Pathway Analysis - T CD8 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 CD8 obtenidas en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_TCD8cells_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
head(jVenn_TCD8cells_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509 GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>    <chr>       <chr>
## 1 TNFSF13B TARP        CD8A
## 2 AIF1     CXCR3       CD2
## 3 XIRP1    PARP15      SIRPG
## 4 CD3D     IDO1        CD3E
## 5 HLA-DMB  SAMD3       IL10RA
## 6 NCF1C    HSH2D       CD27
```

```
jVenn_TCD8cells_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_T
jVenn_TCD8cells_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD8
jVenn_TCD8cells_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD8
jVenn_TCD8cells_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_TCD8
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_TCD8cells <- 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:

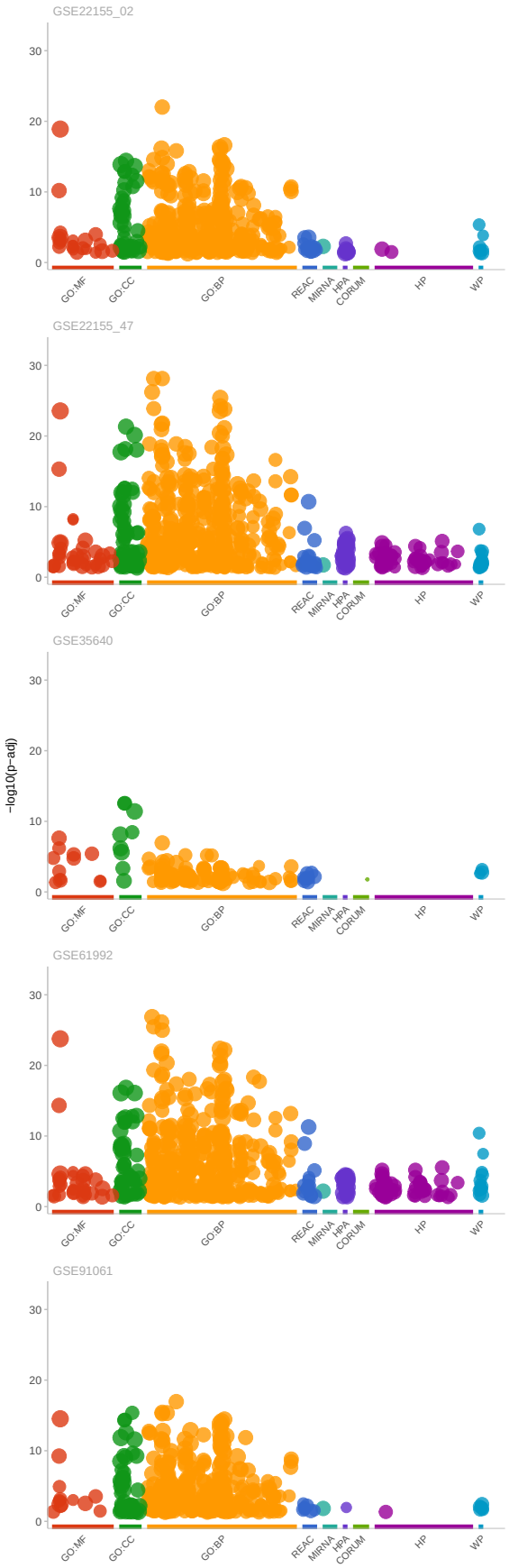
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_TCD8cells_vs_GSE22155_02$GSE50509[which
                                   "GSE22155_47" = jVenn_TCD8cells_vs_GSE22155_47$GSE50509[which(is.na(jVenn_T
                                   "GSE35640" = jVenn_TCD8cells_vs_GSE35640$GSE50509[which(is.na(jVenn_TCD8cel
                                   "GSE91061" = jVenn_TCD8cells_vs_GSE91061$GSE50509[which(is.na(jVenn_TCD8cel
                                   "GSE61992" = jVenn_TCD8cells_vs_GSE61992$GSE50509[which(is.na(jVenn_TCD8cel
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##         query significant      p_value term_size query_size intersection_size
## 1 GSE22155_02        TRUE 9.668334e-23      1512        907               158
## 2 GSE22155_02        TRUE 2.338781e-17      2955        907               231
## 3 GSE22155_02        TRUE 4.810050e-17      6250        907               398
## 4 GSE22155_02        TRUE 7.122774e-17      3855        907               278
## 5 GSE22155_02        TRUE 9.476473e-17      3939        907               282
## 6 GSE22155_02        TRUE 1.512516e-16      1496        907               143
##   precision
## 1 0.1742007
## 2 0.2546858
## 3 0.4388093
## 4 0.3065050
## 5 0.3109151
## 6 0.1576626
```

