# Pathway Analysis - Monocytes
# 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 Monocitos obtenidos en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_Monocytes_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
head(jVenn_Monocytes_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509  GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>     <chr>       <chr>
## 1 CEMIP     TLR2        SAT2
## 2 LINC01091 ERGIC2      KLHL8
## 3 DDX23     CCDC59      SARM1
## 4 SLC50A1   IRAK3       WDR75
## 5 RNASE6    ANKDD1A     EMC6
## 6 DYRK4     PARK7       NEXN
```

```
jVenn_Monocytes_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_
jVenn_Monocytes_vs_tcgaskcm <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Mon
jVenn_Monocytes_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Mon
jVenn_Monocytes_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Mon
jVenn_Monocytes_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn_Mon
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_Monocytes <- read_delim("~/Desktop/ELENA_UOC/TFM/DEGs/Venn_diagrams_cell_type_treatments_compariso
    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:

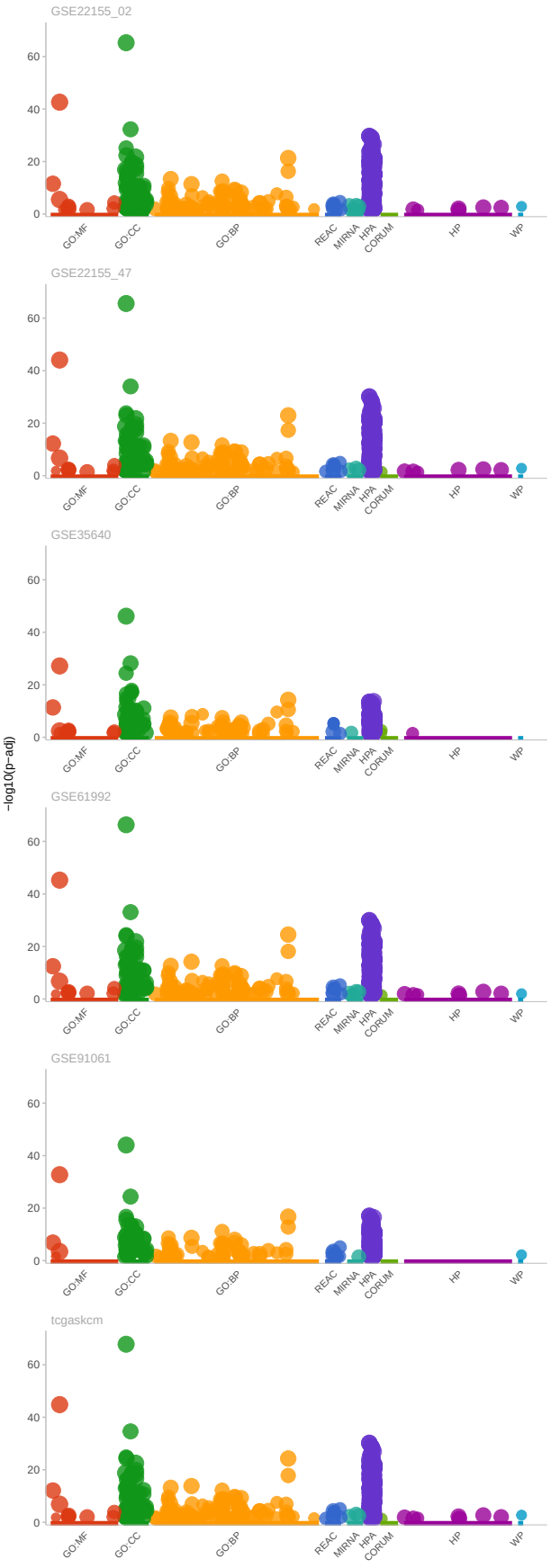
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_Monocytes_vs_GSE22155_02$GSE50509[which
                                   "GSE22155_47" = jVenn_Monocytes_vs_GSE22155_47$GSE50509[which(is.na(jVenn_
                                   "GSE35640" = jVenn_Monocytes_vs_GSE35640$GSE50509[which(is.na(jVenn_Monocy
                                   "GSE91061" = jVenn_Monocytes_vs_GSE91061$GSE50509[which(is.na(jVenn_Monocy
                                   "GSE61992" = jVenn_Monocytes_vs_GSE61992$GSE50509[which(is.na(jVenn_Monocy
                                   "tcgaskcm" = jVenn_Monocytes_vs_tcgaskcm$GSE50509[which(is.na(jVenn_Monocy
                                   ),
                      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 4.732670e-22      5986       1774              699
## 2 GSE22155_02        TRUE 5.229098e-17      1761       1774              260
## 3 GSE22155_02        TRUE 4.188504e-14      3613       1774              439
## 4 GSE22155_02        TRUE 2.620437e-13      1194       1774              185
## 5 GSE22155_02        TRUE 4.039479e-12      4912       1774              553
## 6 GSE22155_02        TRUE 1.690752e-10       911       1774              144
##    precision
## 1 0.39402480
## 2 0.14656144
## 3 0.24746336
## 4 0.10428410
## 5 0.31172492
## 6 0.08117249
```

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


GSE22155_02
GSE22155_47
GSE35640
GSE61992
GSE91061
tcgaskcm
−log10(p-adj)
0
20
40
60
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_Monocytes_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    tcgaskcm
##         422         406         278         408         287         419
```

