Pathway Analysis - Macrophages M1
Input data: DEGs

Elena Eyre Sánchez, PhD

2025-01-04

Contents

1 Introducción y Objetivo **1**

2 Paquetes y datos **1**

3 Datos **1**

4 Uncovered **2**

5 Exclusive **9**

6 Máximo solapamiento **13**

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 Macrófagos tipo M1 obtenidos en jVenn.

```
setwd("~/Desktop/ELENA_UOC/TFM")
jVenn_MacrophagesM1_vs_GSE22155_02 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jV
head(jVenn_MacrophagesM1_vs_GSE22155_02) # sección de los resultados
```

```
## # A tibble: 6 x 3
##   GSE50509 GSE22155_02 `GSE50509|GSE22155_02`
##   <chr>    <chr>       <chr>
## 1 LIF      C5AR1       HAVCR2
## 2 MIR155HG CLEC7A      HCK
## 3 PTPRE    KCNJ15      RAB31
## 4 SOD2     OSM         CD163
## 5 SRSF10   APOBR       ITGAX
## 6 SH2B3    PILRA       C1orf162
```

```
jVenn_MacrophagesM1_vs_GSE22155_47 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jV
jVenn_MacrophagesM1_vs_tcgaskcm <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn
jVenn_MacrophagesM1_vs_GSE91061 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn
jVenn_MacrophagesM1_vs_GSE35640 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn
jVenn_MacrophagesM1_vs_GSE61992 <- read_csv("DEGs/Venn_diagrams_cell_type_treatments_vs_untreated/jVenn
setwd("~/Desktop/ELENA_UOC/TFM/GOBPs") # Directorio de trabajo
jVenn_MacrophagesM1 <- read_delim("~/Desktop/ELENA_UOC/TFM/DEGs/Venn_diagrams_cell_type_treatments_compa
    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:

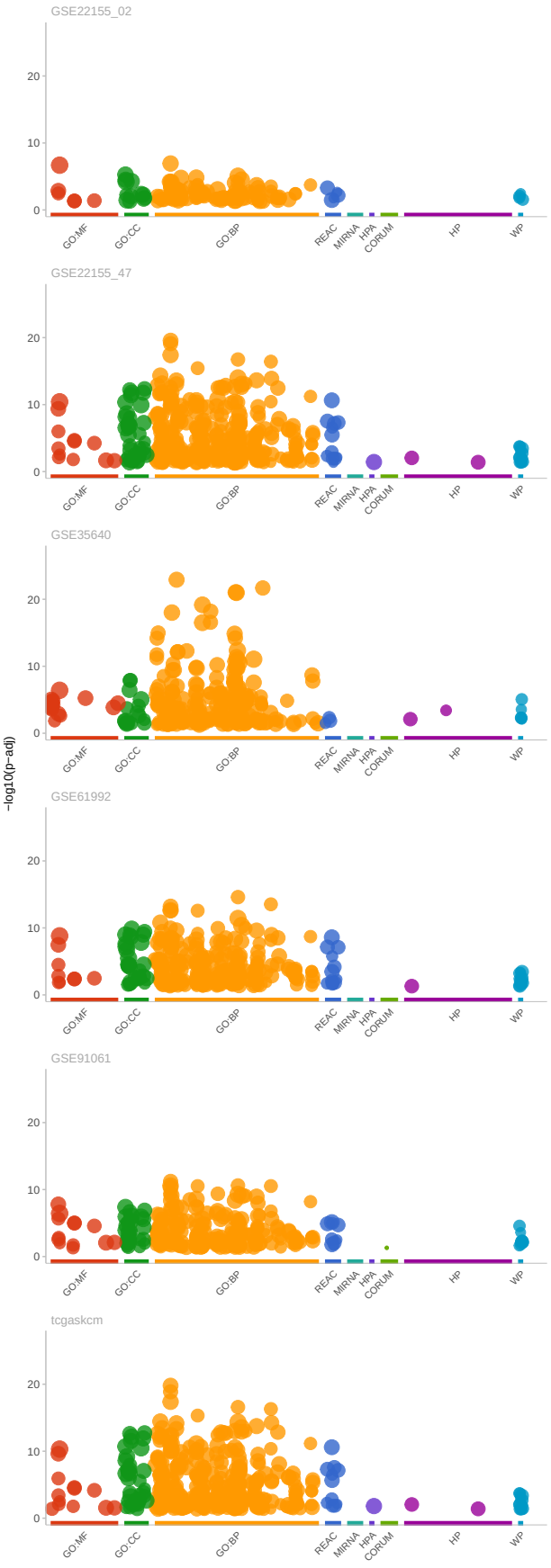
En esta sección contemplo la posibilidad que los DEGs únicos en el dataset sin tratar se traten de mecanismos que se escapan al tratamiento pero que igualmente se ven afectados por la enfermedad.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list("GSE22155_02" = unique(jVenn_MacrophagesM1_vs_GSE22155_02$GSE50509[wh
                                   "GSE22155_47" = jVenn_MacrophagesM1_vs_GSE22155_47$GSE50509[which(is.na(jV
                                   "GSE35640" = jVenn_MacrophagesM1_vs_GSE35640$GSE50509[which(is.na(jVenn_Ma
                                   "GSE91061" = jVenn_MacrophagesM1_vs_GSE91061$GSE50509[which(is.na(jVenn_Ma
                                   "GSE61992" = jVenn_MacrophagesM1_vs_GSE61992$GSE50509[which(is.na(jVenn_Ma
                                   "tcgaskcm" = jVenn_MacrophagesM1_vs_tcgaskcm$GSE50509[which(is.na(jVenn_Ma
                                   ),
                      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 1.164604e-07      3855        320              107
## 2 GSE22155_02        TRUE 8.748051e-06      8976        320              188
## 3 GSE22155_02        TRUE 1.455333e-05      3437        320               93
## 4 GSE22155_02        TRUE 1.589636e-05      3443        320               93
## 5 GSE22155_02        TRUE 3.189032e-05      7394        320              161
## 6 GSE22155_02        TRUE 6.050687e-05      1791        320               58
##   precision
## 1  0.334375
## 2  0.587500
## 3  0.290625
## 4  0.290625
## 5  0.503125
## 6  0.181250
```