```
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_TCD8cells_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
##         463         683         122         649         381
```

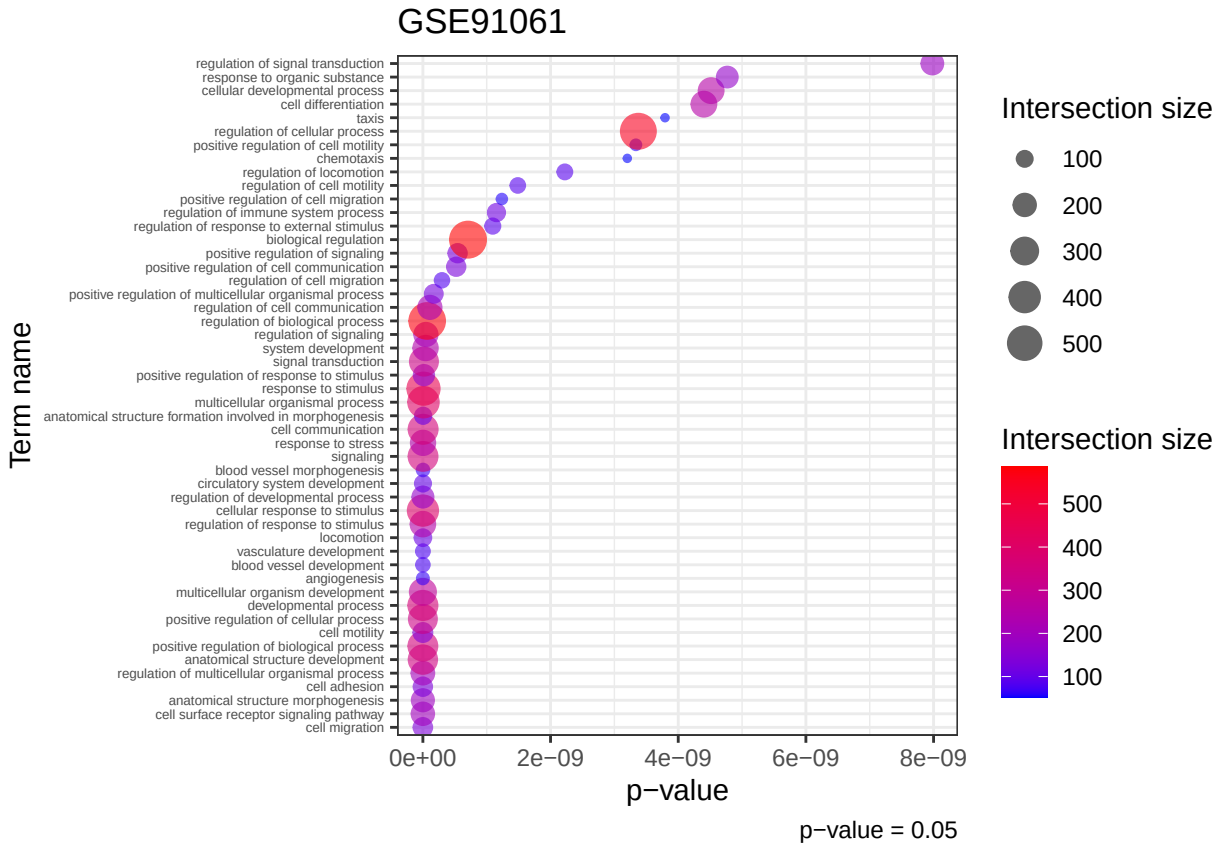
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061
##   20.147955   29.721497    5.308964   28.241950   16.579634
```

```
#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/TCD8cells_GOBPs.txt", sep = "
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/TCD8cells_GOBPs_freq.txt", s
#rm(jVenn_TCD8cells_vs_GSE22155_02, jVenn_TCD8cells_vs_GSE22155_47, jVenn_TCD8cells_vs_GSE35640, jVenn_
```

Barplot of the top GO-BPs:

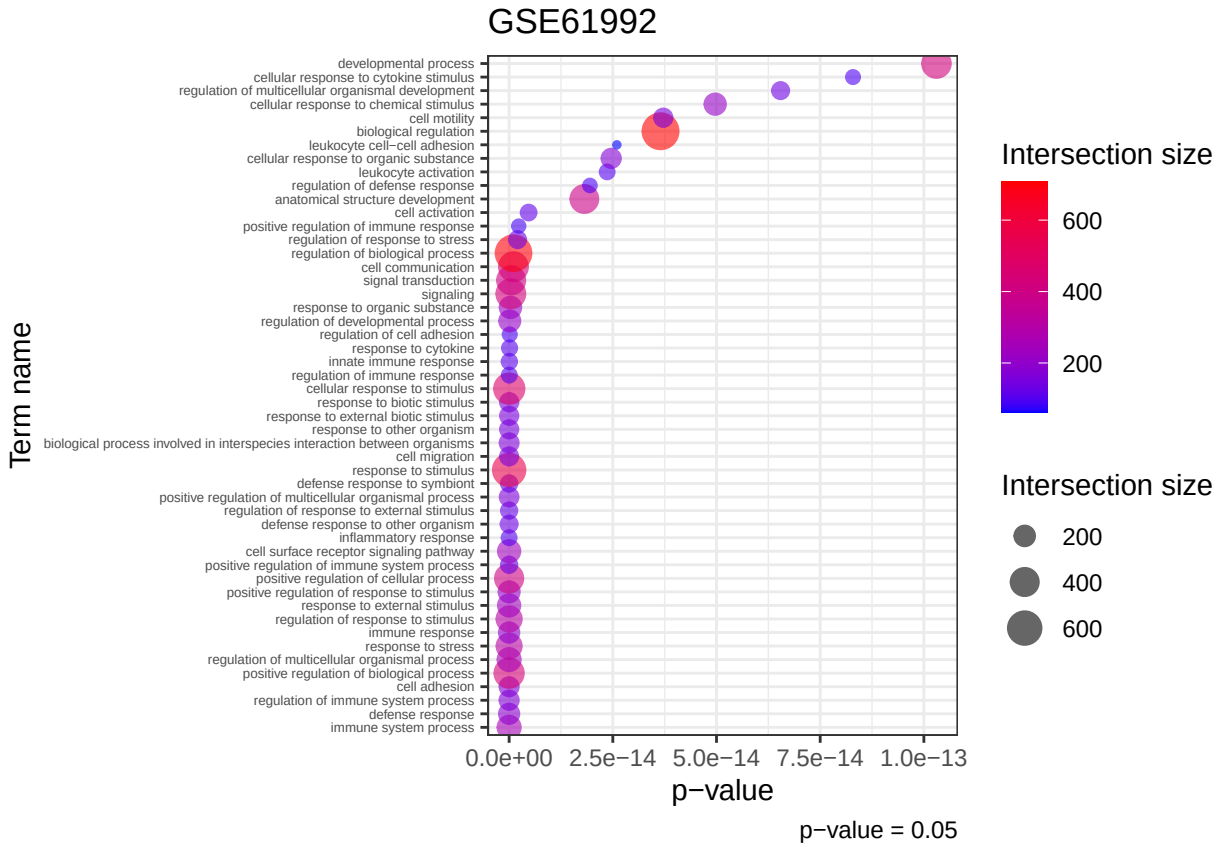
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size,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