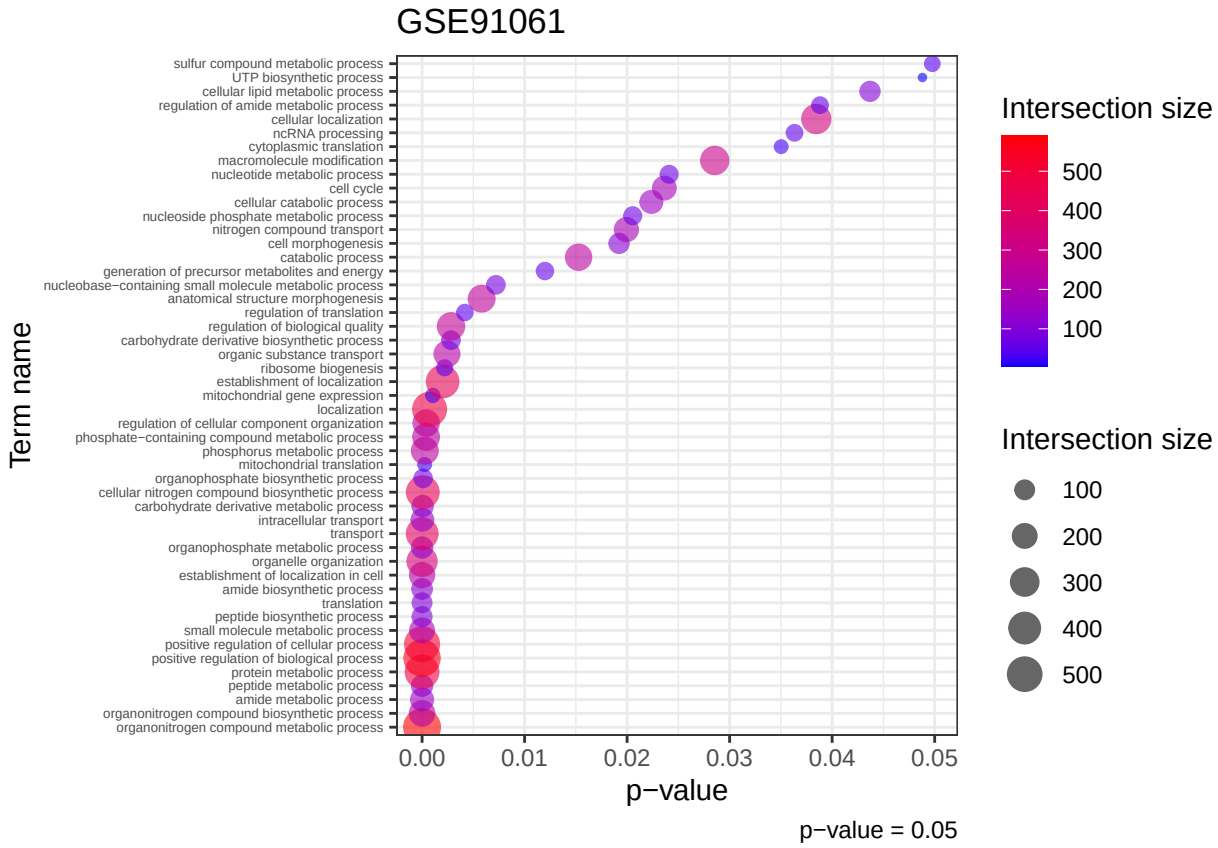
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061    tcgaskcm
##    19.00901    18.28829    12.52252    18.37838    12.92793    18.87387
```

```
#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/Monocytes_GOBPs.txt", sep = "
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/Monocytes_GOBPs_freq.txt", s
#rm(jVenn_Monocytes_vs_GSE22155_02, jVenn_Monocytes_vs_GSE22155_47, jVenn_Monocytes_vs_GSE35640, jVenn_
```

Barplot of the top GO-BPs:

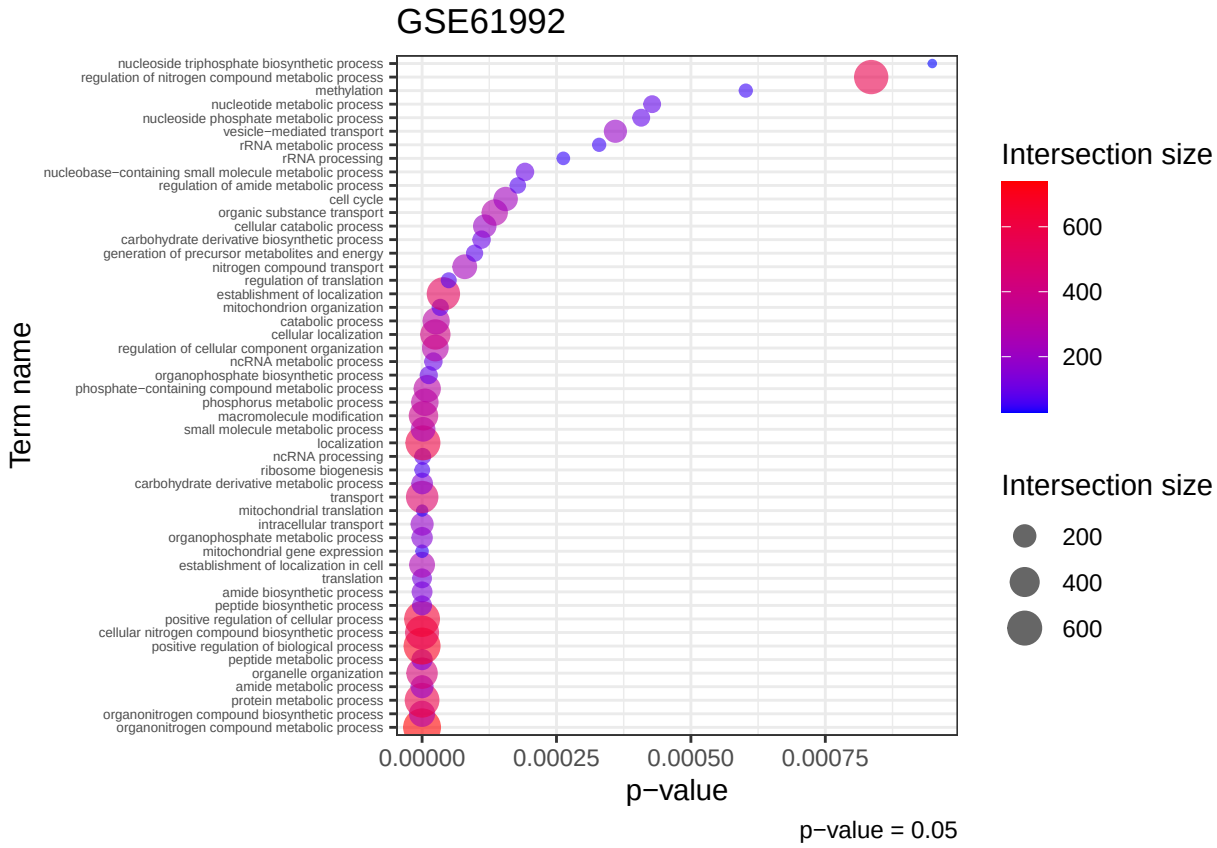
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size, size = intersection_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", n=49)
```




