Pathway Analysis - NK 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 NK obtenidas en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_NKcells_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_NK
head(jVenn_NKcells_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509 GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>    <chr>       <chr>
## 1 LPIN1    AFP         AUTS2
## 2 TRAPPC2  RNPS1       BCL2
## 3 DKC1     TRIP11      ZNF318
## 4 SLC2A12  PHF1        PUM1
## 5 RCBTB2   CDK6        NABP2
## 6 RPS6KA3  SRD5A3      GNG2
```

```
jVenn_NKcells_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_NK
jVenn_NKcells_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_NKCel
jVenn_NKcells_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_NKCel
jVenn_NKcells_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_NKCel
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_NKcells <- read_delim("~/Desktop/ELENA_UOC/TFM/DEGs/Venn_diagrams_cell_type_treatments_comparison/
    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:

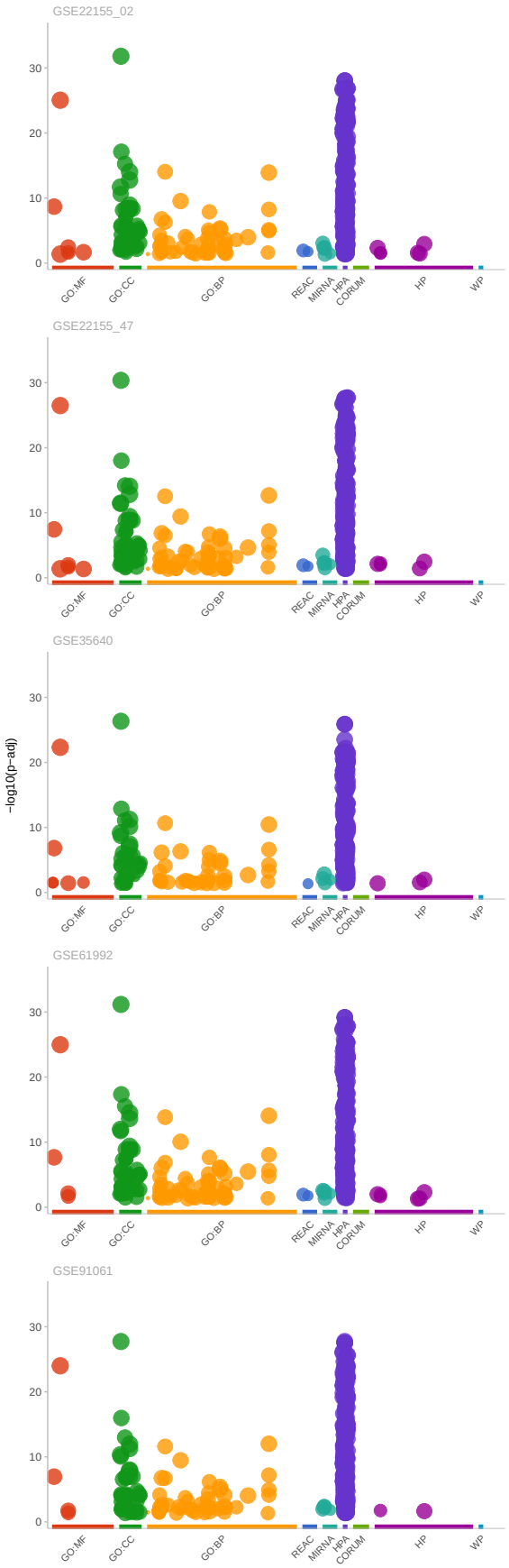
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_NKcells_vs_GSE22155_02$GSE50509[which(is
                              "GSE22155_47" = jVenn_NKcells_vs_GSE22155_47$GSE50509[which(is.na(jVenn_NK
                              "GSE35640" = jVenn_NKcells_vs_GSE35640$GSE50509[which(is.na(jVenn_NKcells_
                              "GSE91061" = jVenn_NKcells_vs_GSE91061$GSE50509[which(is.na(jVenn_NKcells_
                              "GSE61992" = jVenn_NKcells_vs_GSE61992$GSE50509[which(is.na(jVenn_NKcells_
                              ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##         query significant      p_value term_size query_size intersection_size
## 1 GSE22155_02        TRUE 9.128453e-15      2556       1325              268
## 2 GSE22155_02        TRUE 1.231616e-14      5986       1325              518
## 3 GSE22155_02        TRUE 2.919690e-10      4912       1325              423
## 4 GSE22155_02        TRUE 5.560949e-09      2074       1325              209
## 5 GSE22155_02        TRUE 1.320218e-08      1600       1325              170
## 6 GSE22155_02        TRUE 1.789227e-07      3613       1325              317
##   precision
## 1 0.2022642
## 2 0.3909434
## 3 0.3192453
## 4 0.1577358
## 5 0.1283019
## 6 0.2392453
```

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


GSE22155_02
GSE22155_47
GSE35640
GSE61992
GSE91061
−log10(p-adj)
0
10
20
30
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_NKcells_Only_
                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
##         335         335         283         338         316
```