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


GSE22155_02
GSE22155_47
GSE35640
GSE61992
GSE91061
tcgaskcm
−log10(p−adj)
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_MacrophagesM
                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
##         122         424         295         303         265         422
```

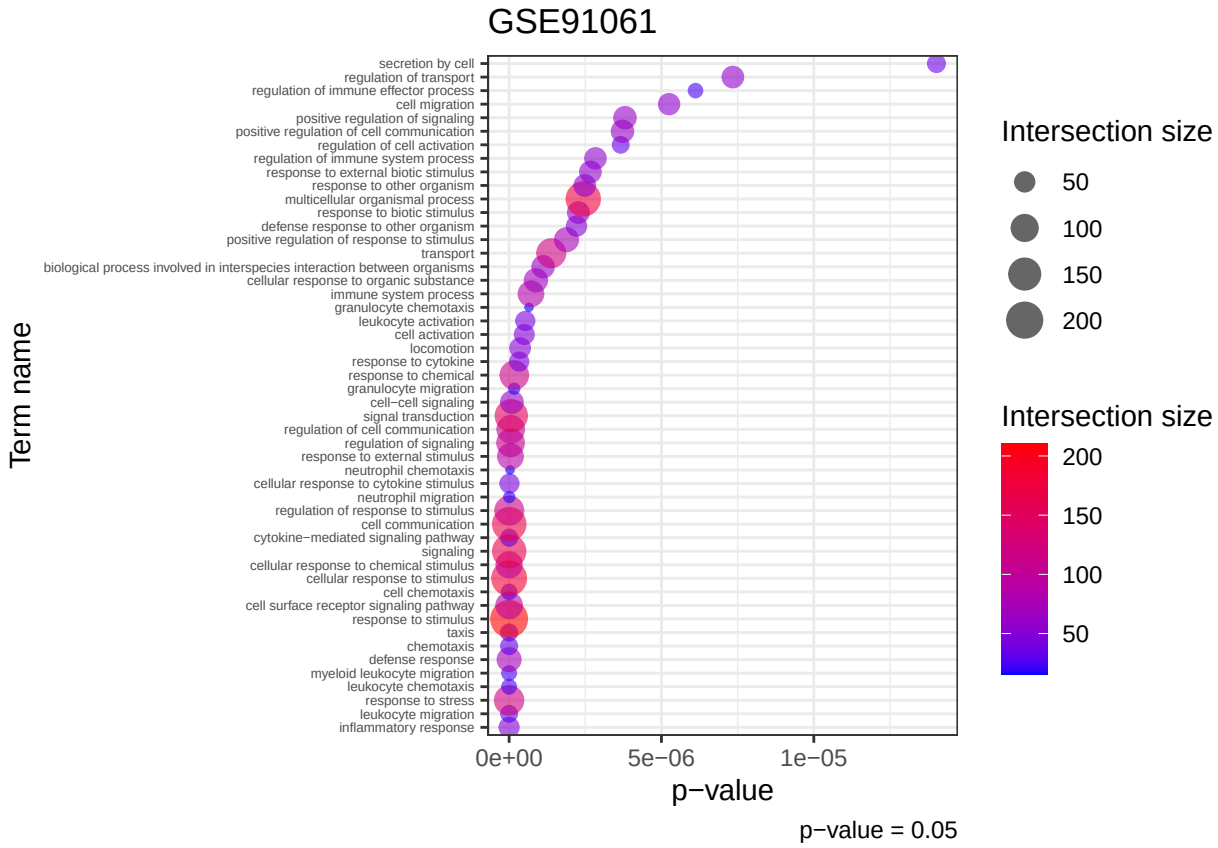
```
prop.table(table(df$query))*100
```

```
##
## GSE22155_02 GSE22155_47    GSE35640    GSE61992    GSE91061    tcgaskcm
##    6.663026   23.156745   16.111415   16.548334   14.472966   23.047515
```