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

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

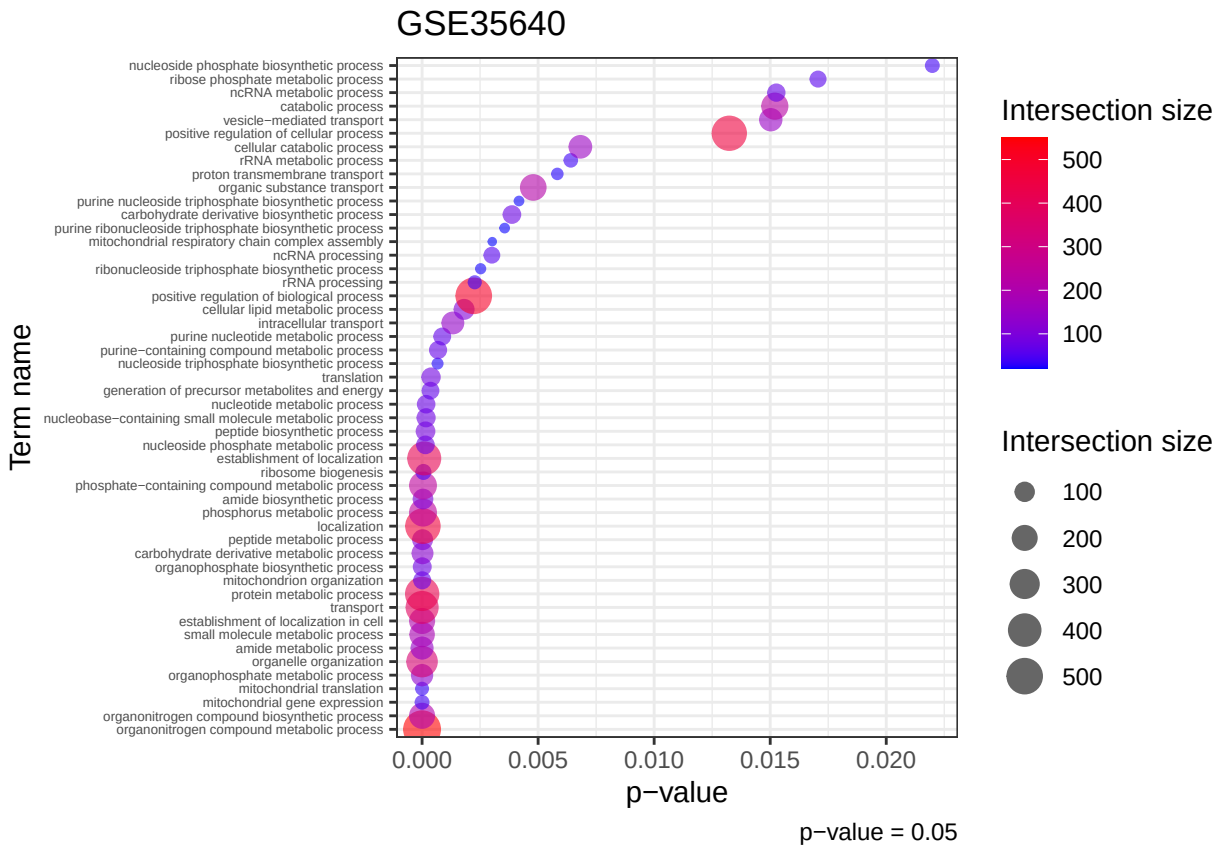
```
plot_gobps(study = "GSE61992")
```




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

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