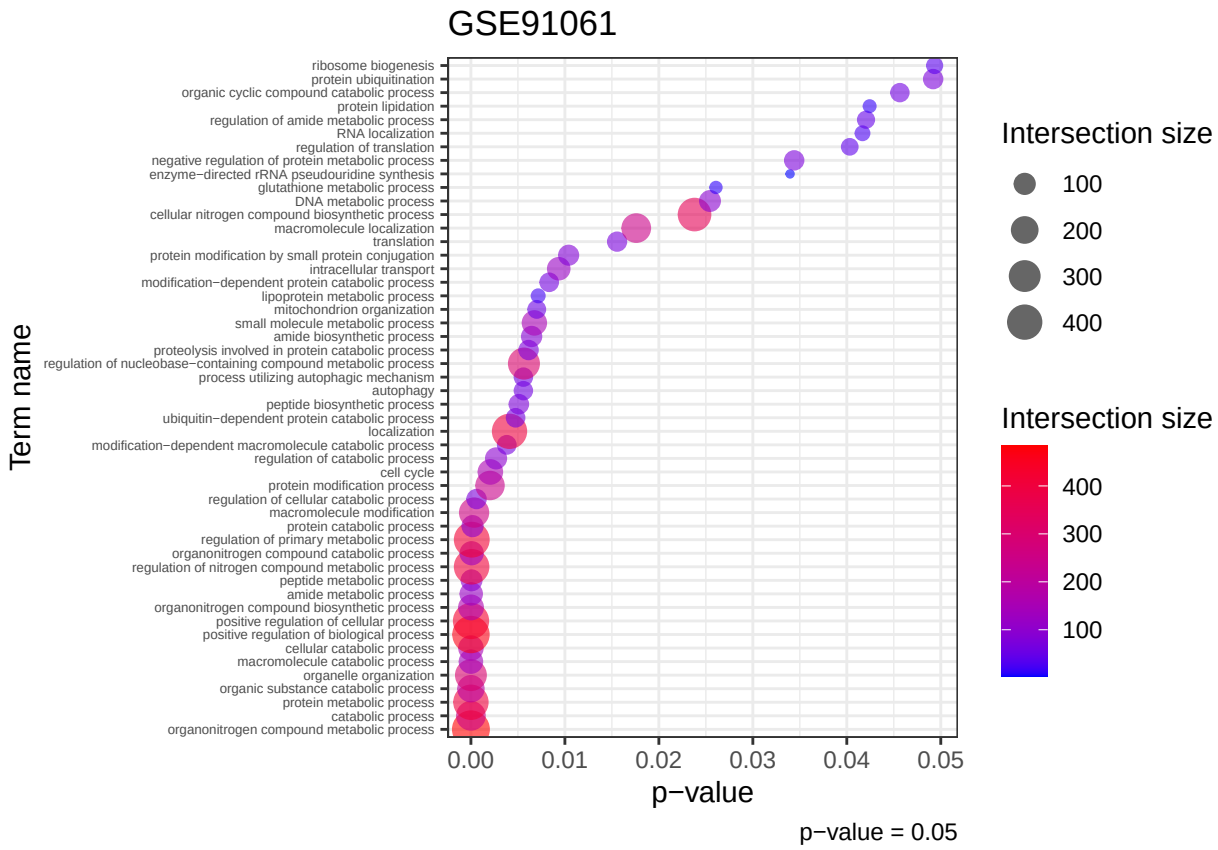
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061
##    20.84630    20.84630    17.61045    21.03298    19.66397
```

```
#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/NKcells_GOBPs.txt", sep = "\t
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/NKcells_GOBPs_freq.txt", sep
#rm(jVenn_NKcells_vs_GSE22155_02, jVenn_NKcells_vs_GSE22155_47, jVenn_NKcells_vs_GSE35640, jVenn_NKcell
```

Barplot of the top GO-BPs:

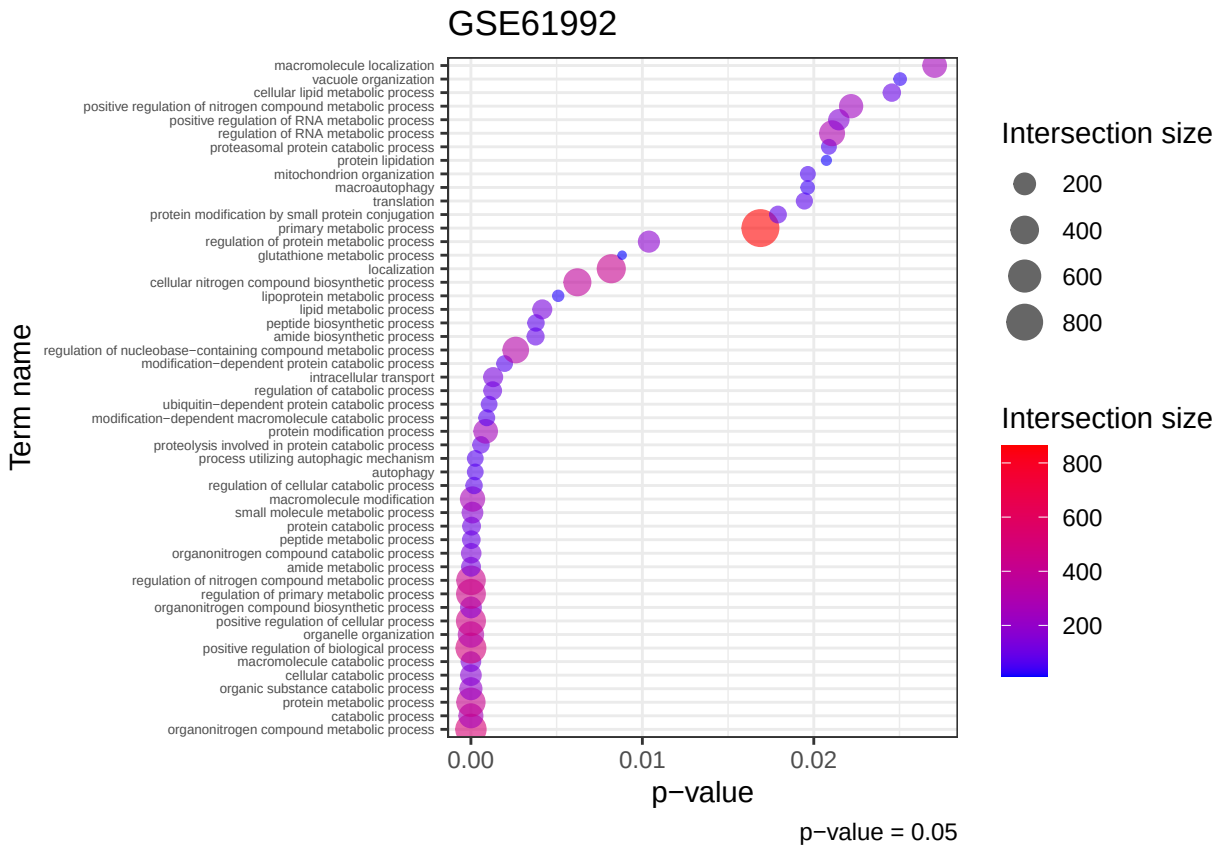
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size, size = intersection_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 = "GSE91061")
```