```
#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/MacrM1_GOBPs.txt", sep = "\t"
write.table(df2, file = "./Venn_diagrams_cell_type_treatments_vs_untreated/MacrM1_GOBPs_freq.txt", sep =
#rm(jVenn_MacrophagesM1_vs_GSE22155_02, jVenn_MacrophagesM1_vs_GSE22155_47, jVenn_MacrophagesM1_vs_GSE3
```

Barplot of the top GO-BPs:

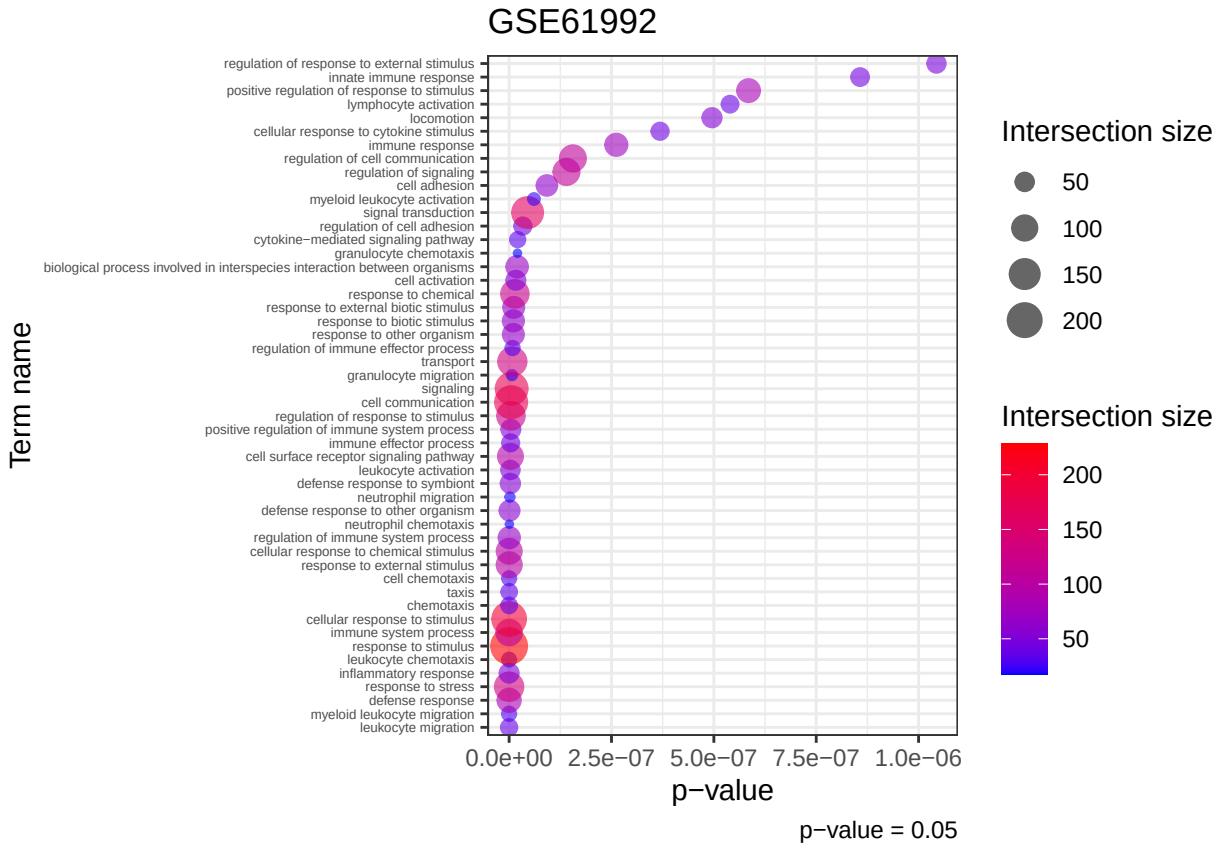
```
plot_gobps <- function(study, n = 50){
df2 <- df[df$source == "GO:BP",]
df2_GSE91061 <- df2[df2$query == study,]
df2_GSE91061 <- df2_GSE91061[1:n,]
ggplot(df2_GSE91061, aes(p_value, reorder(term_name,p_value))) + ggtitle(study) +
  geom_point(aes(p_value, reorder(term_name,p_value), colour = intersection_size,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