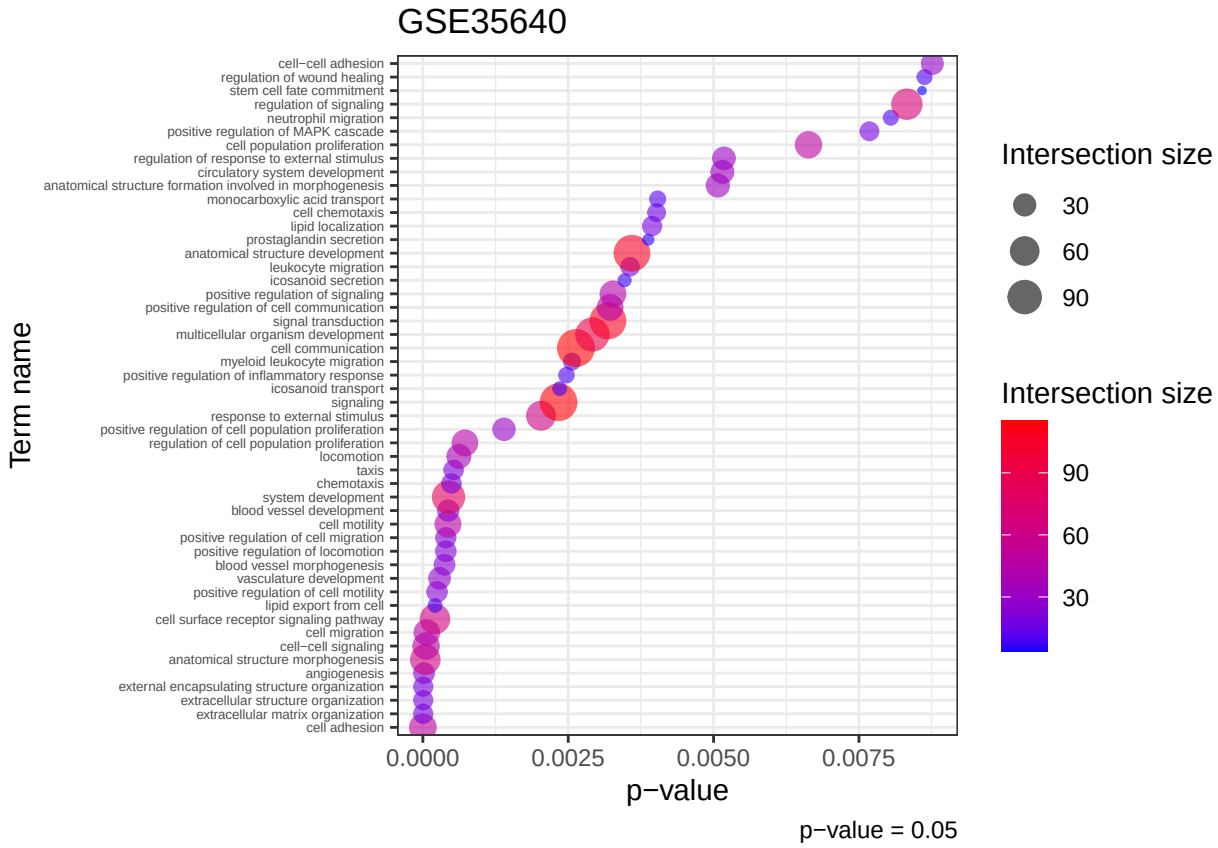
```
plot_gobps(study = "GSE61992")
```




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

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

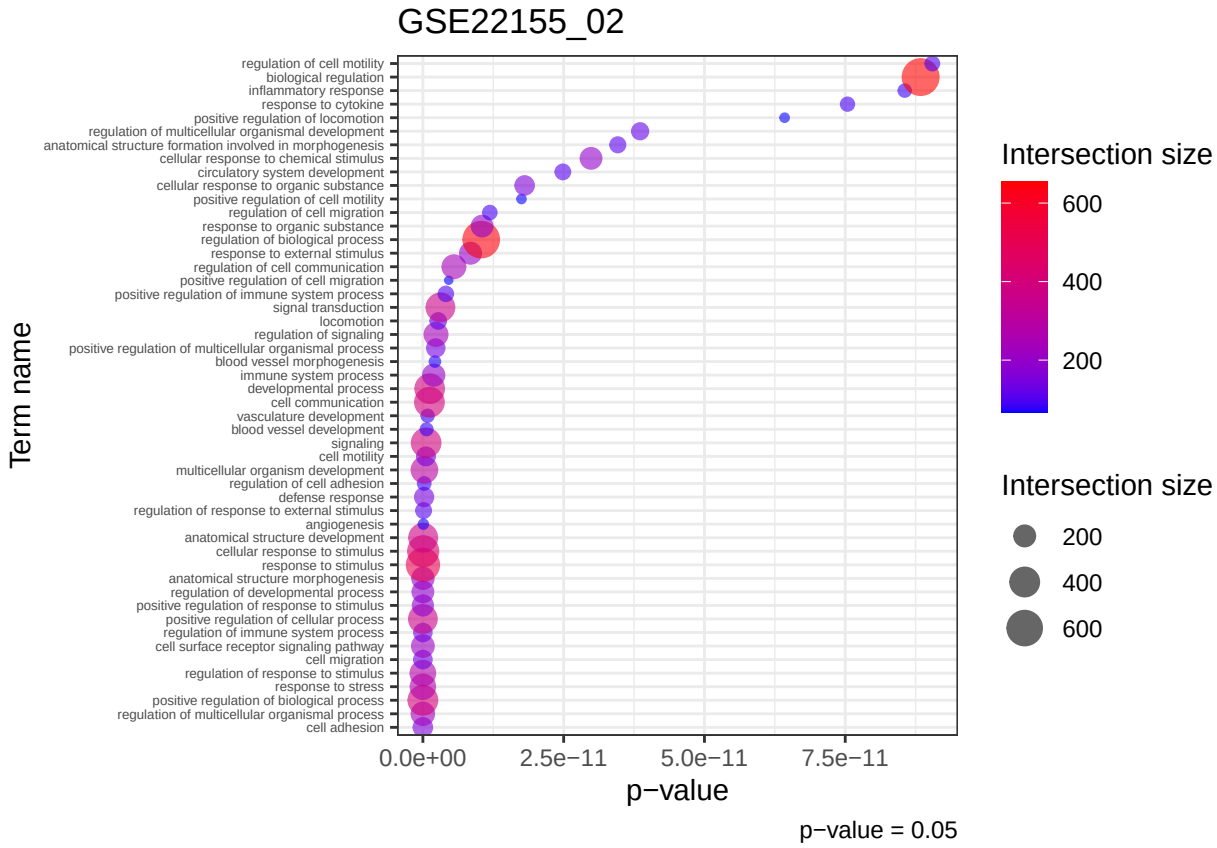
```
plot_gobps(study = "GSE35640")
```




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

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

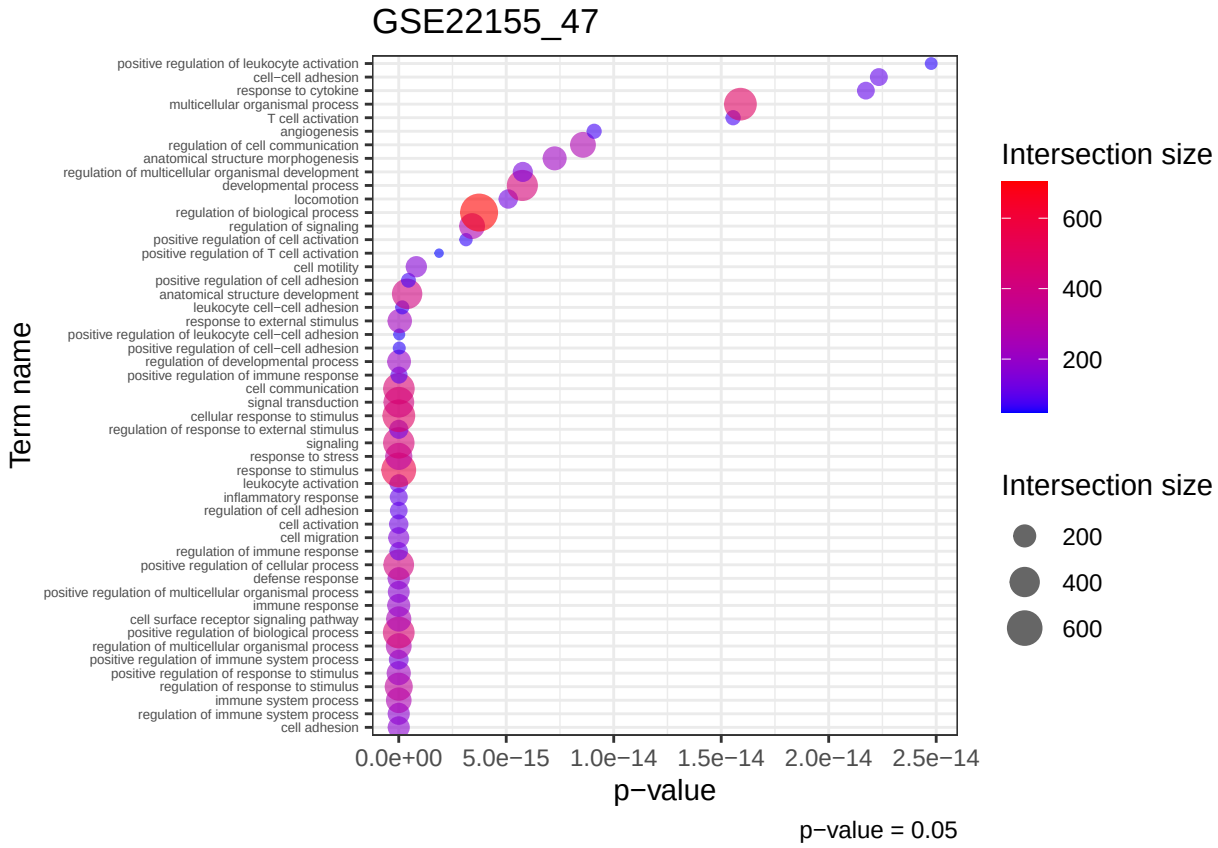
```
plot_gobps(study = "GSE22155_02")
```