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

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

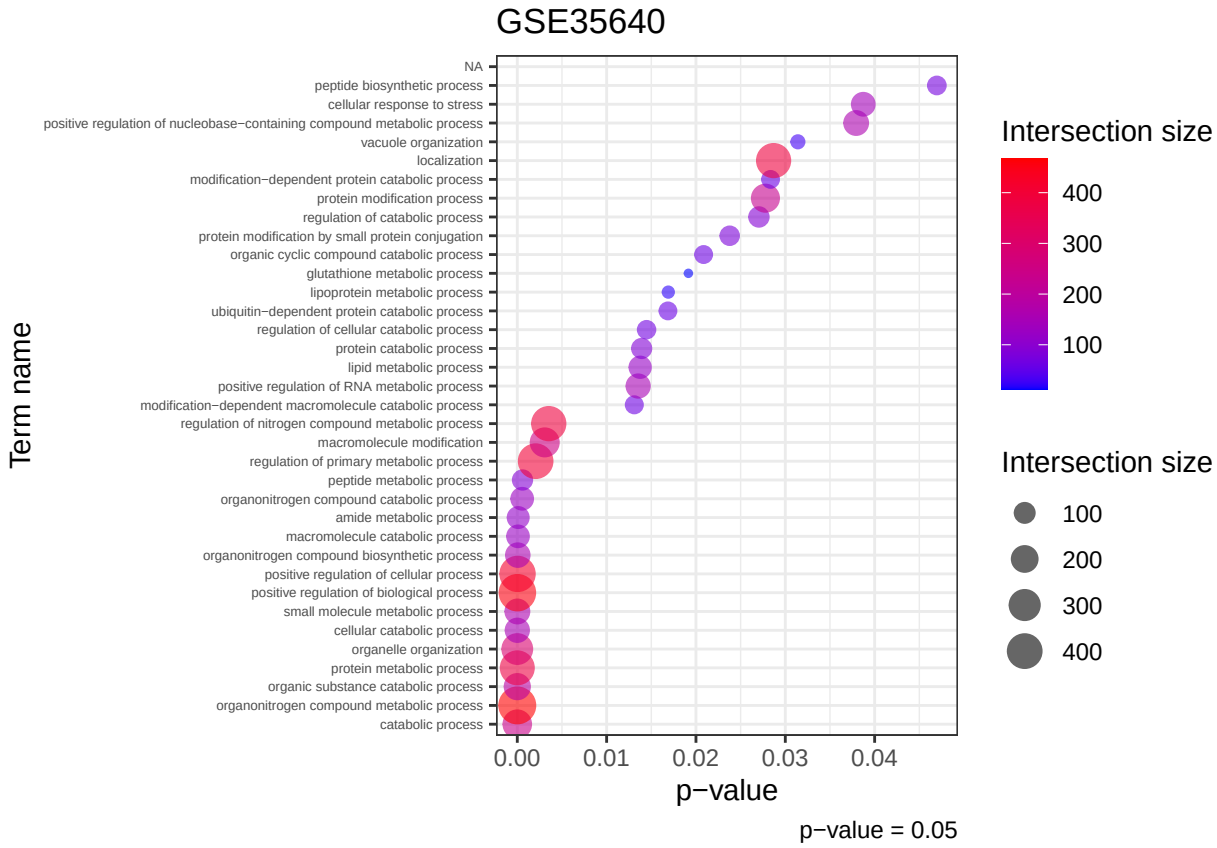
```
plot_gobps(study = "GSE61992")
```




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

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

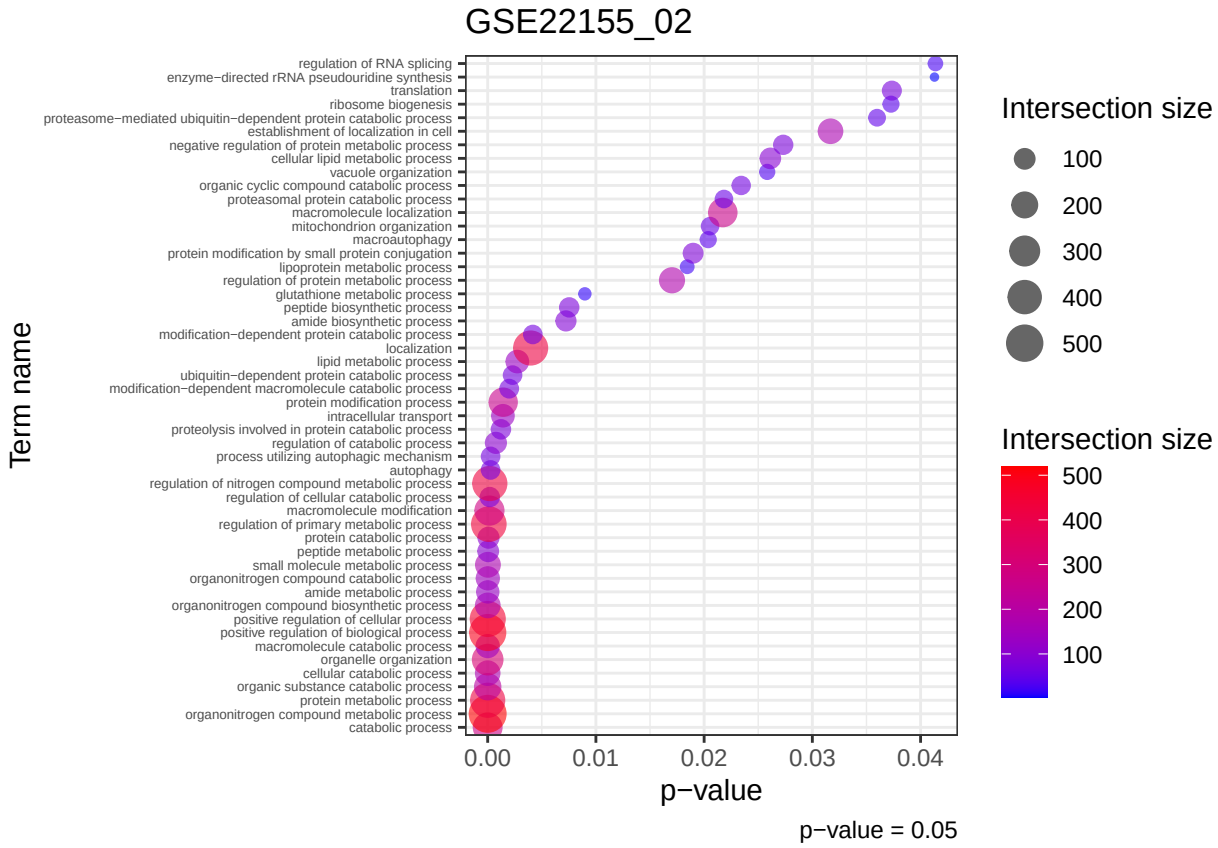
```
plot_gobps(study = "GSE35640")
```




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

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

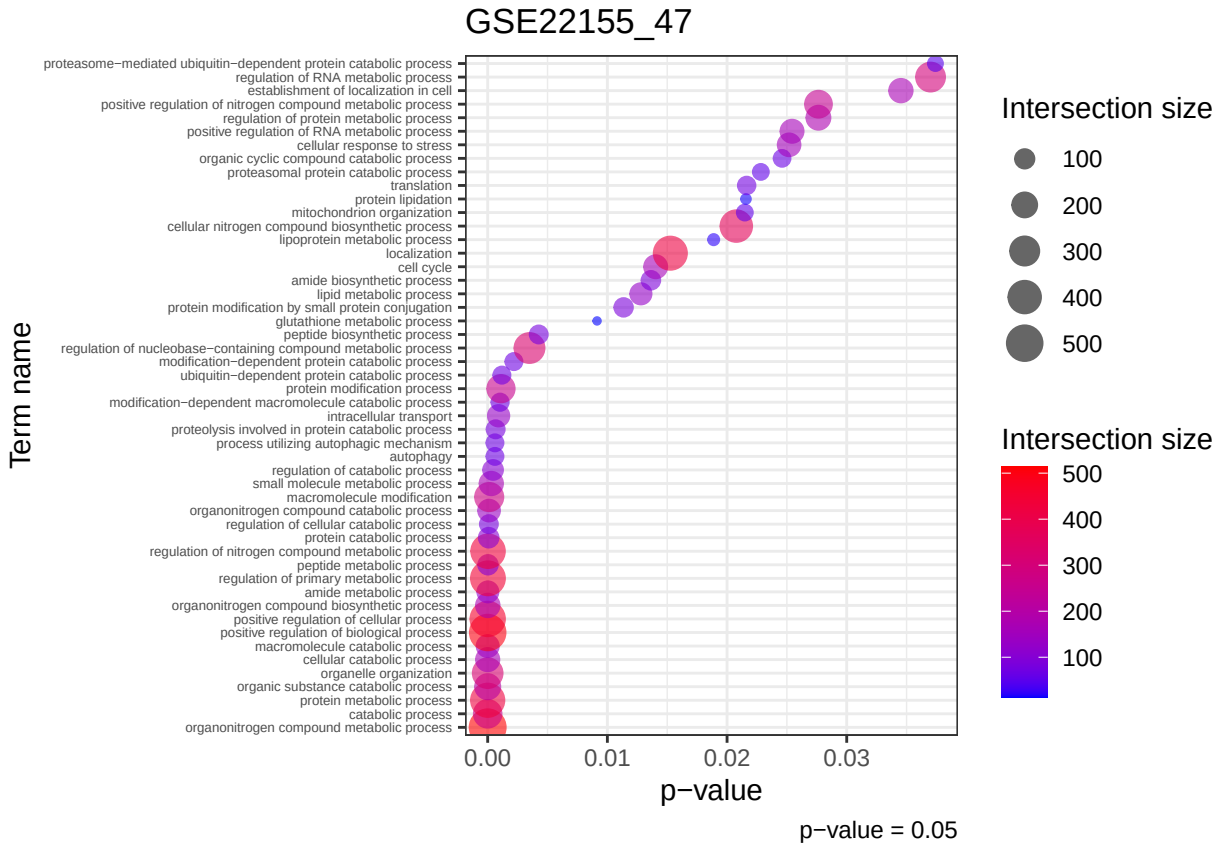
```
plot_gobps(study = "GSE22155_02")
```




```