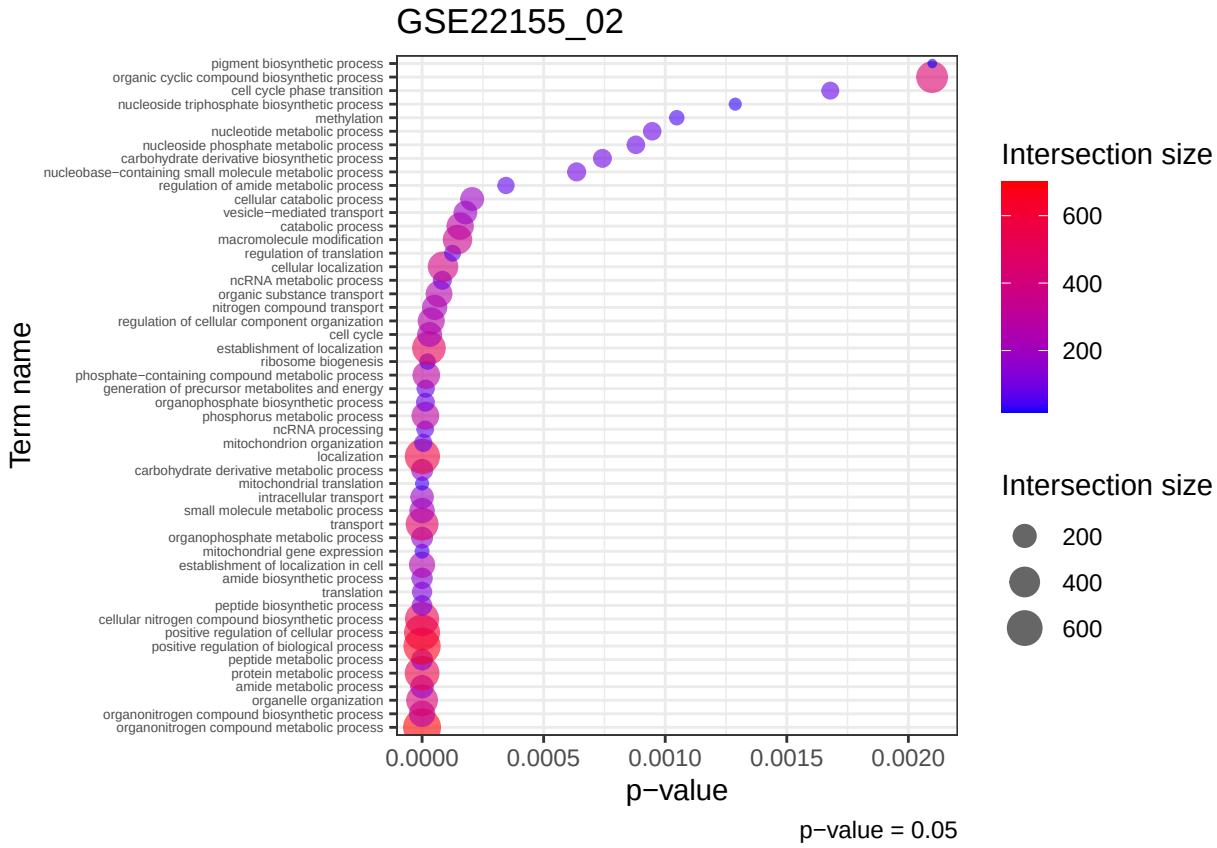
```
plot_gobps(study = "GSE35640")
```




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

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

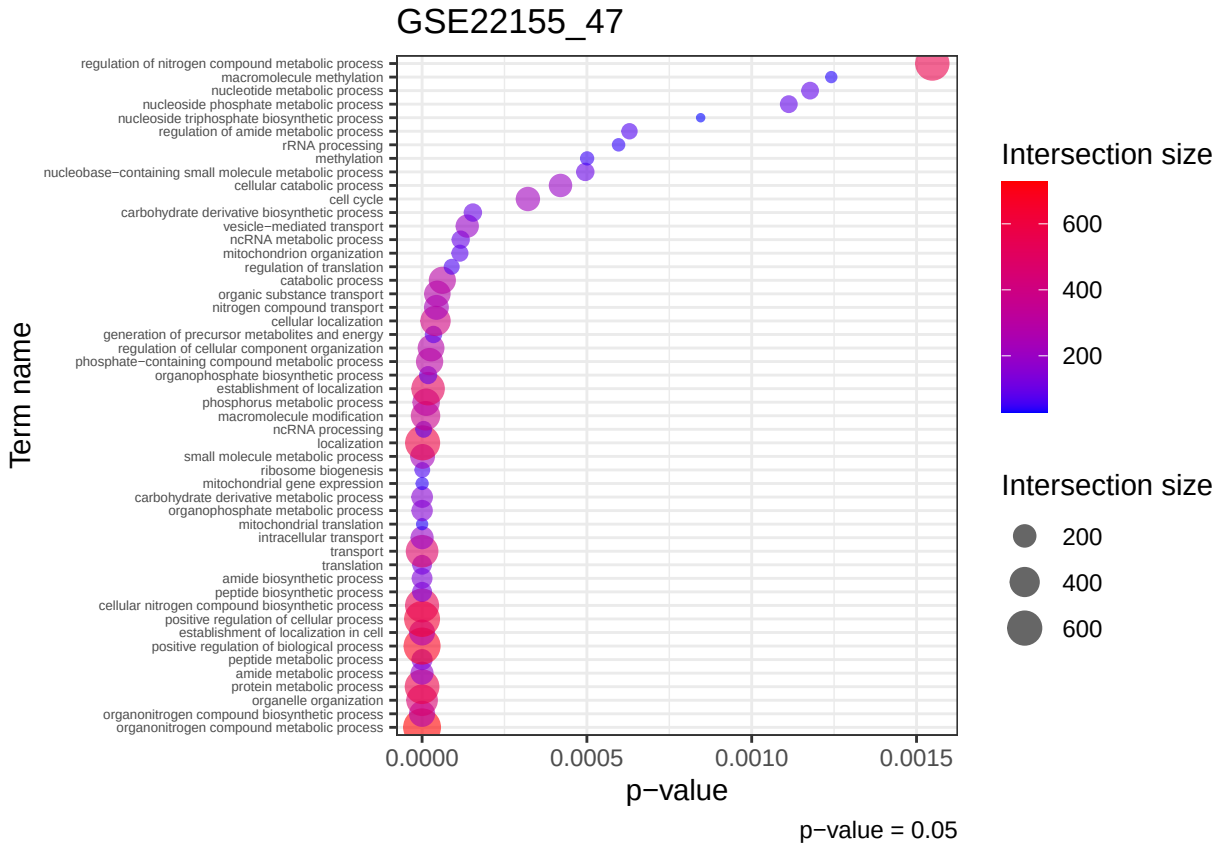
```
plot_gobps(study = "GSE22155_02")
```