```
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_TCD8cells$GSE61992[which(is.na(jVenn_TCD8cells$GSE
                                   "GSE91061" = jVenn_TCD8cells$GSE91061[which(is.na(jVenn_TCD8cells$GSE91061
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

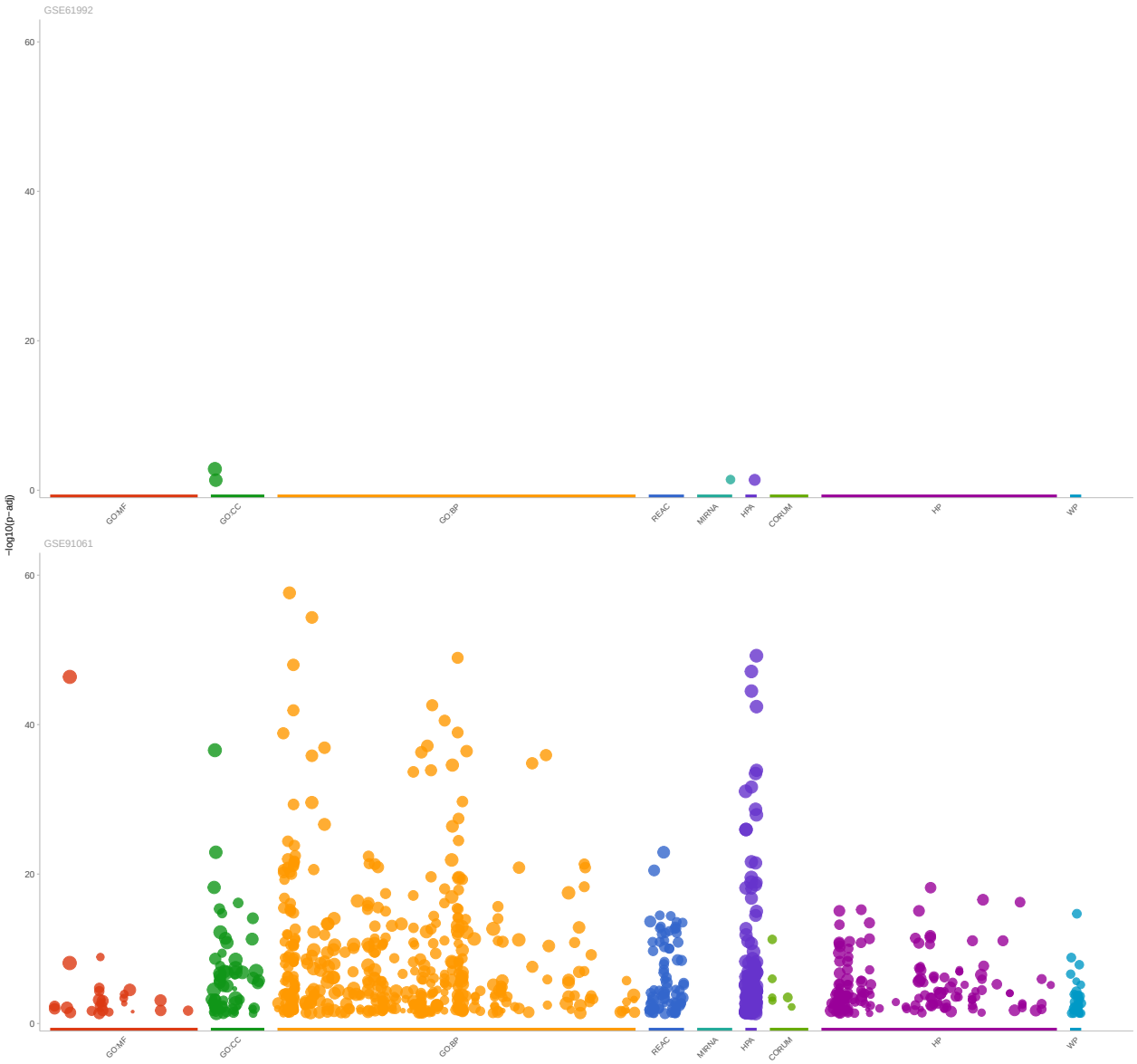
```
##      query significant      p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 1.426609e-03     12345         46               40
## 2 GSE61992        TRUE 4.486260e-02      5487         46               23
## 3 GSE61992        TRUE 4.068265e-02       789         37               10
## 4 GSE61992        TRUE 3.711198e-02        76         44                4
## 5 GSE91061        TRUE 5.489894e-12        74        382               34
## 6 GSE91061        TRUE 1.024985e-06        43        382               20
##    precision
## 1 0.86956522
## 2 0.50000000
## 3 0.27027027
## 4 0.09090909
## 5 0.08900524
```