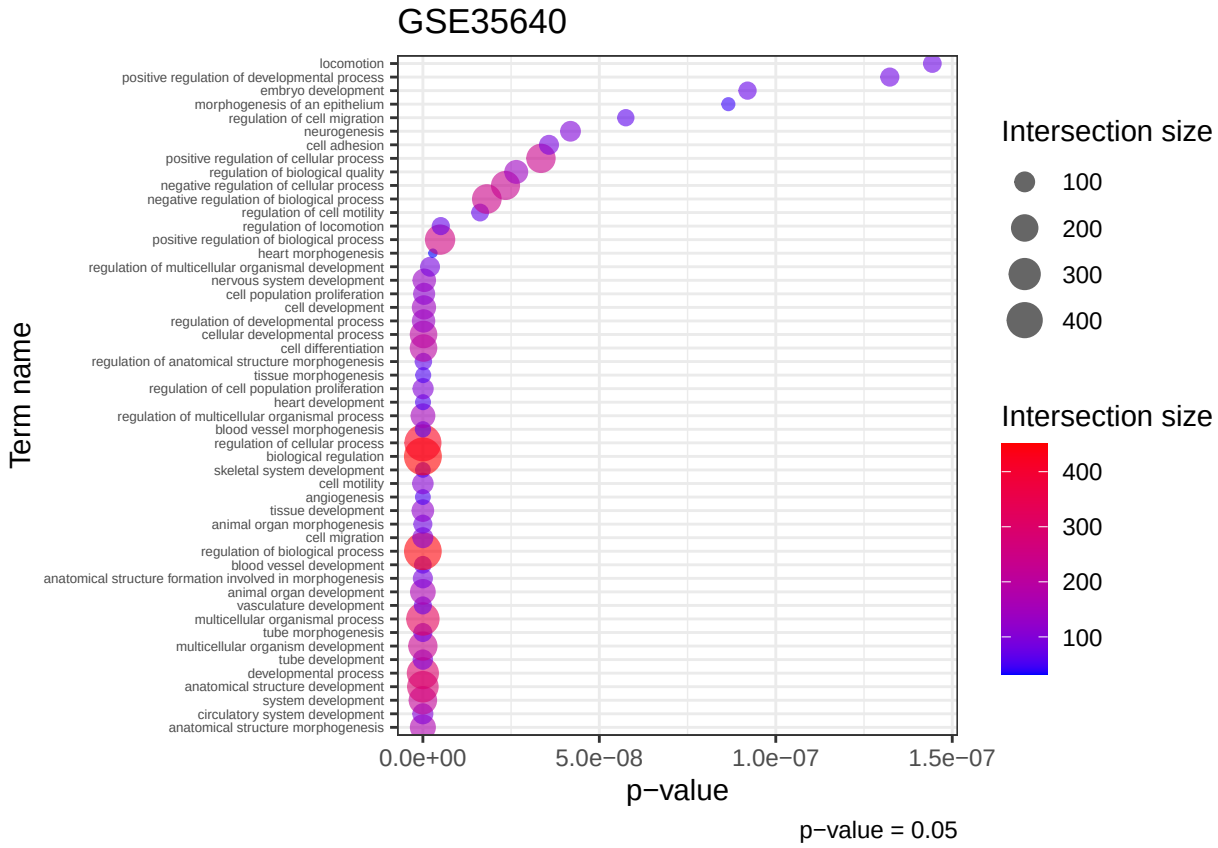
```
plot_gobps(study = "GSE61992")
```




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

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

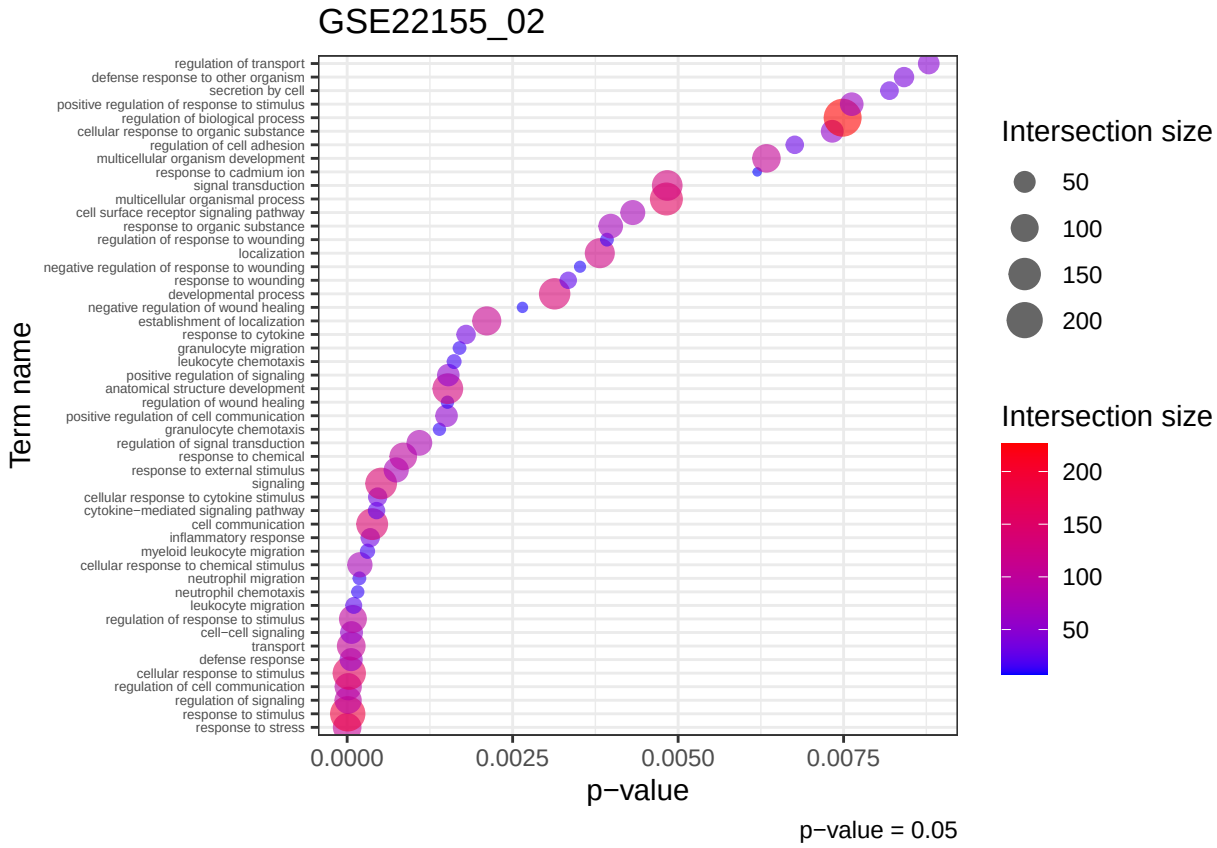
```
plot_gobps(study = "GSE35640")
```




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

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

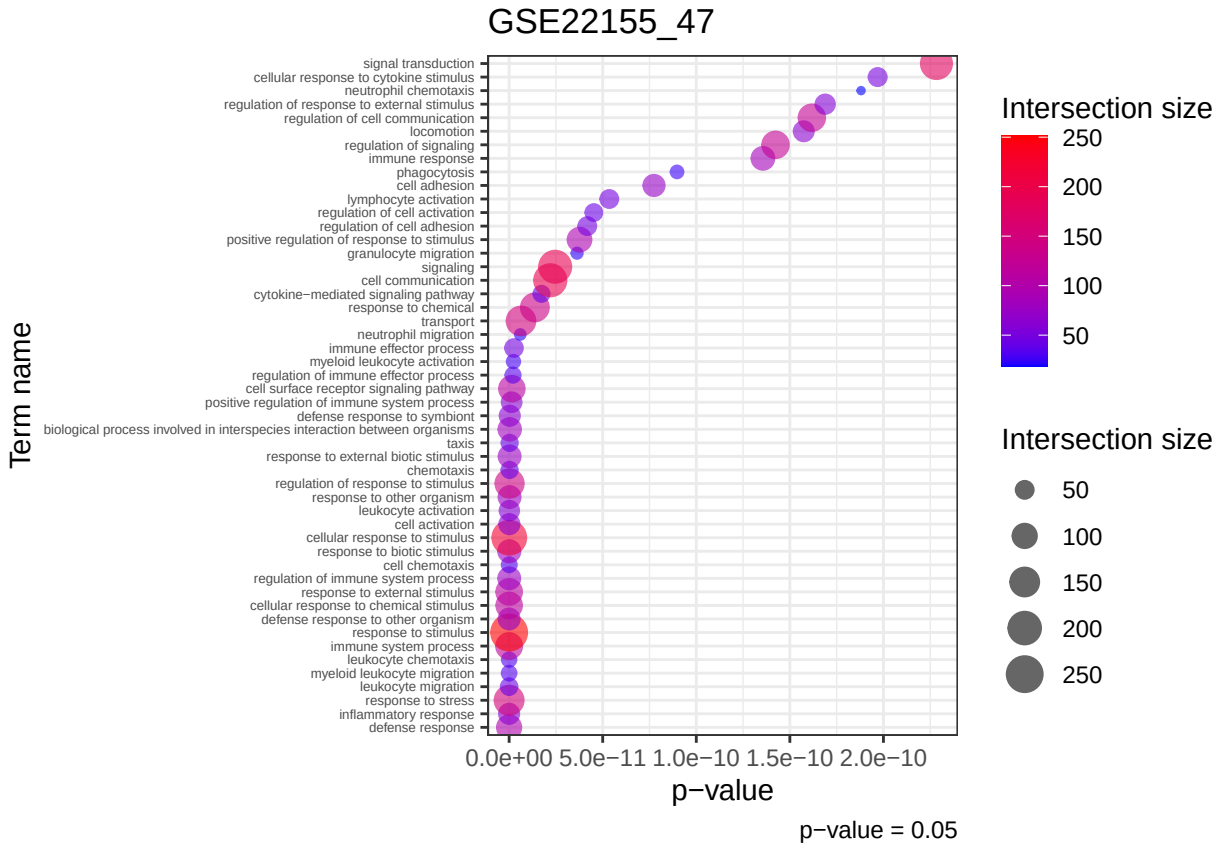
```
plot_gobps(study = "GSE22155_02")
```




```
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_MacrophagesM1$GSE61992[which(is.na(jVenn_Macropha
                                   "GSE91061" = jVenn_MacrophagesM1$GSE91061[which(is.na(jVenn_MacrophagesM1$
                                   "tcgaskcm" = jVenn_MacrophagesM1$`TCGA-SKCM`[which(is.na(jVenn_MacrophagesM