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

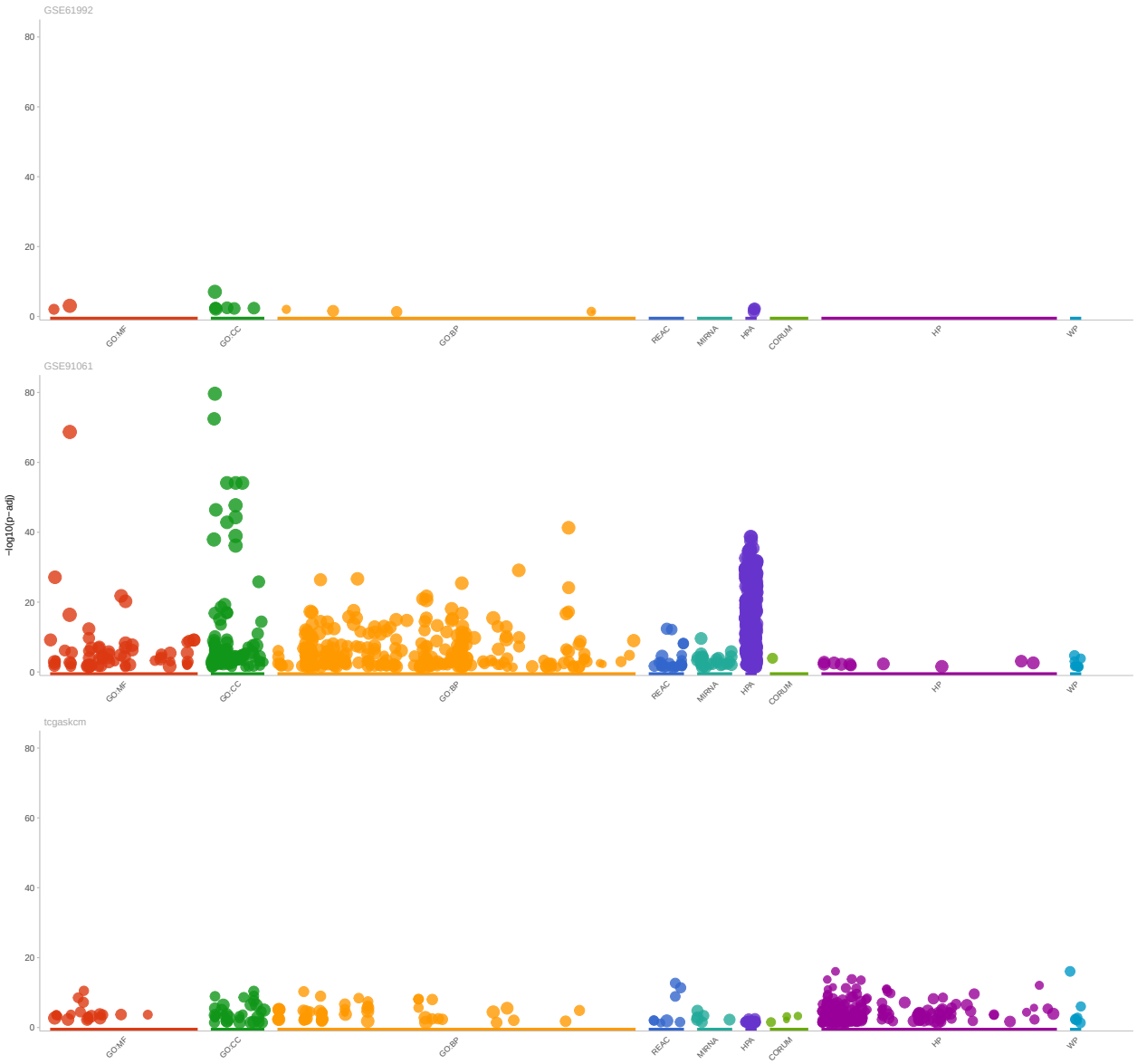
```
##      query significant      p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 9.247703e-03        22        113                 4
## 2 GSE61992        TRUE 2.554621e-02       496        113                12
## 3 GSE61992        TRUE 3.832242e-02        31        113                 4
## 4 GSE61992        TRUE 4.343346e-02       220        113                 8
## 5 GSE61992        TRUE 4.933661e-02         2        113                 2
## 6 GSE61992        TRUE 8.613609e-08     12345        115                96
##    precision
## 1 0.03539823
## 2 0.10619469
## 3 0.03539823
## 4 0.07079646
```

```
## 5 0.01769912
## 6 0.83478261
```

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_Monocytes_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:

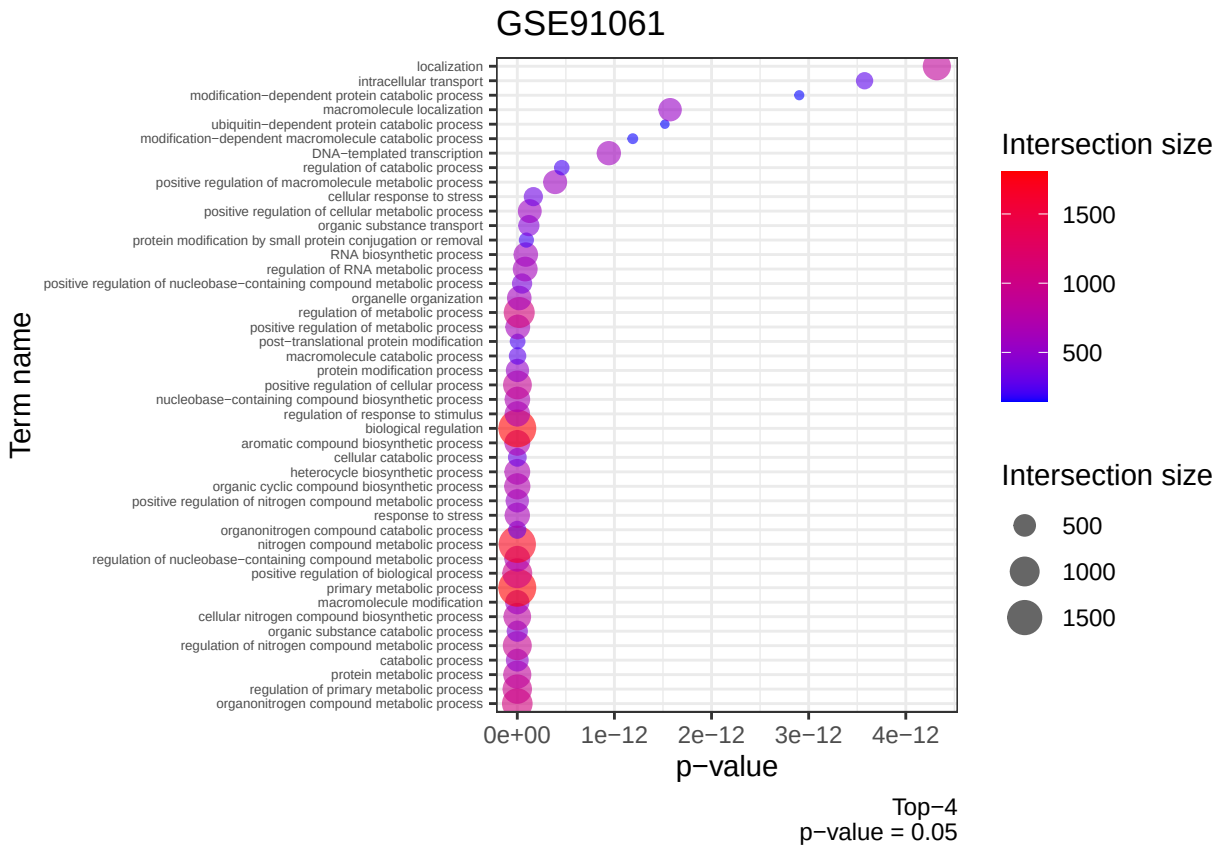
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061 tcgaskcm
##       16      795      328
```

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

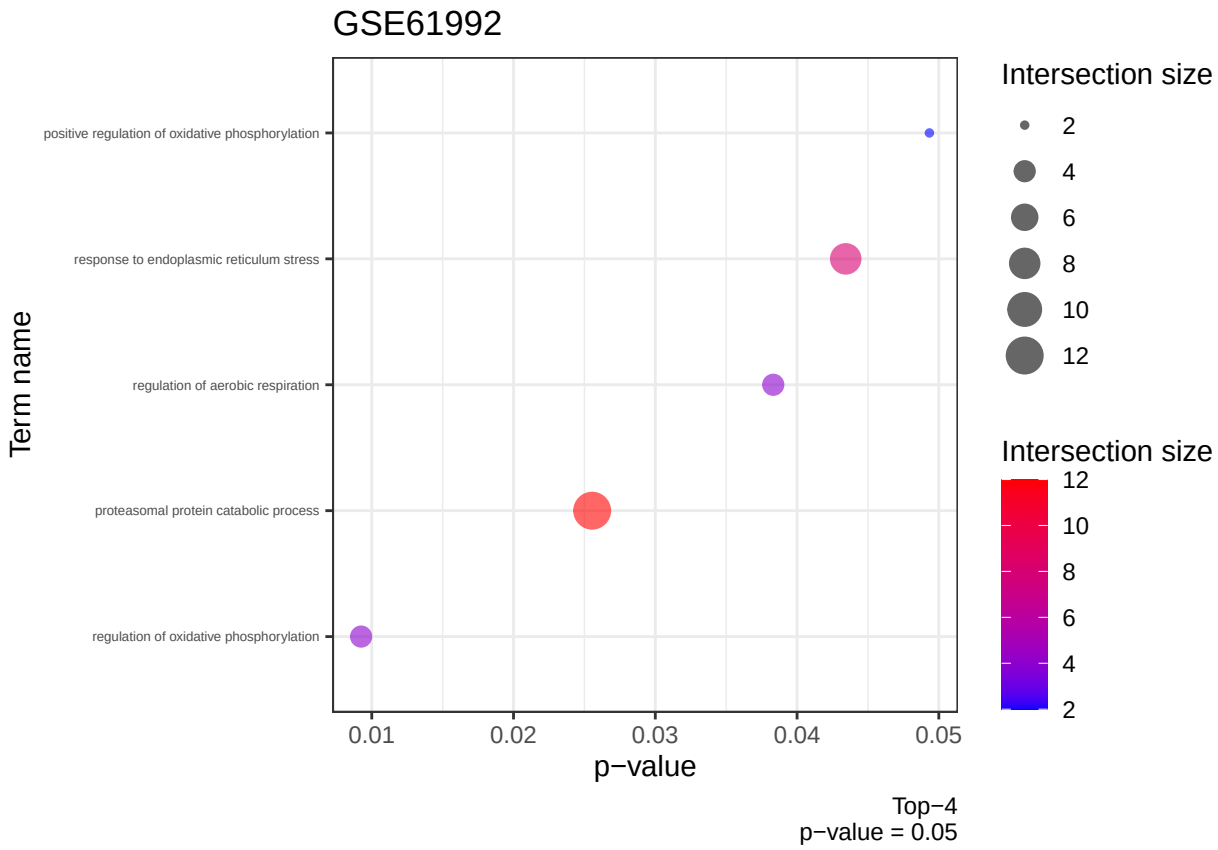
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




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

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

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




# 6 Máximo solapamiento

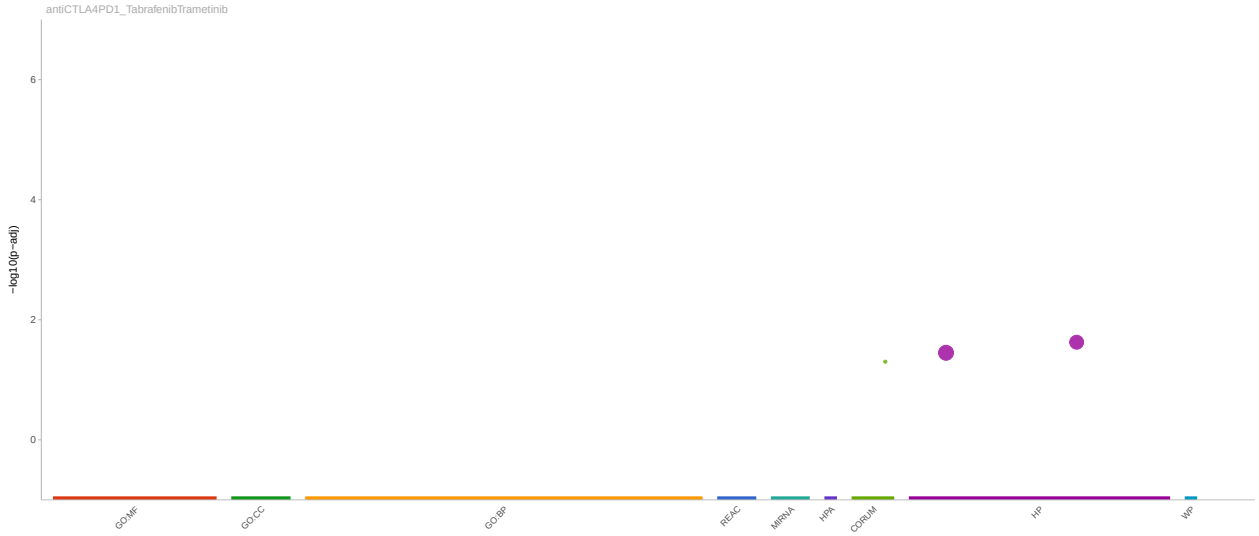
En esta sección analizo los DEGs que más se comparten entre los estudios para observar los mecanismos que pudieran haber en común entre los tratamientos.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("antiCTLA4PD1_TabrafenibTrametinib" = jVenn_Monocytes$`GSE91061|GSE6