```
## 6 0.05235602
```

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_TCD8cells_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:

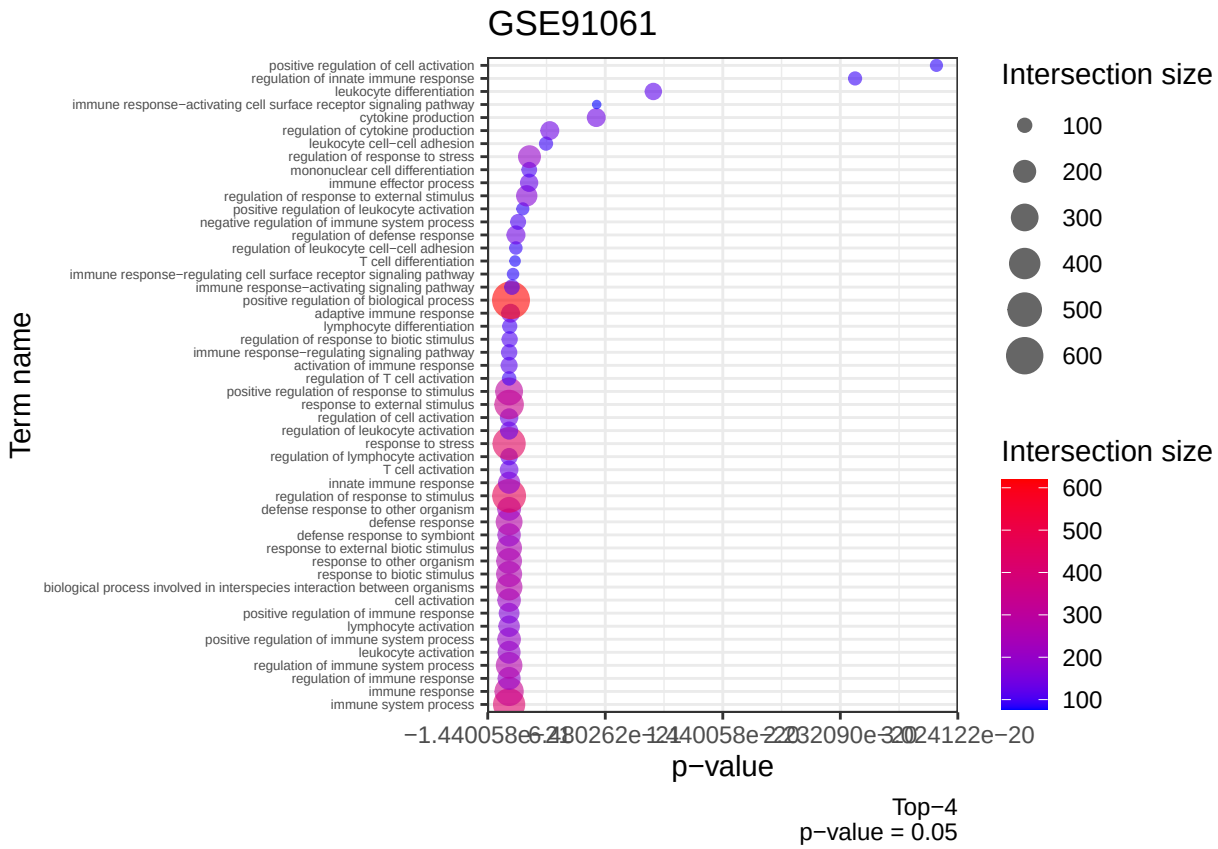
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061
##        4      965
```

```
write.table(gem, file = "./Treatment_comparisons/gProfiler_TCD8cells_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")
```

