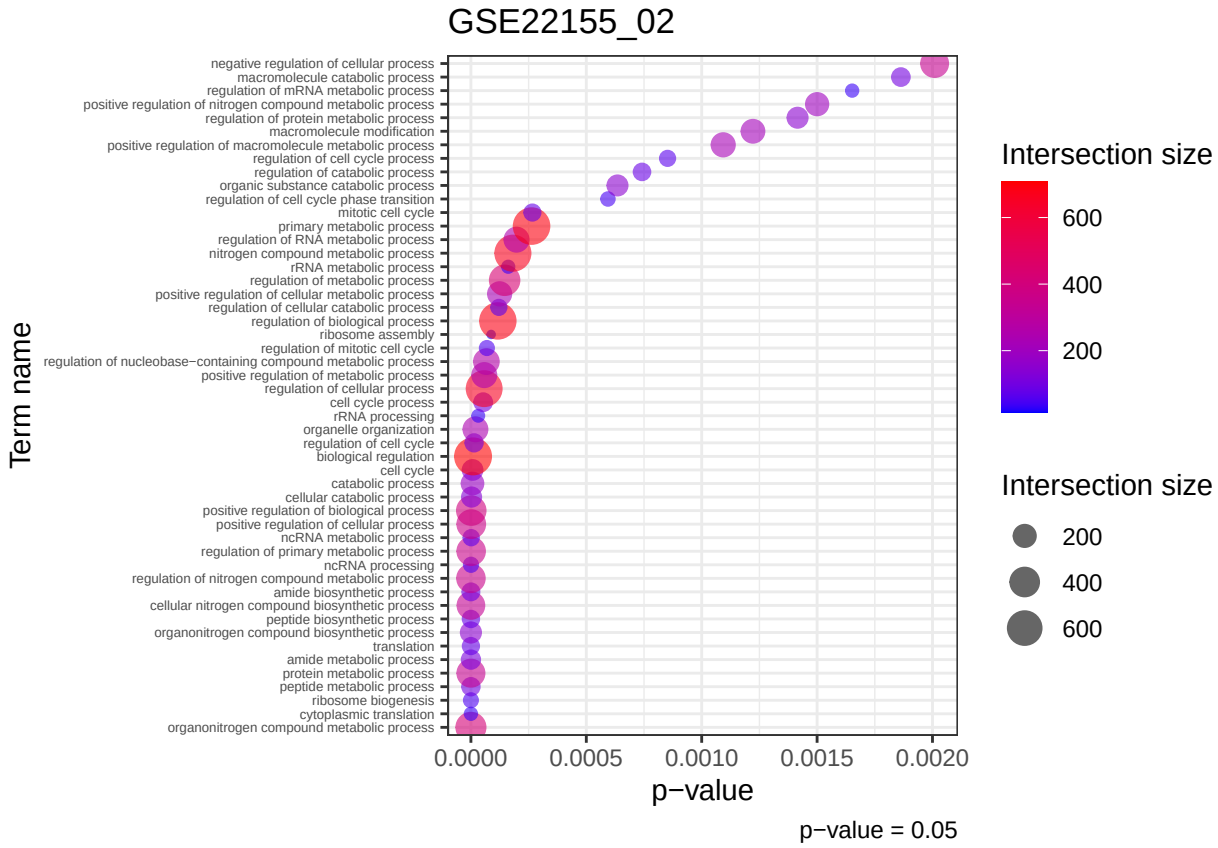
```
plot_gobps(study = "GSE35640")
```




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

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

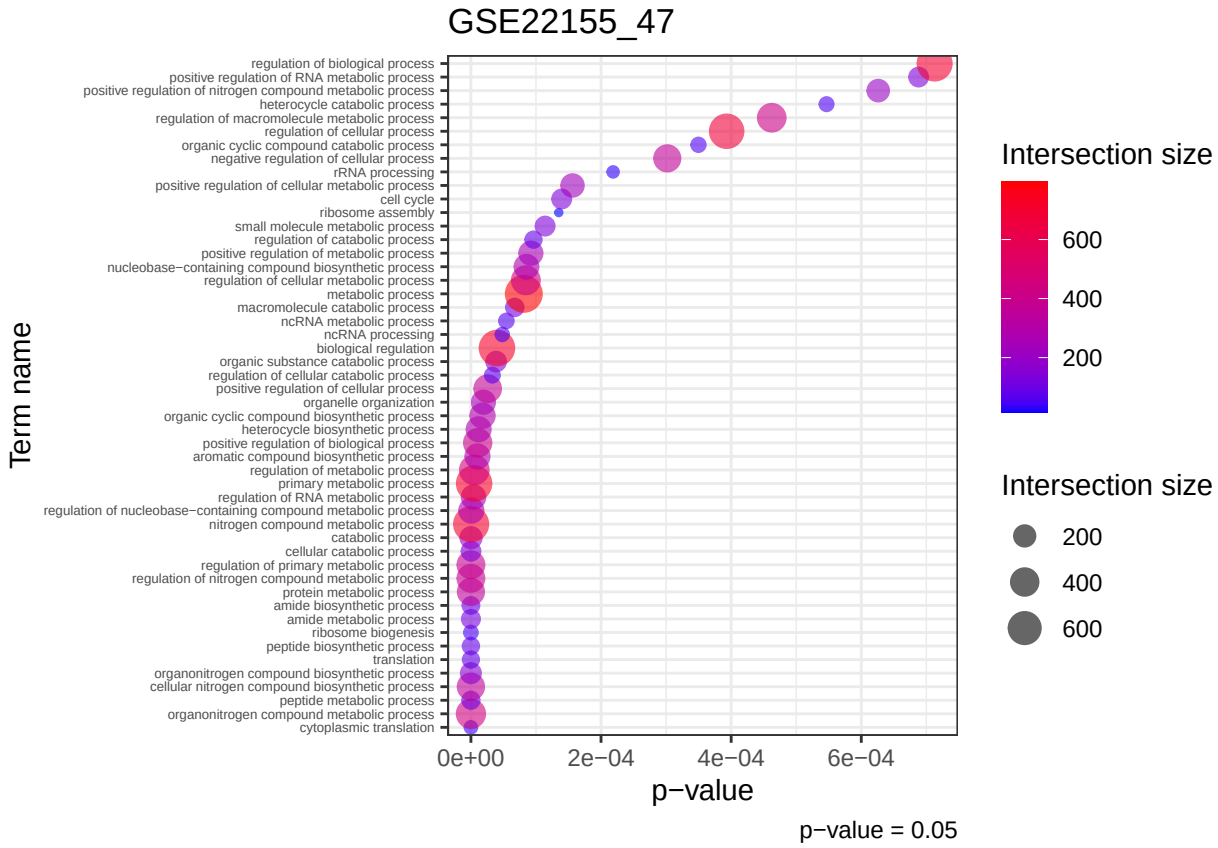
```
plot_gobps(study = "GSE22155_02")
```




```
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_TCD4cells$GSE61992[which(is.na(jVenn_TCD4cells$GSE