),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##                              query significant    p_value term_size query_size
## 1 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.04993169         2          1
## 2 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.02367769         5          8
## 3 antiCTLA4PD1_TabrafenibTrametinib        TRUE 0.03548786         6          8
##   intersection_size precision
## 1                 1      1.00
## 2                 2      0.25
## 3                 2      0.25
```

```
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.

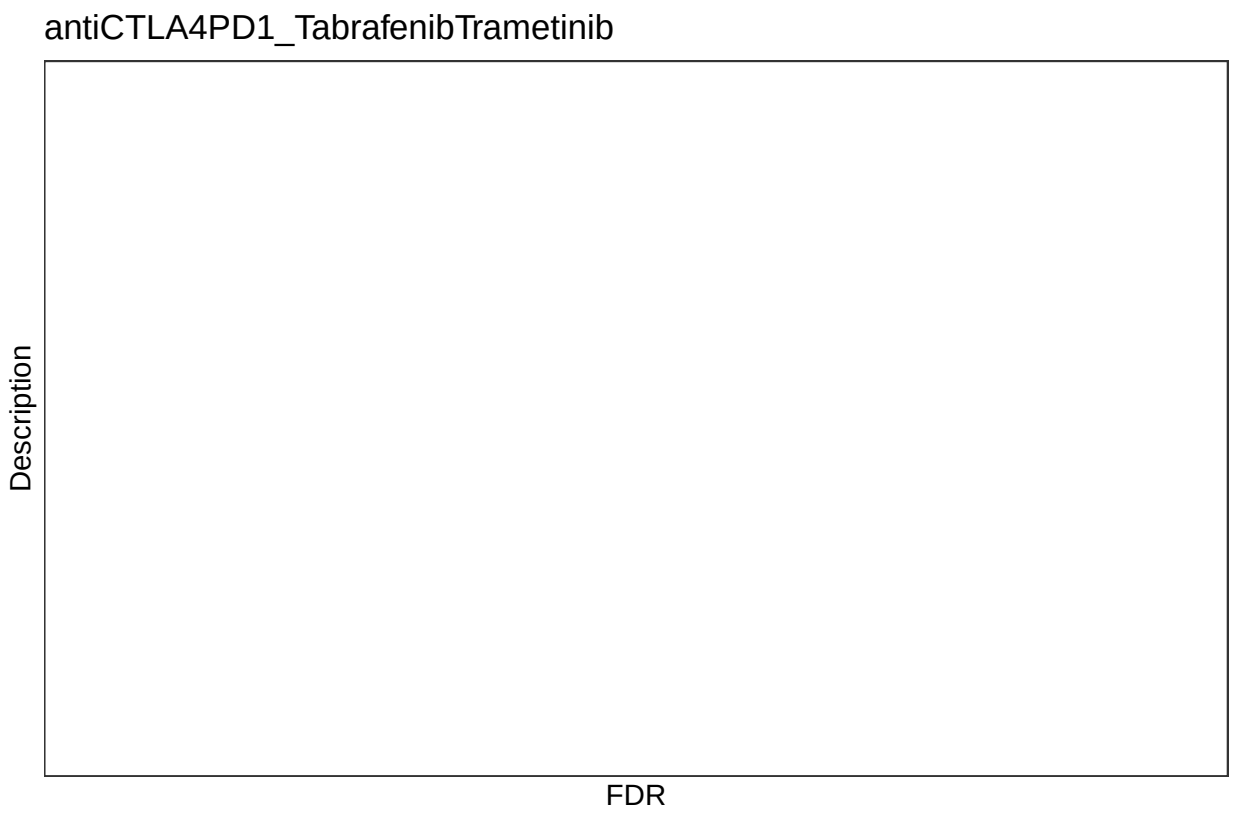
```
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