                                   ),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

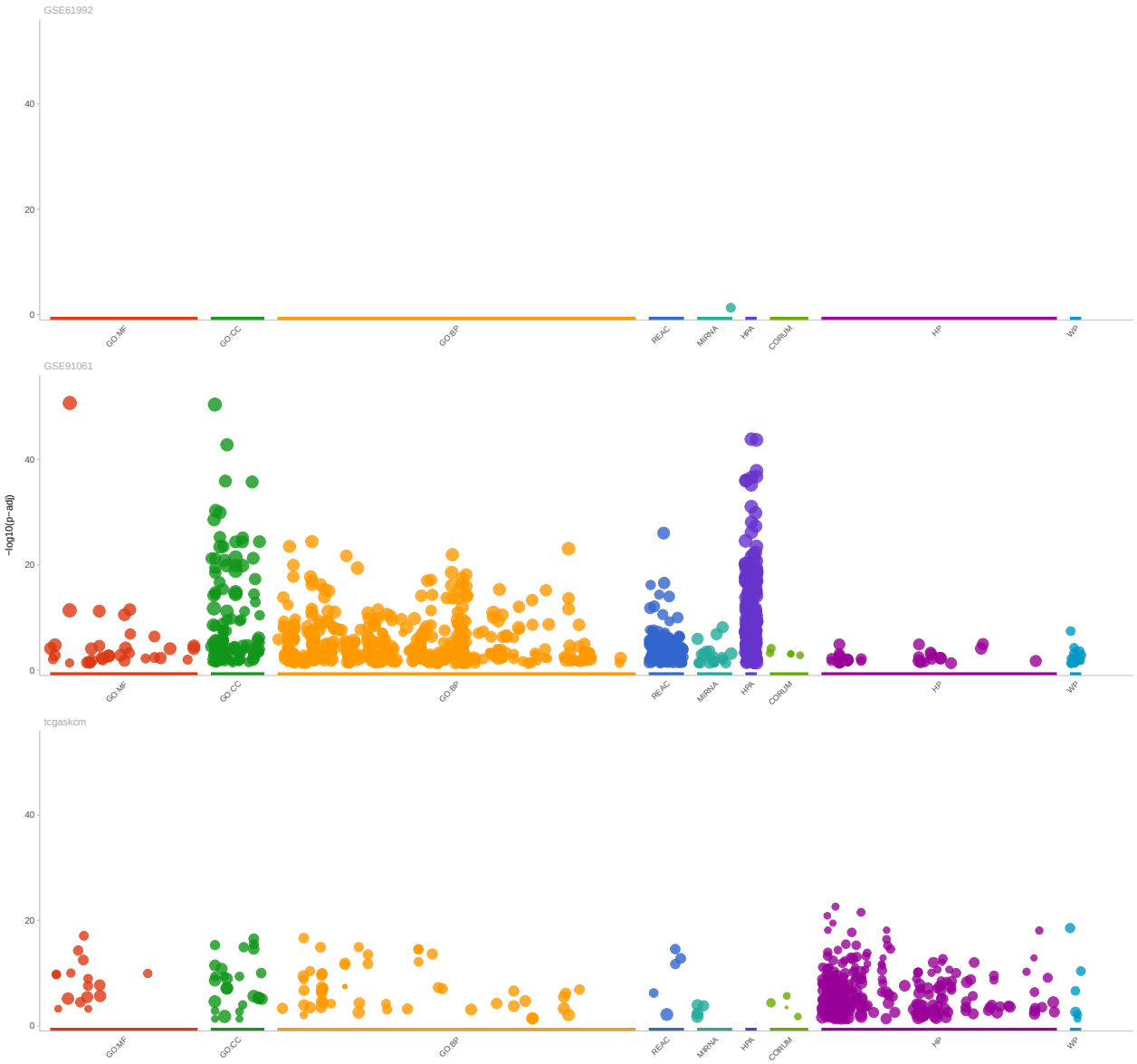
```
##      query significant      p_value term_size query_size intersection_size
## 1 GSE61992        TRUE 4.773516e-02        69         20                3
## 2 GSE91061        TRUE 7.384767e-05        33        320               14
## 3 GSE91061        TRUE 5.343907e-04        20        320               10
## 4 GSE91061        TRUE 8.141399e-04        10        320                7
## 5 GSE91061        TRUE 8.141399e-04        10        320                7
## 6 GSE91061        TRUE 1.456767e-03        14        320                8
##   precision
## 1  0.150000
## 2  0.043750
## 3  0.031250
## 4  0.021875
```

```
## 5  0.021875
## 6  0.025000
```

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