                                   "GSE91061" = jVenn_TCD4cells$GSE91061[which(is.na(jVenn_TCD4cells$GSE91061
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

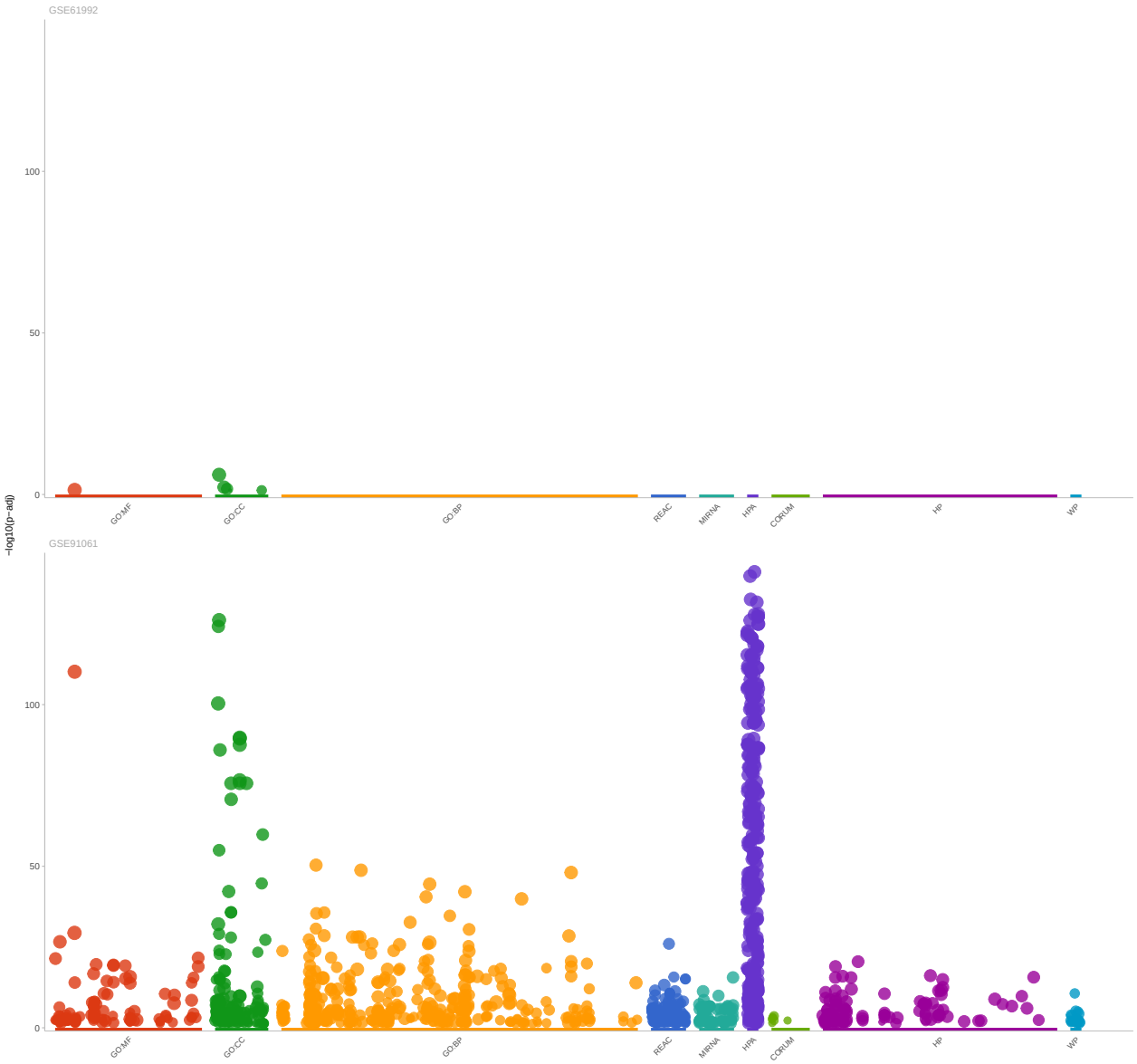
```
##      query significant      p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 7.639095e-07     12345         89              76
## 2 GSE61992        TRUE 5.445643e-03      4777         89              37
## 3 GSE61992        TRUE 1.861715e-02       896         89              13
## 4 GSE61992        TRUE 4.923287e-02       126         89               5
## 5 GSE61992        TRUE 4.973386e-02         5         89               2
## 6 GSE61992        TRUE 3.909250e-02     14838         91              82
##    precision
## 1 0.85393258
## 2 0.41573034
## 3 0.14606742
## 4 0.05617978
## 5 0.02247191
```