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_NKcells$GSE61992[which(is.na(jVenn_NKcells$GSE6199
                                   "GSE91061" = jVenn_NKcells$GSE91061[which(is.na(jVenn_NKcells$GSE91061) ==
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

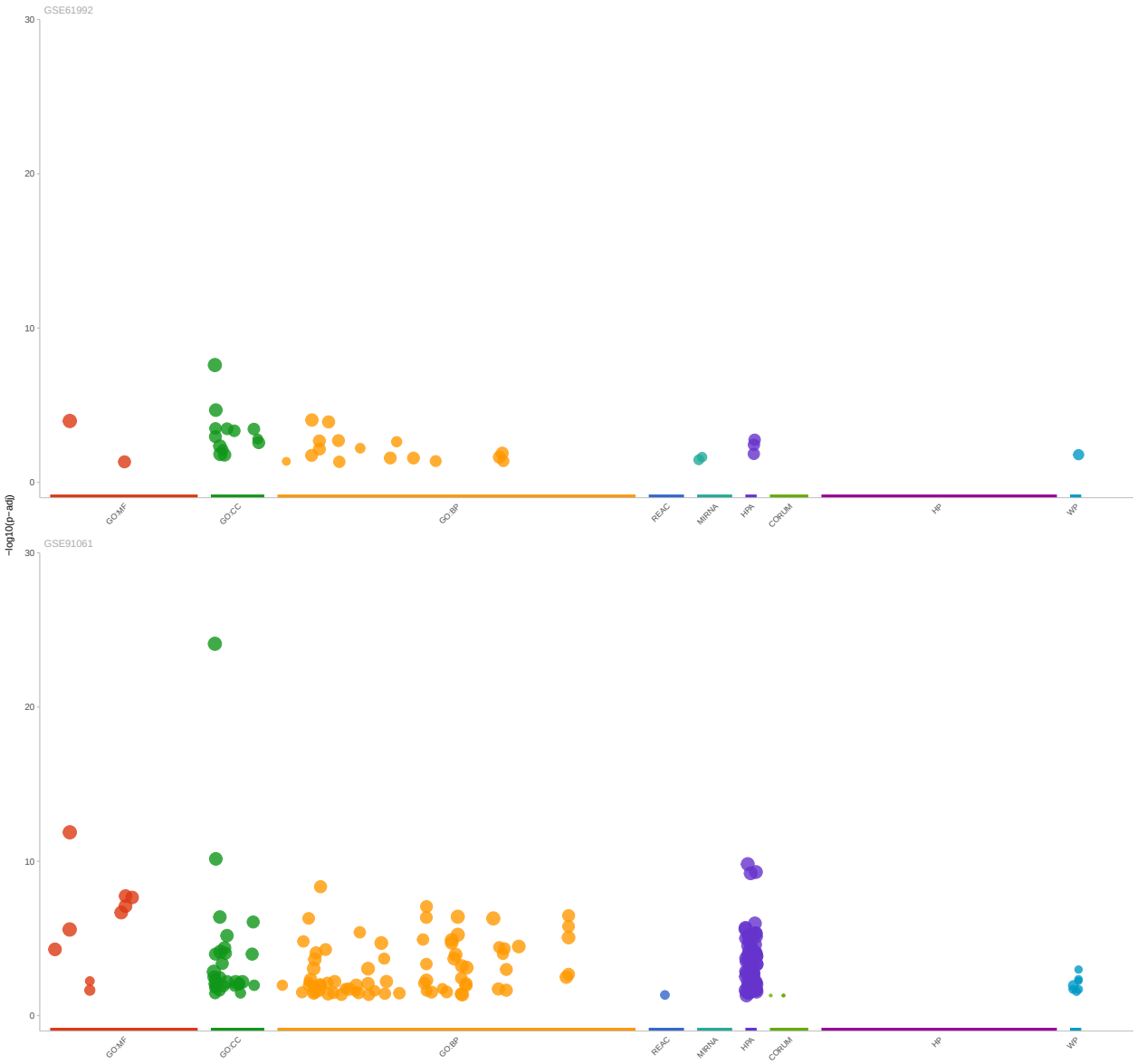
```
##      query significant      p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 9.193374e-05      3855        163               59
## 2 GSE61992        TRUE 1.217385e-04      2637        163               46
## 3 GSE61992        TRUE 1.980644e-03      1984        163               36
## 4 GSE61992        TRUE 2.077135e-03      1988        163               36
## 5 GSE61992        TRUE 2.343299e-03       220        163               11
## 6 GSE61992        TRUE 6.173393e-03       115        163                8
##    precision
## 1 0.36196319
## 2 0.28220859
## 3 0.22085890
## 4 0.22085890
## 5 0.06748466
```

```
## 6 0.04907975
```

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_NKcells_Only_geneset_", unique(.y$query
                sep = "\t", quote = F, row.names = F))
```

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



Proporciones de los pathways para cada estudio:

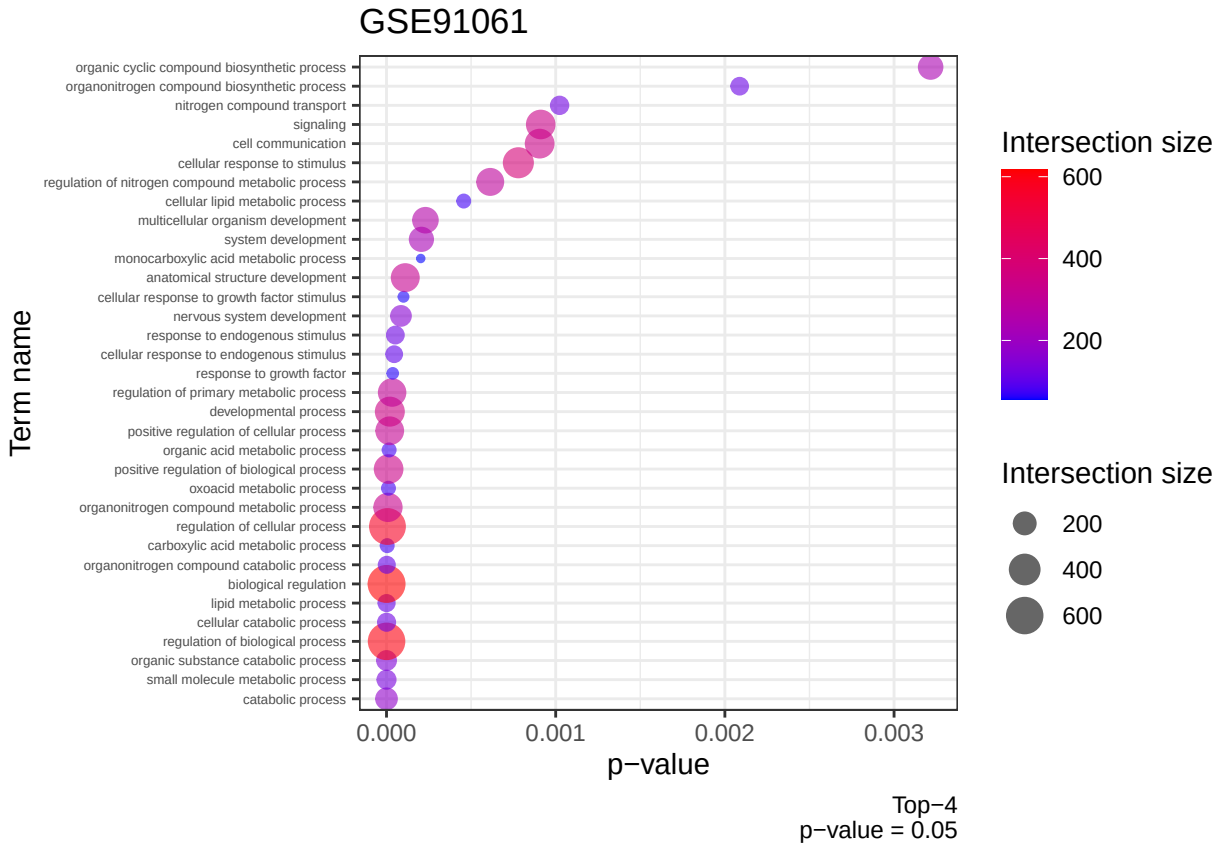
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061
##       37      211
```

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