```
# format to GEM
# 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_MacrophagesM1_Only_geneset_", unique(.y
                sep = "\t", quote = F, row.names = F))
```

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



Proporciones de los pathways para cada estudio:

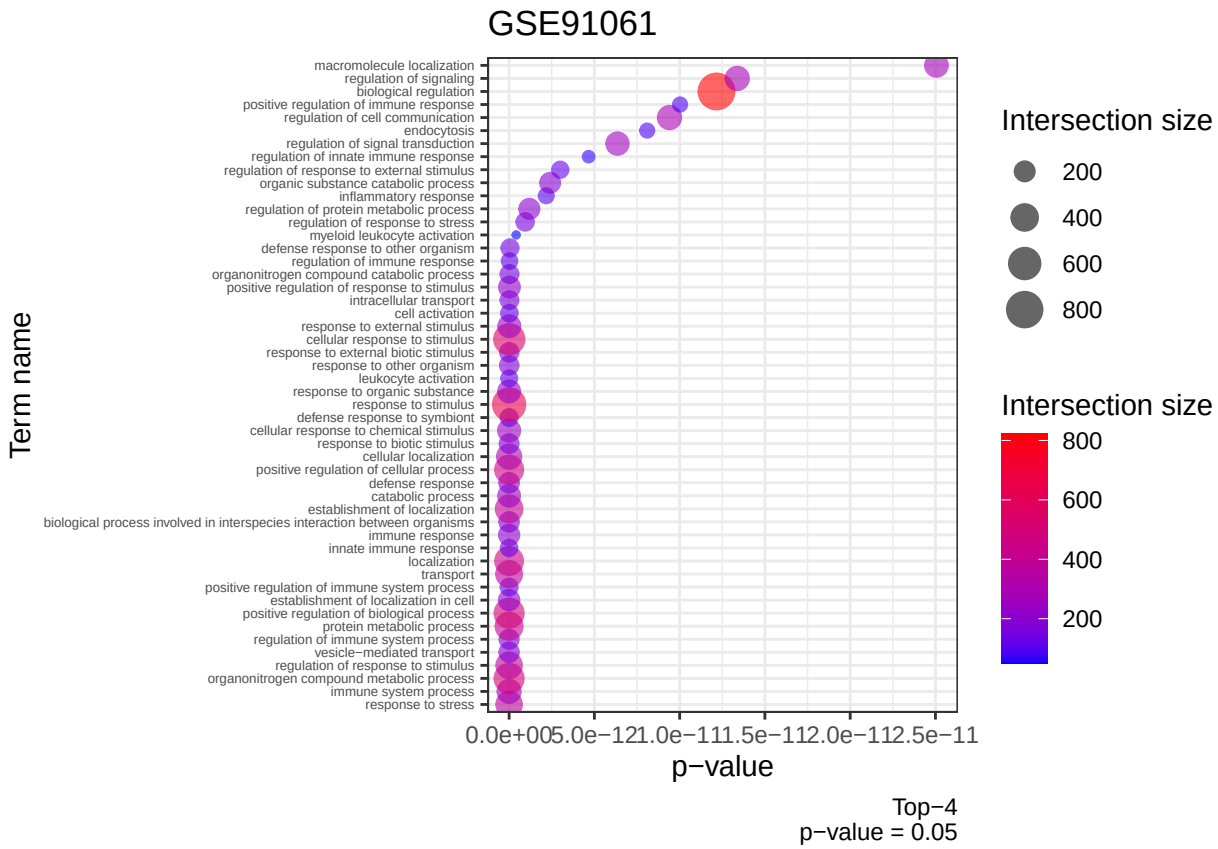
```
df <- as.data.frame(multi_gostres$result[,1:13])
table(df$query)
```

```
##
## GSE61992 GSE91061 tcgaskcm
##        1     1039      371
```

```
write.table(gem, file = "./Treatment_comparisons/gProfiler_MacrophagesM1_Only_genesets.txt", sep = "\t"
```