```
## 6 0.90109890
```

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_TCD4cells_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:

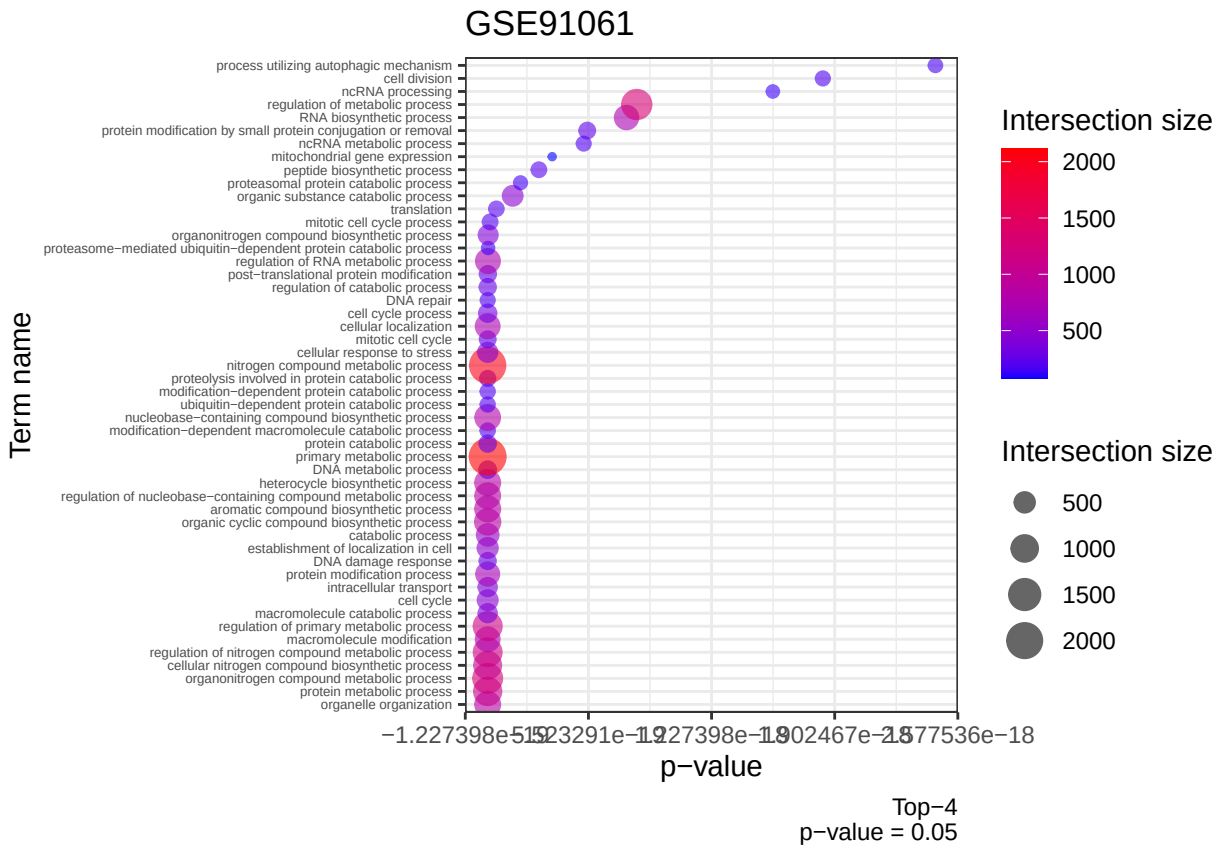
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061
##        6     1345
```

```
write.table(gem, file = "./Treatment_comparisons/gProfiler_TCD4cells_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")
```