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_Monocytes_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")
}

pdf(file = paste0("./Treatment_comparisons/gProfiler_Monocytes_maxOverlap_geneset_antiCTLA4PD1_Tabrafeni
plot_gobps("antiCTLA4PD1_TabrafenibTrametinib")
dev.off()
```

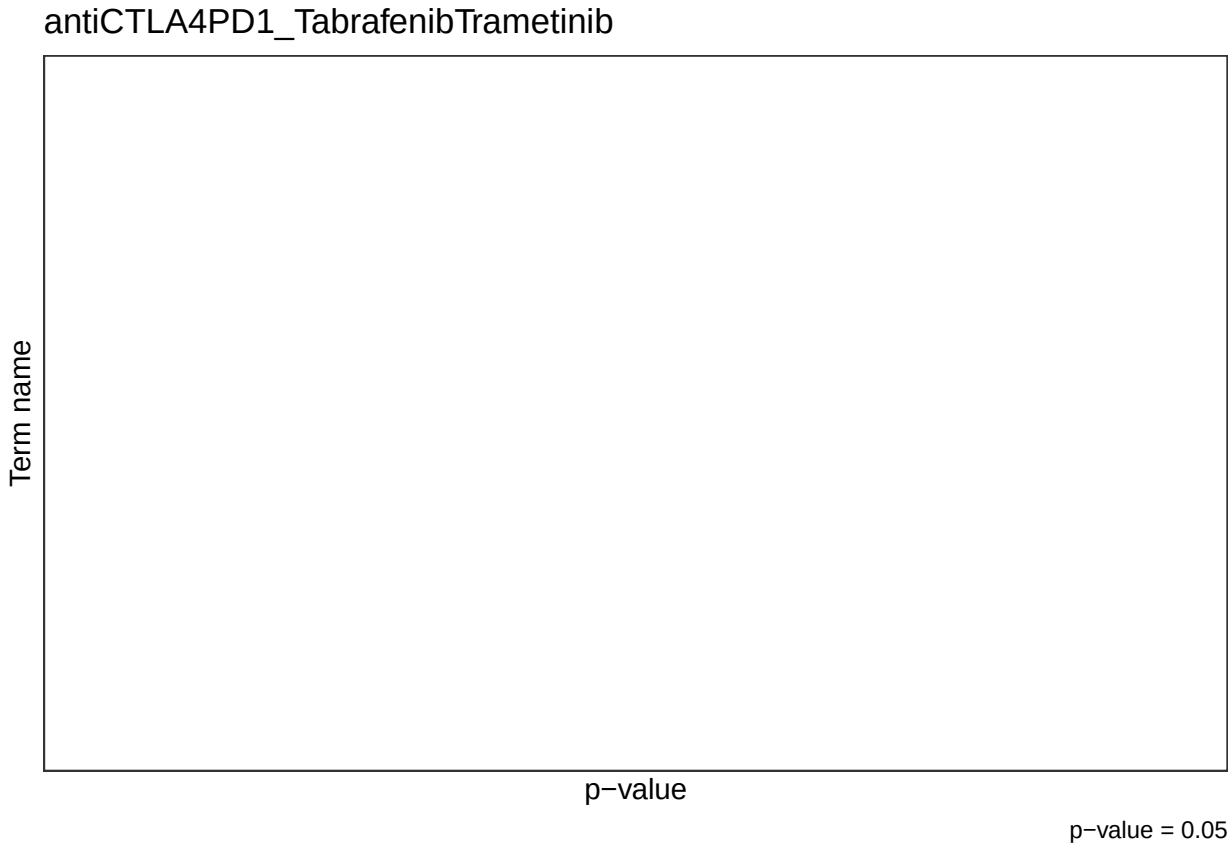
```
## pdf
##   2
```

```
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