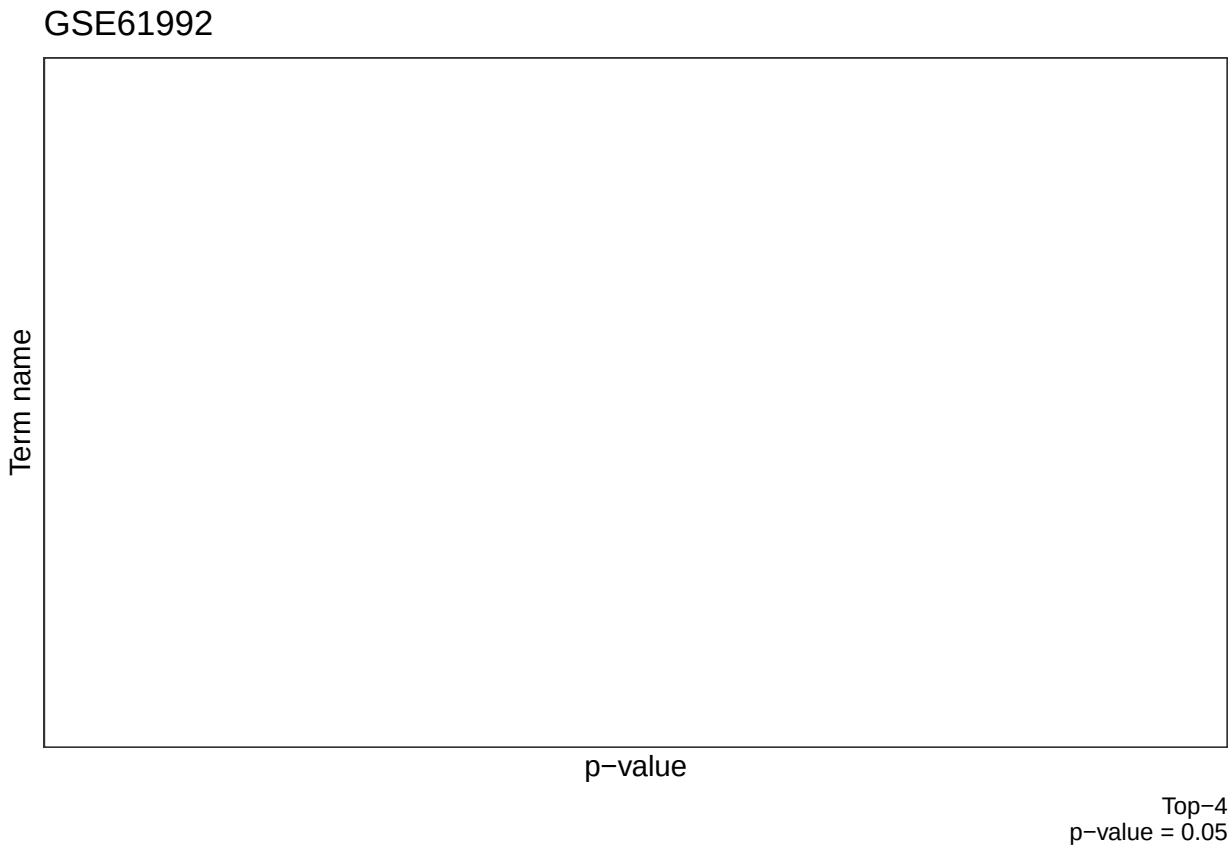
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




```
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.

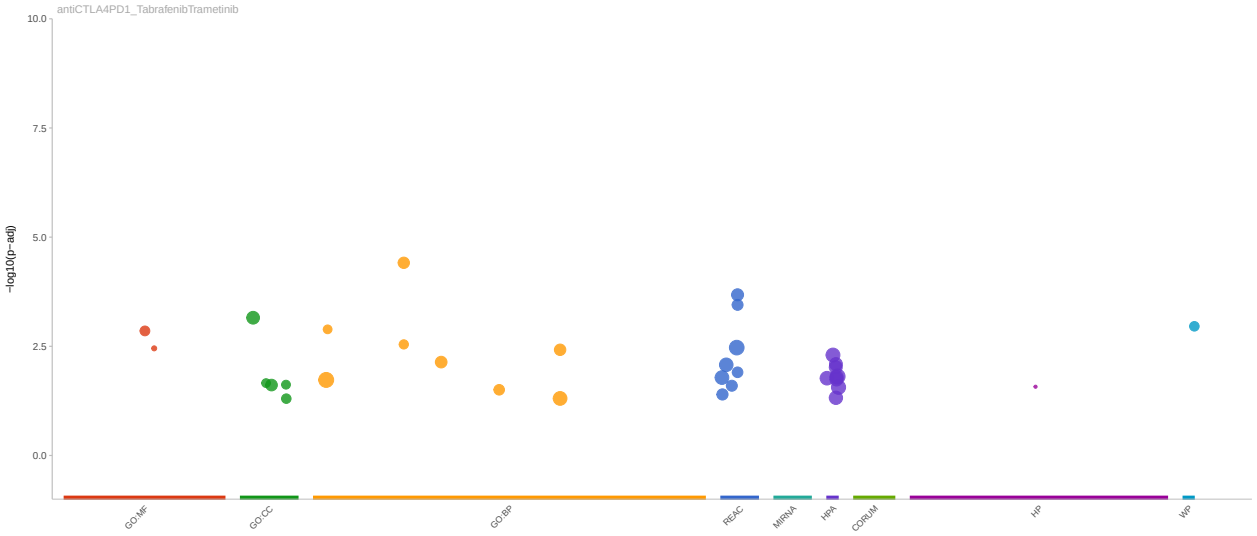
Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list(
"antiCTLA4PD1_TabrafenibTrametinib" = jVenn_TCD8cells$`GSE91061|GSE61992`[which(is.na(jVenn_TCD8cells$`(
),
                      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 antiCTLA4PD1_TabrafenibTrametinib        TRUE 3.867326e-05       115
## 2 antiCTLA4PD1_TabrafenibTrametinib        TRUE 1.292078e-03        24
## 3 antiCTLA4PD1_TabrafenibTrametinib        TRUE 2.859494e-03        31
## 4 antiCTLA4PD1_TabrafenibTrametinib        TRUE 3.796755e-03       124
## 5 antiCTLA4PD1_TabrafenibTrametinib        TRUE 7.271458e-03       146
## 6 antiCTLA4PD1_TabrafenibTrametinib        TRUE 1.861769e-02      2776
##   query_size intersection_size precision
## 1         17                5 0.2941176
## 2         17                3 0.1764706
## 3         17                3 0.1764706
## 4         17                4 0.2352941
## 5         17                4 0.2352941
```

```
## 6         17                 10 0.5882353
```