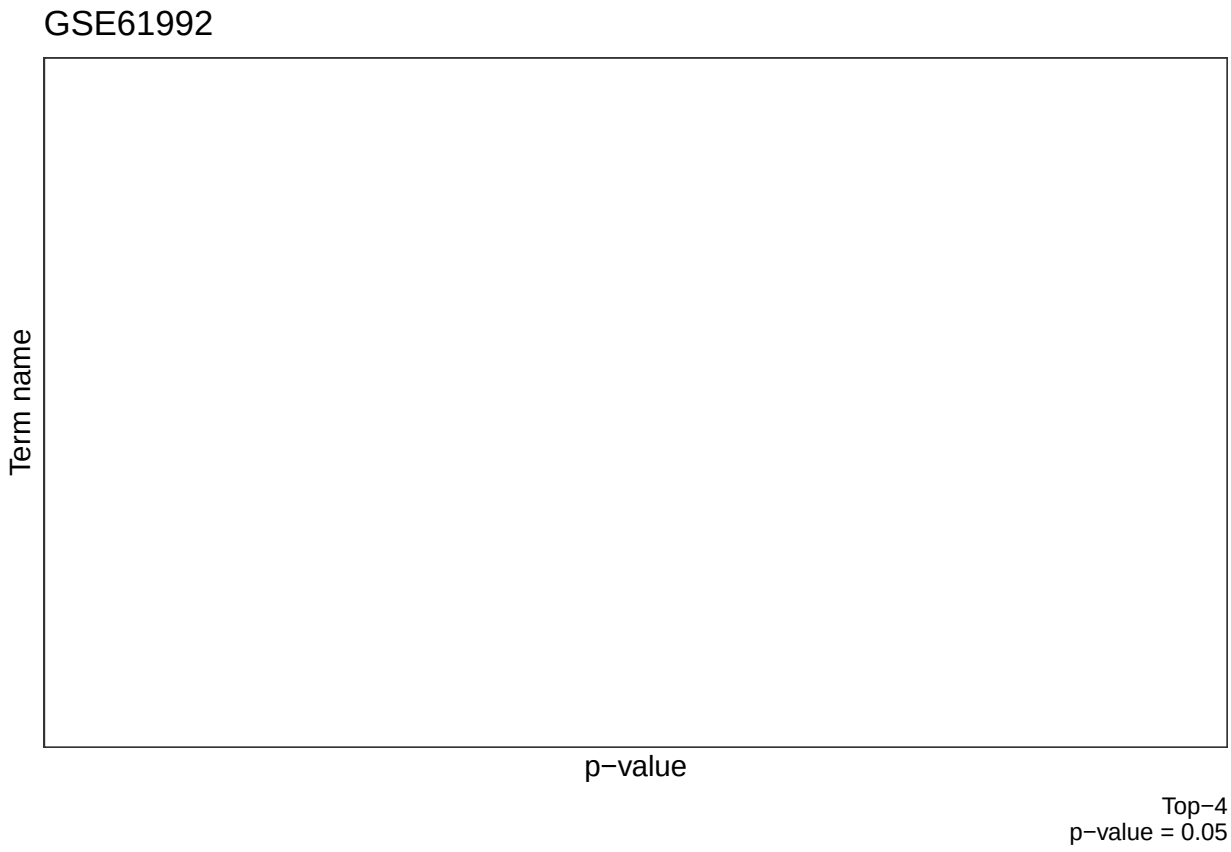
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