```
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")
```

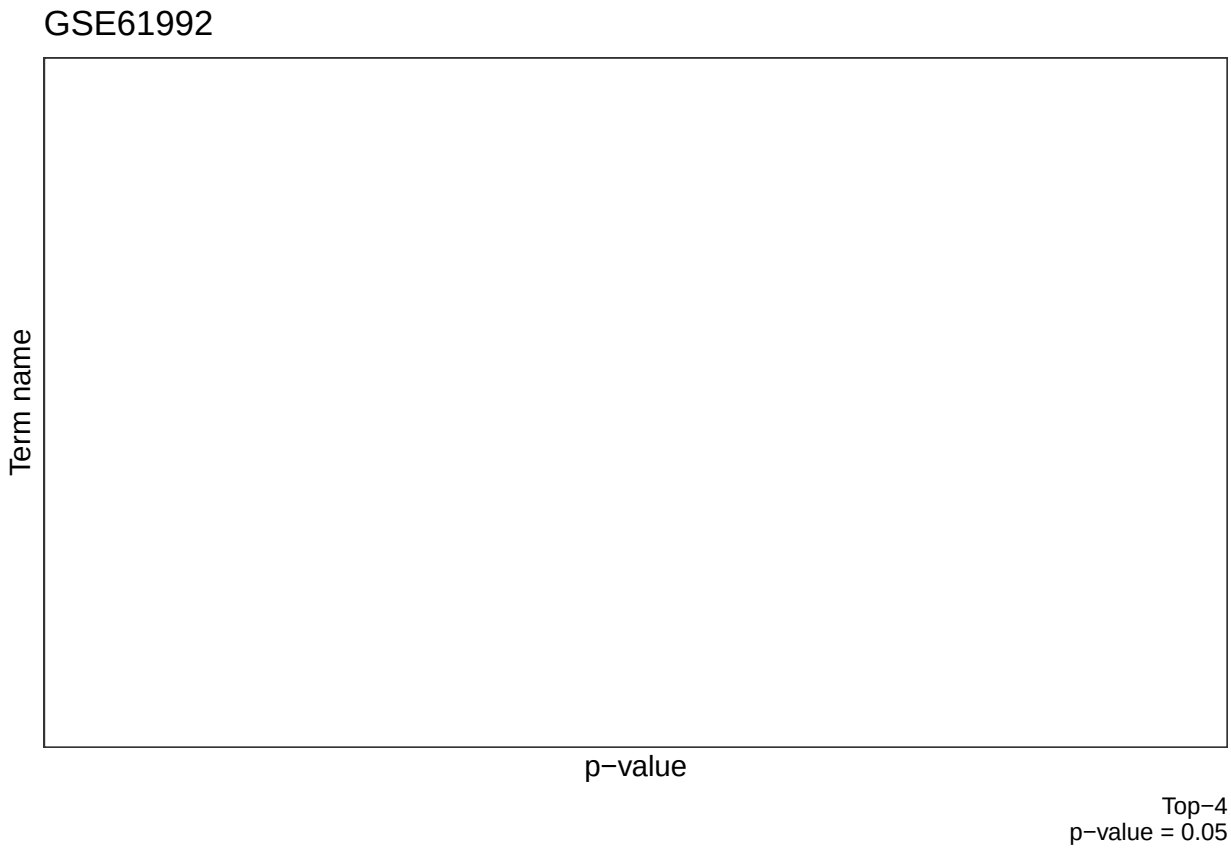
```
}
df2 <- df[df$source == "GO:BP",]
df2 <- df2[1:50,]
plot_gobps(study = "GSE91061")
```




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

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

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



6 Máximo solapamiento

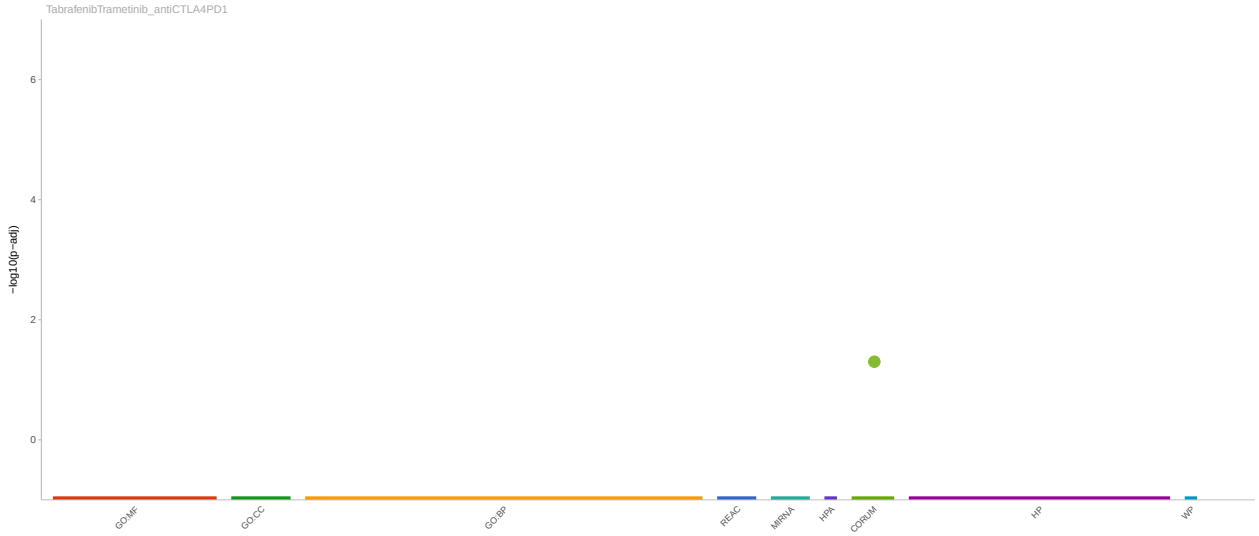
En esta sección analizo los DEGs que más se comparten entre los estudios para observar los mecanismos que pudieran haber en común entre los tratamientos.

Primero realizo el pathway multianálisis:

```
multi_gostres <- gost(query = list(
  "TabrafenibTrametinib_antiCTLA4PD1" = jVenn_MacrophagesM1$`GSE91061|GSE61992`[which(is.na(jVenn_Macro
),
                      evcodes = TRUE, multi_query = FALSE,
                      sources = c("GO", "REAC", "MIRNA", "CORUM", "HP", "HPA", "WP"))
head(multi_gostres$result[,1:7])
```

```
##                               query significant    p_value term_size query_size
## 1 TabrafenibTrametinib_antiCTLA4PD1        TRUE 0.04998799         1          3
##   intersection_size precision
## 1                 1 0.3333333
```

```
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.