```
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")
```

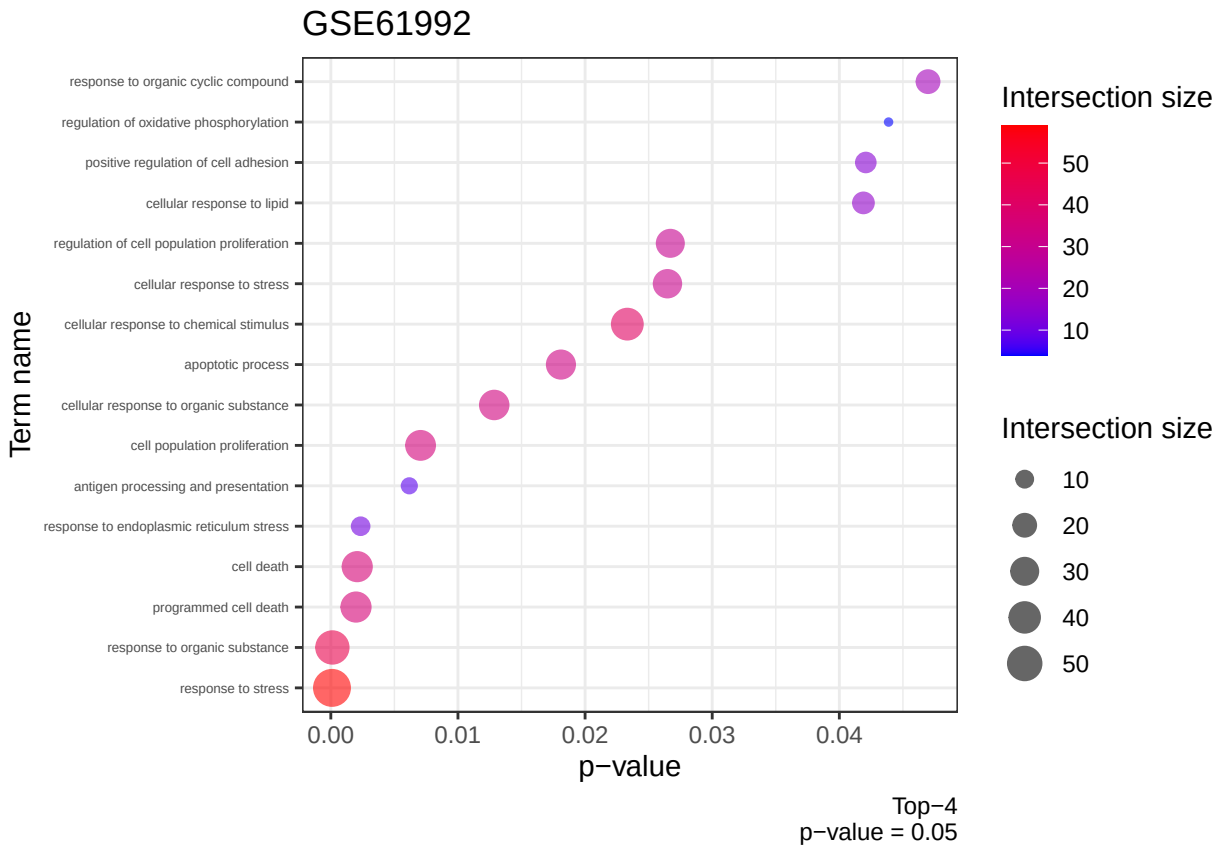
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




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

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

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



6 Máximo solapamiento

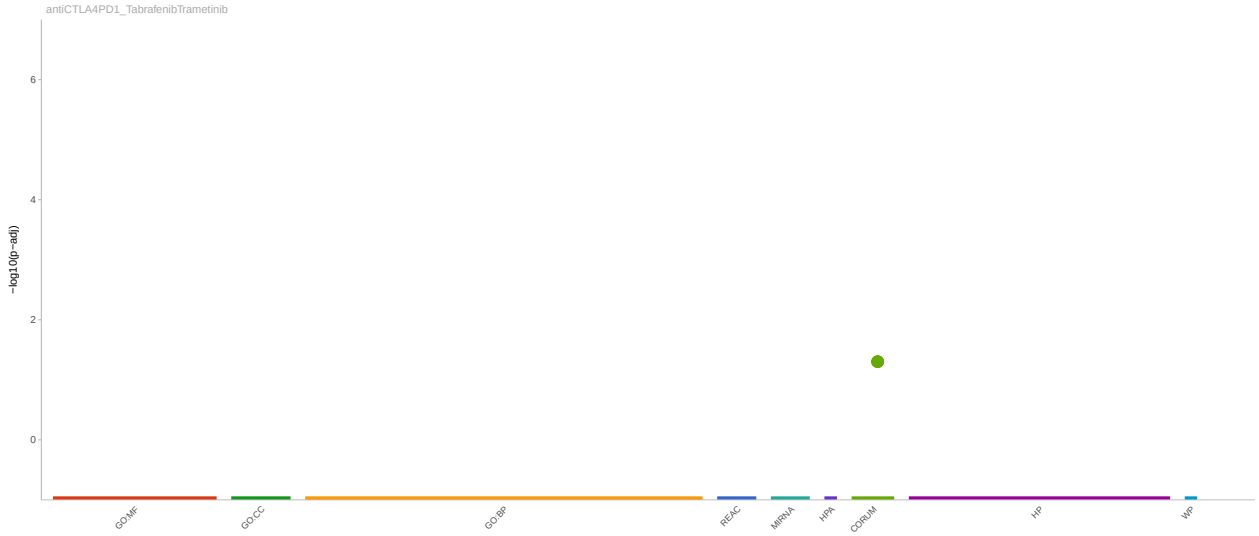
En esta sección analizo los DEGs que más se comparten entre los estudios para observar los mecanismos que pudieran haber en común entre los tratamientos.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list(
                                   "antiCTLA4PD1_TabrafenibTrametinib" = jVenn_NKcells$`GSE91061|GSE6199
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##                                query significant    p_value term_size query_size
## 1 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.04993169         2          1
## 2 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.04993169         2          1
## 3 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.04993169         2          1
## 4 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.04993169         2          1
##   intersection_size precision
## 1                 1         1
## 2                 1         1
## 3                 1         1
## 4                 1         1
```