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

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

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



6 Máximo solapamiento

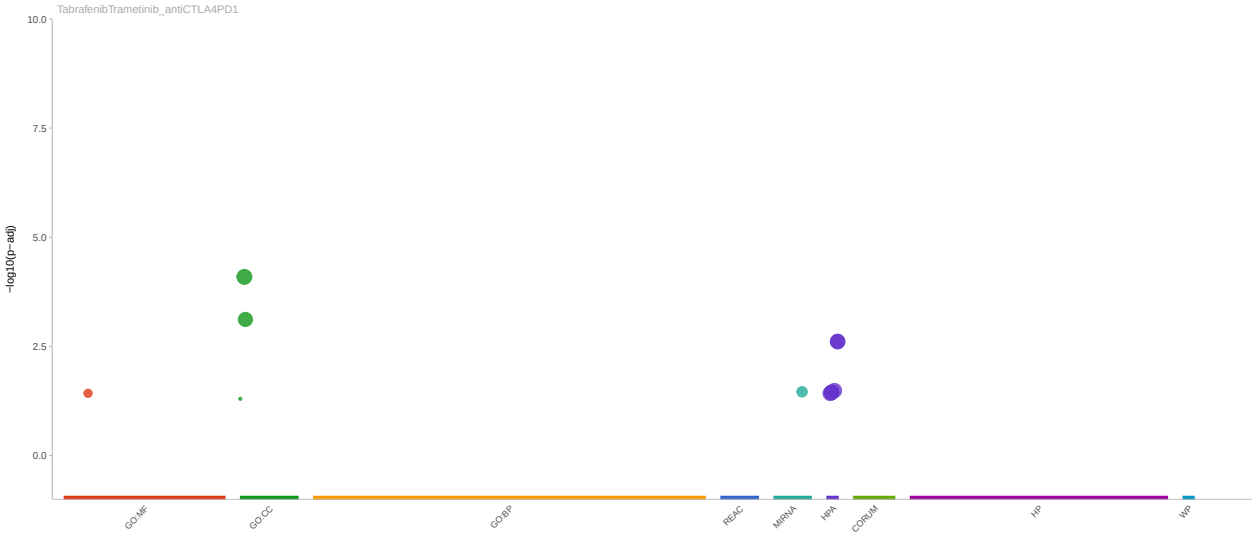
En esta sección analizo los DEGs que más se comparten entre los estudios para observar los mecanismos que pudieran haber en común entre los tratamientos.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("TabrafenibTrametinib_antiCTLA4PD1" = jVenn_TCD4cells$`GSE91061|GSE6
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##                              query significant      p_value term_size
## 1 TabrafenibTrametinib_antiCTLA4PD1        TRUE 8.070222e-05     12345
## 2 TabrafenibTrametinib_antiCTLA4PD1        TRUE 7.613922e-04      5487
## 3 TabrafenibTrametinib_antiCTLA4PD1        TRUE 4.995269e-02        16