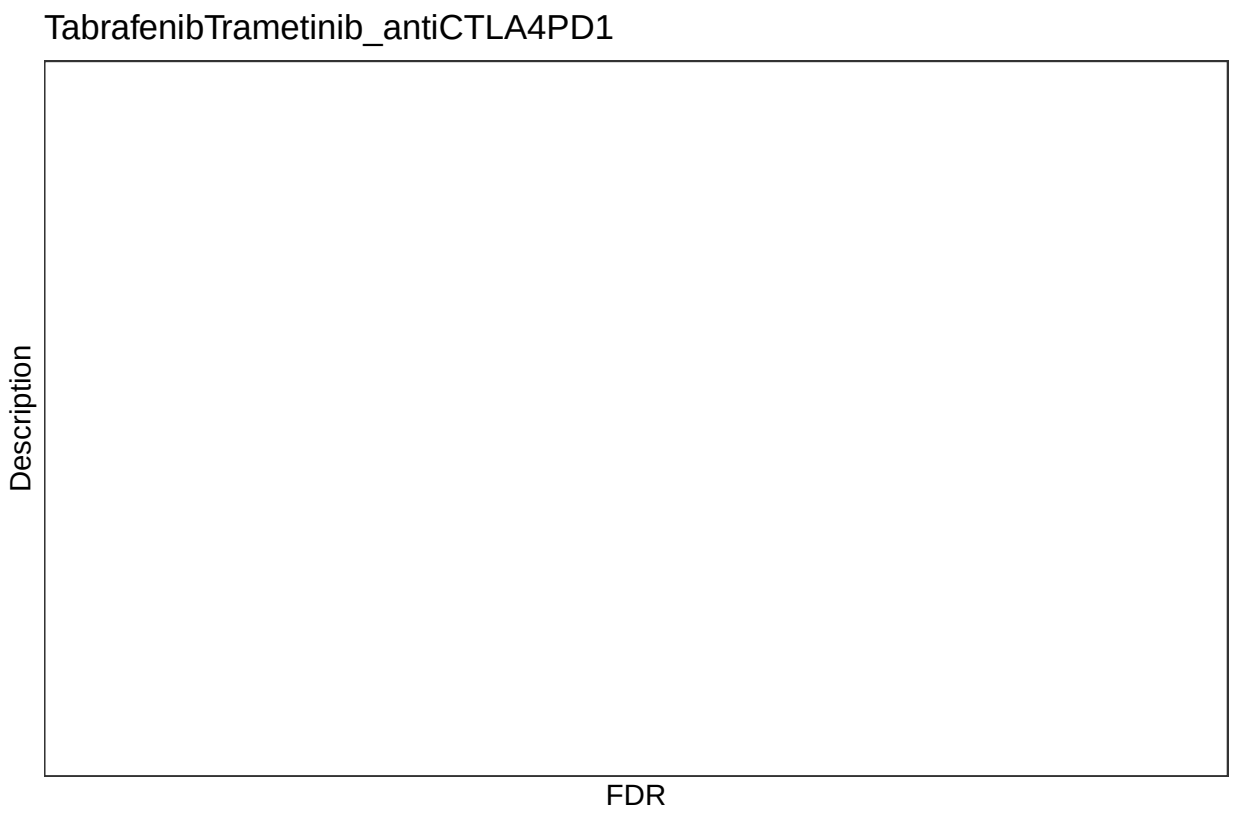
```
# format to GEM
# format to GEM
gem <- multi_gostres$result[,c("query", "term_id", "term_name", "p_value", "intersection_size")]
colnames(gem) <- c("query", "GO.ID", "Description", "p.Val", "Genes")
gem$FDR <- gem$p.Val
gem$Phenotype = "+1"

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_MacrophagesM1_maxOverlap_geneset_", un
                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.

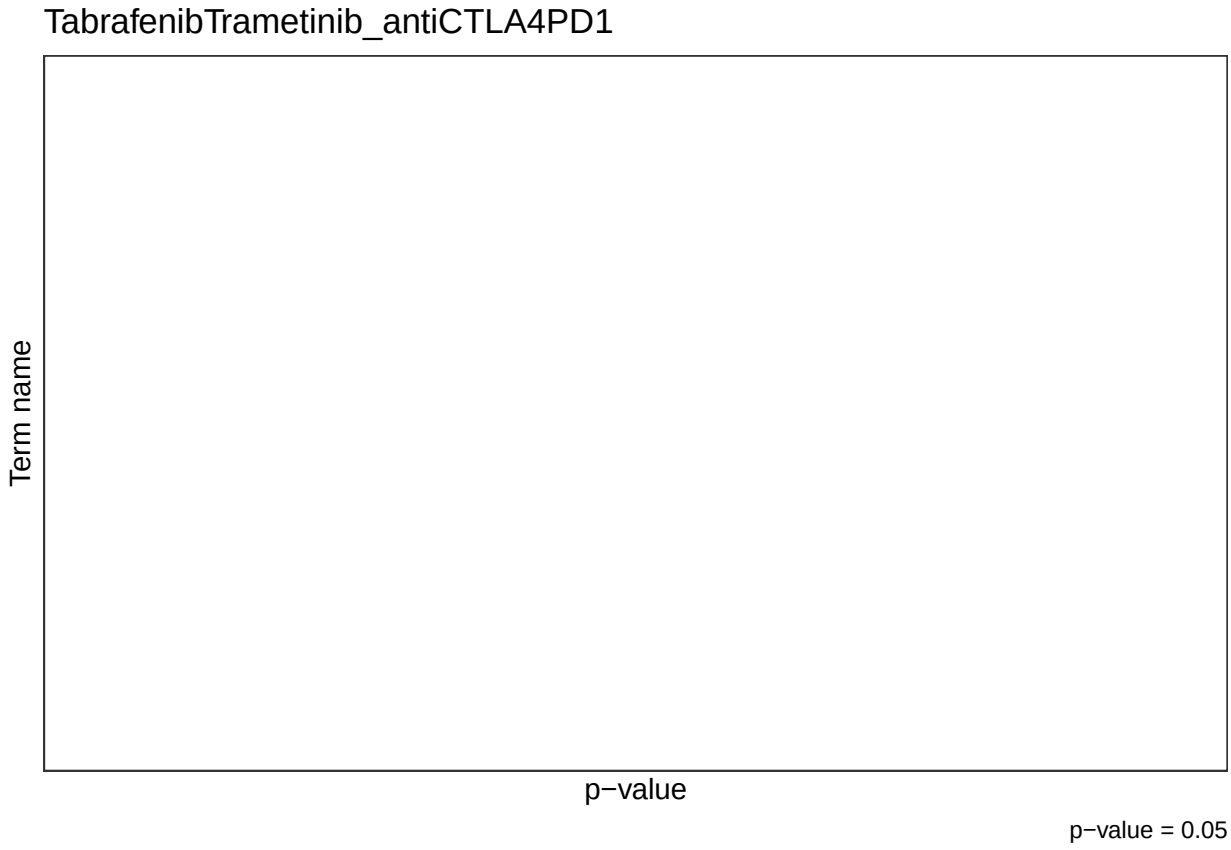
```
#unique(gem$query)

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_MacrophagesM1_maxOverlap_geneset_TabrafenibTrametin
plot_gobps("TabrafenibTrametinib_antiCTLA4PD1")
dev.off()
```

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

```
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