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

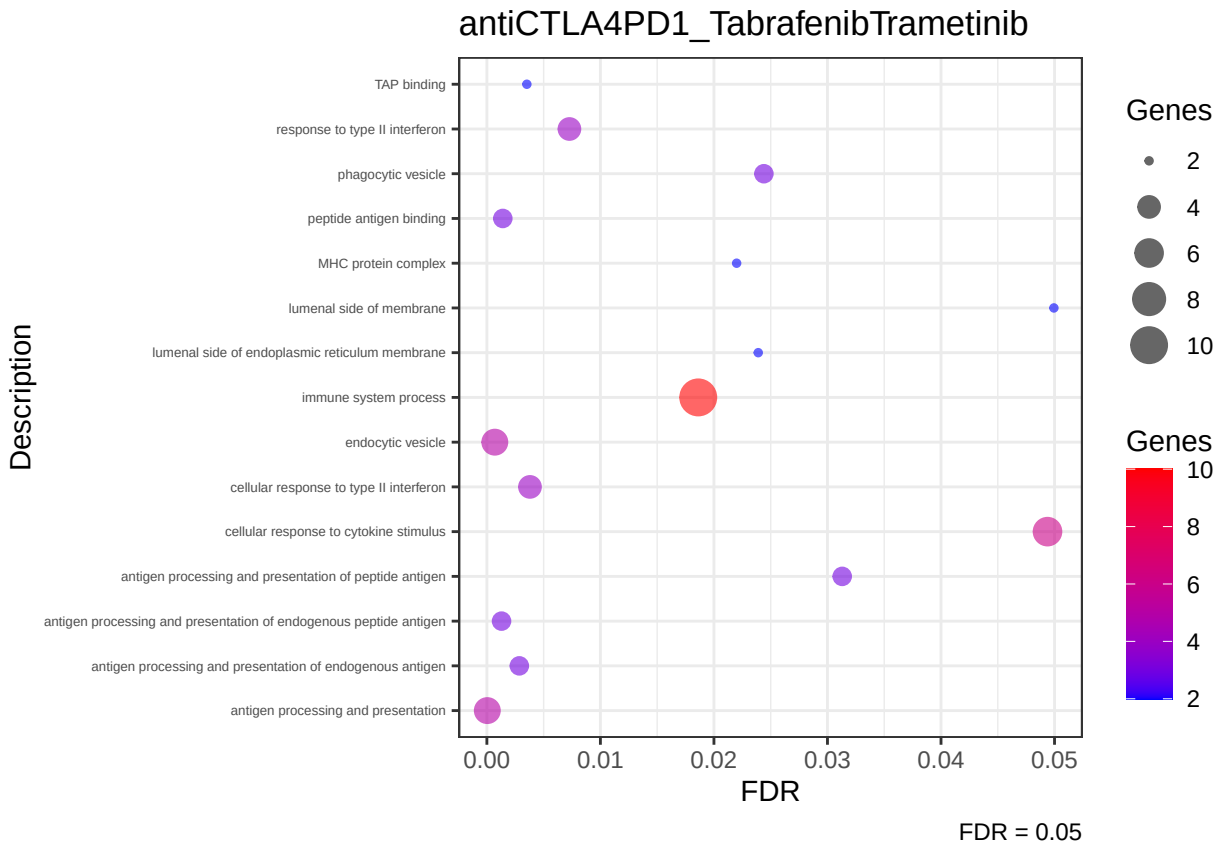


Estos resultados son los que utilizo para generar los archivos gem que archivo para usar más tarde en Cytoscape.

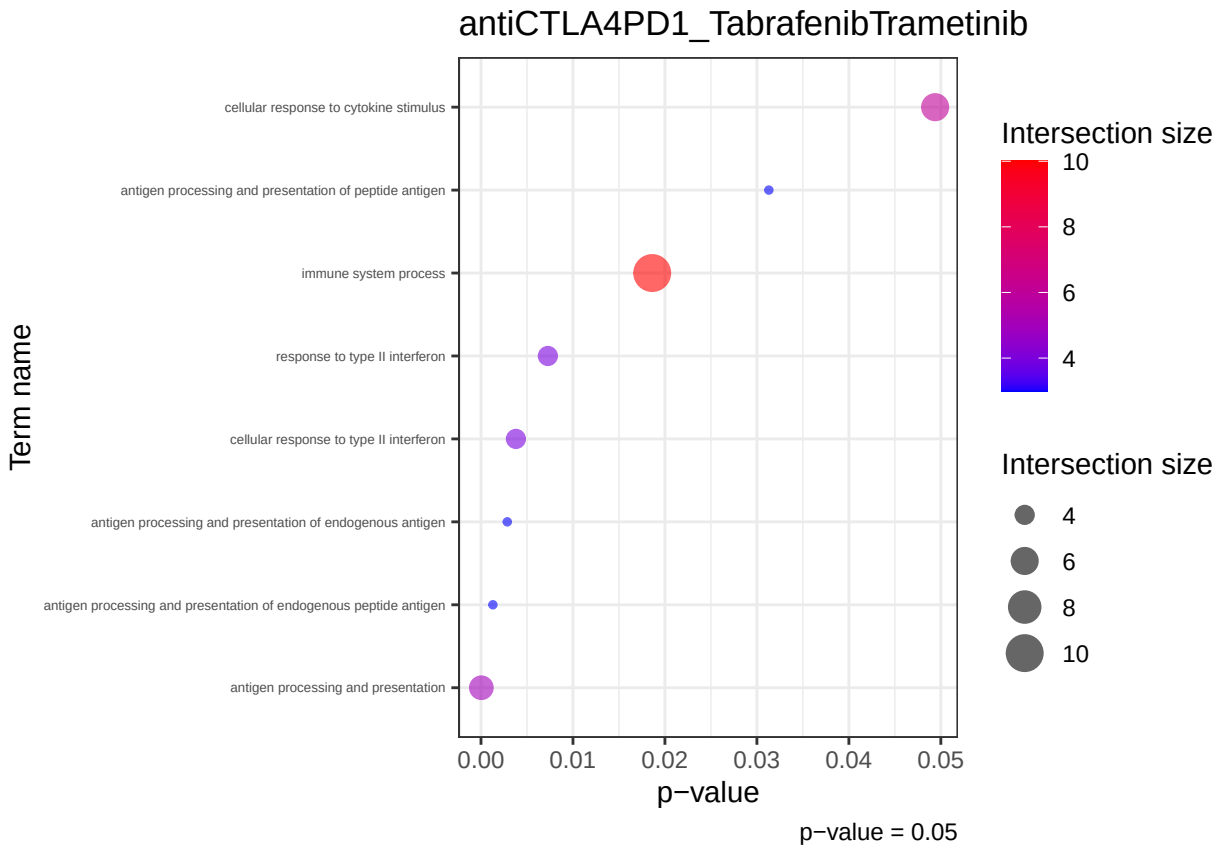
```
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

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_TCD8cells_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("antiCTLA4PD1_TabrafenibTrametinib")
```




```
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 = "antiCTLA4PD1_TabrafenibTrametinib")
```


antiCTLA4PD1_TabrafenibTrametinib
Term name
cellular response to cytokine stimulus
antigen processing and presentation of peptide antigen
immune system process
response to type II interferon
cellular response to type II interferon
antigen processing and presentation of endogenous antigen
antigen processing and presentation of endogenous peptide antigen
antigen processing and presentation
0.00
0.01
0.02
0.03
0.04
0.05
p−value
p−value = 0.05
Intersection size
10
8
6
4
Intersection size
4
6
8
10