## 4 TabrafenibTrametinib_antiCTLA4PD1        TRUE 3.741367e-02        57
## 5 TabrafenibTrametinib_antiCTLA4PD1        TRUE 2.441565e-03      7664
## 6 TabrafenibTrametinib_antiCTLA4PD1        TRUE 2.441565e-03      7664
##   query_size intersection_size  precision
## 1         31                30 0.96774194
## 2         31                20 0.64516129
## 3         31                 2 0.06451613
## 4         32                 3 0.09375000
## 5         31                31 1.00000000
## 6         31                31 1.00000000
```

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

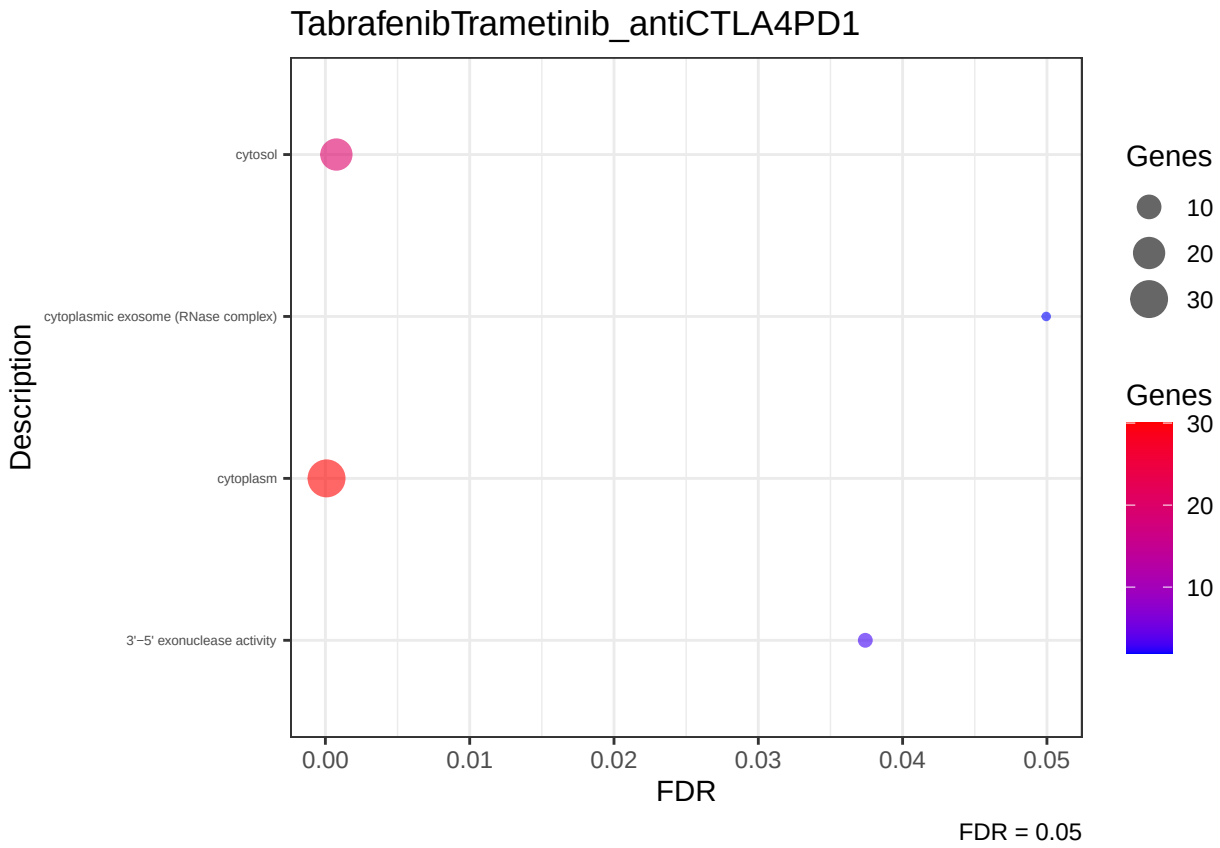


Estos resultados son los que utilizo para generar los archivos gem que archivo para usar más tarde en Cytoscape.

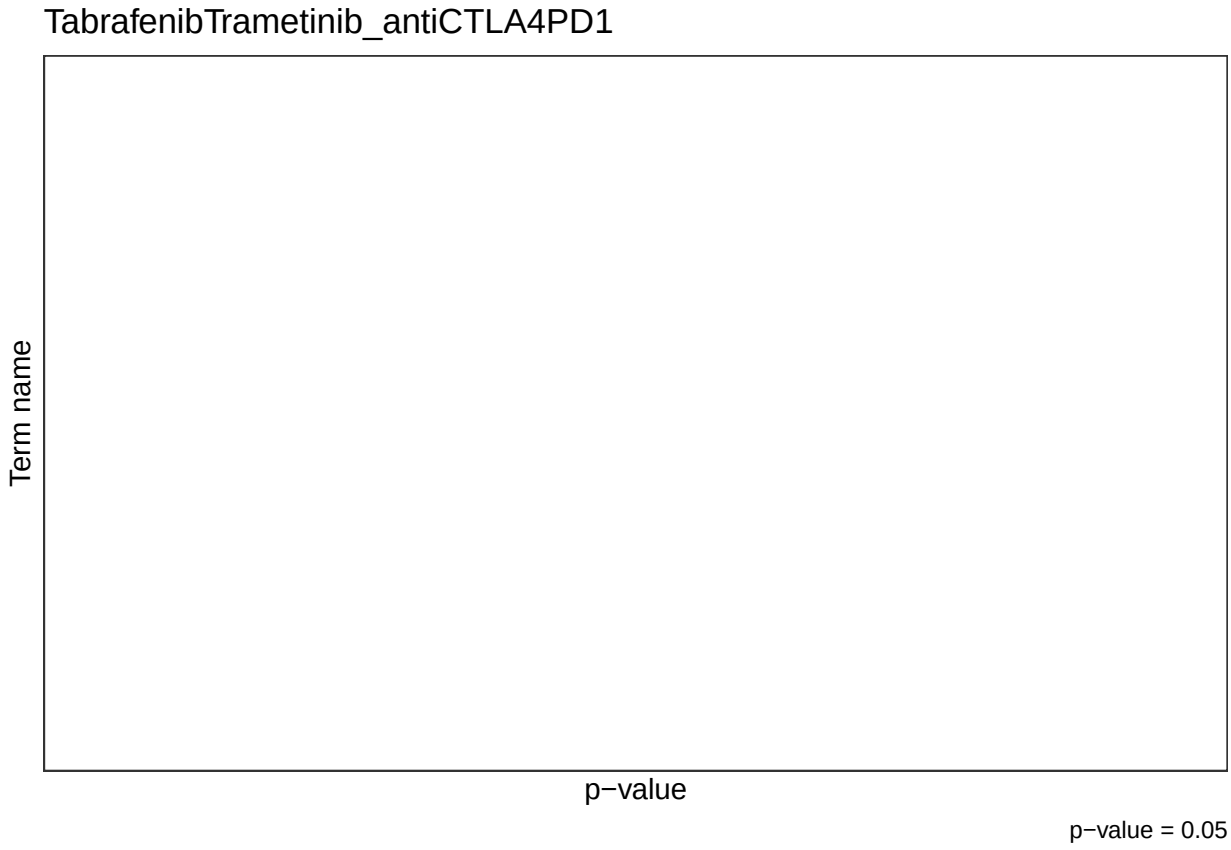
```
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