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



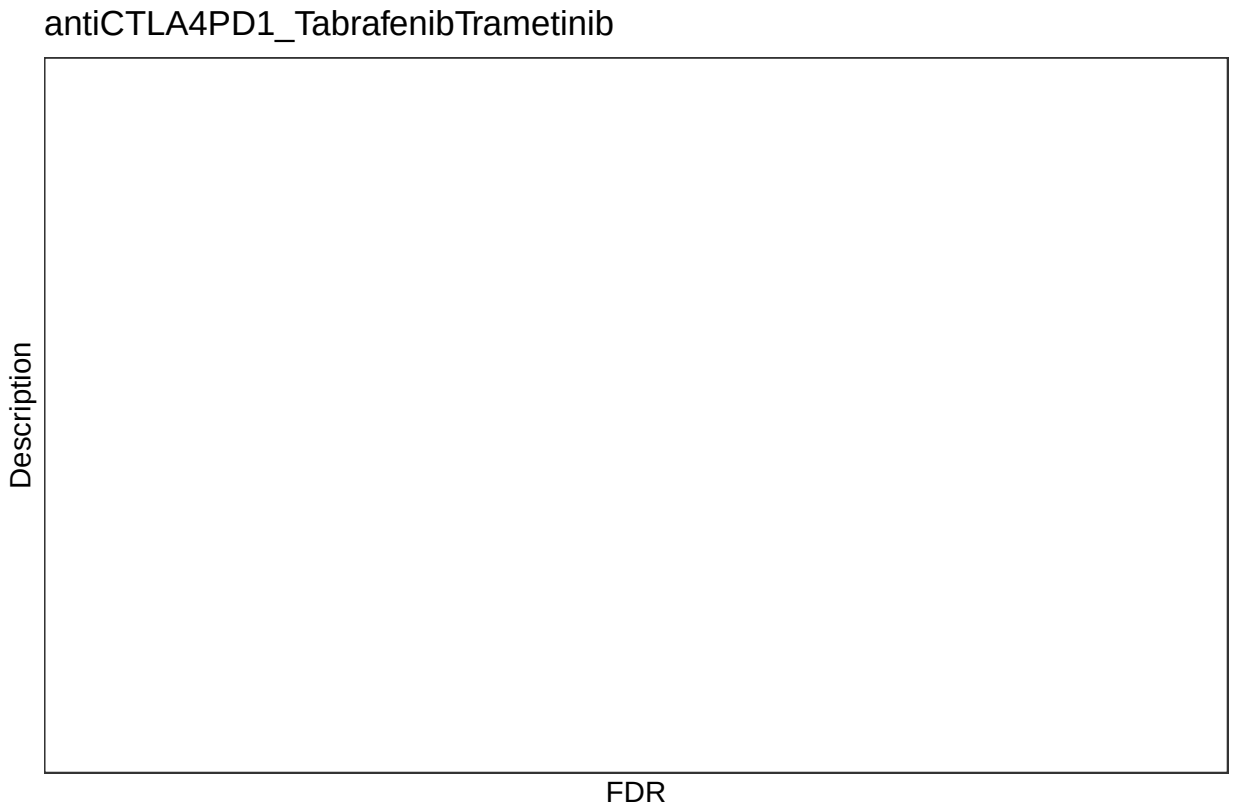
Estos resultados son los que utilizo para generar los archivos gem que archivo para usar más tarde en Cytoscape.

```
# 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_NKcells_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.

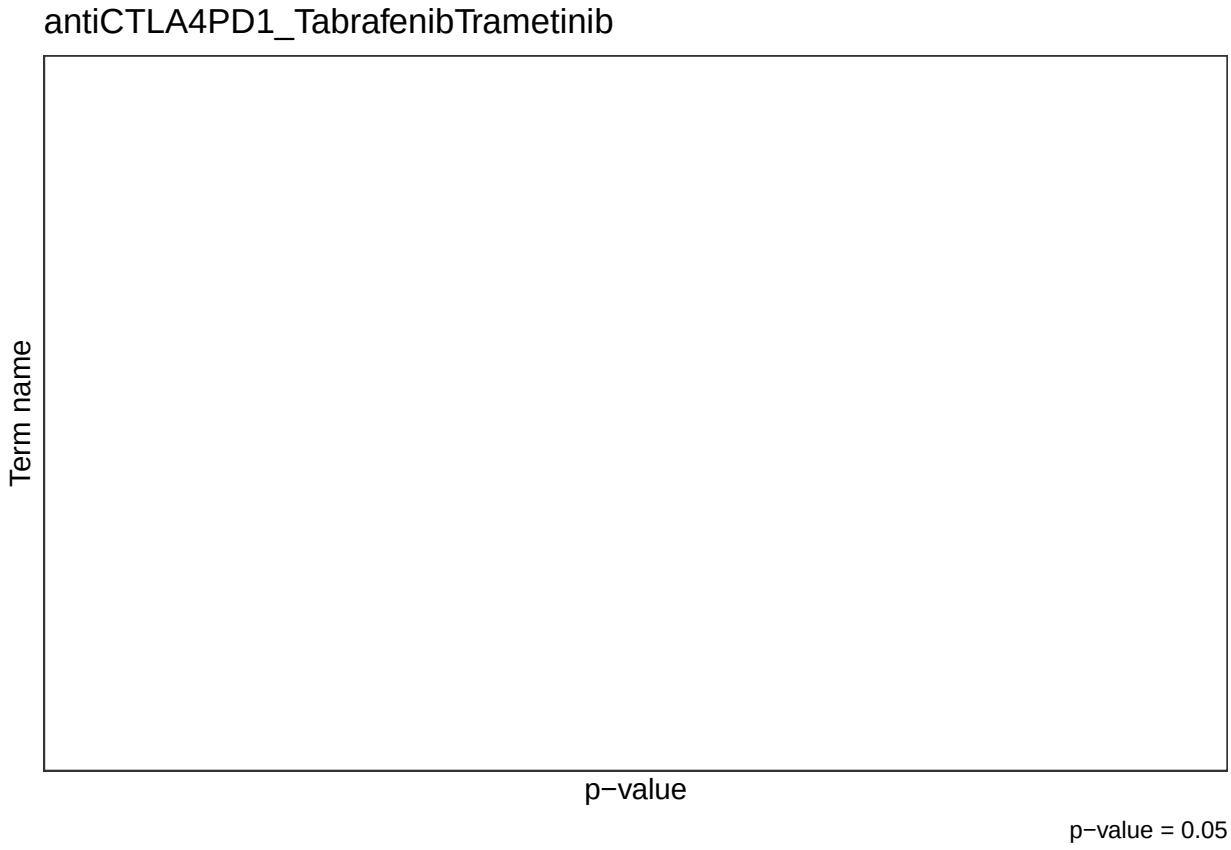
En este caso no hay resultado de GOBPs.

```
#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")
```



FDR = 0.05

```
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
p−value
p−value = 0.05