gem %>% group_by(query) %>%
  group_walk(~
                   write.table(data.frame(.x[,c("GO.ID", "Description", "p.Val", "FDR", "Phenotype", "G
                file = paste0("./Treatment_comparisons/gProfiler_TCD4cells_maxOverlap_geneset_", unique
                sep = "\t", quote = F, row.names = F))
```

Al mirar los procesos por cada comparación salieron demasiados para poder visualizarlos bien, por lo que decidí restringir a los procesos GO-BPs únicamente y guardarlos en pdf.

```
#unique(gem$query)

gem2 <- gem[grep("GO",gem$GO.ID),]
plot_gobps <- function(query, cutoff = 0.05) {
gem2 <- gem2[gem2$FDR <= cutoff,]
gem2_1 <- gem2[gem2$query == query, ]
ggplot(gem2_1, aes(FDR, Description)) + ggtitle(query) +
  geom_point(aes(FDR, Description, colour = Genes, size = Genes), alpha = 0.6)+
  theme_bw() + theme(axis.text.y.left = element_text(size = 5)) +
  labs(caption = paste0("FDR = ",cutoff))+
  scale_color_gradient(low="blue", high="red")
}

plot_gobps("TabrafenibTrametinib_antiCTLA4PD1")
```




```
df <- as.data.frame(multi_gostres$result[,1:13])
df2 <- df[df$source == "GO:BP",]
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")
}
plot_gobps(study = "TabrafenibTrametinib_antiCTLA4PD1")
```


TabrafenibTrametinib_antiCTLA4PD1
Term name
p−value
p−value = 0.05