

# elt2\_target\_gene\_ontology

Robert Williams

4/11/2022

```
library("topGO")
```

```
## Warning: package 'S4Vectors' was built under R version 4.1.1
```

```
library(tidyverse)
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

## Gene ontology analysis

```
source("../04_promoters/02_scripts/GOfxns.R")
```

```
WORMGO_nonObsolete <- read_tsv(file = "../06_intestine_enriched_genes/01_input/Celegans_G0terms_NonO
```

```
## Rows: 130120 Columns: 5
```

```
## -- Column specification -----
```

```
## Delimiter: "\t"
```

```
## chr (5): Gene.symbol, Gene.goAnnotation.qualifier, Gene.goAnnotation.ontolog...
```

```
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
WORMGO <- WORMGO_nonObsolete %>% dplyr::select(wbps_gene_id = "Gene.primaryIdentifier",  
  external_gene_id = "Gene.symbol",  
  go_accession = "Gene.goAnnotation.ontologyTerm.identifier"  
)
```

```
fisherGOplot <- function(in.df){
```

```
in.df$BP.result %>% mutate(GO = "BP") %>% bind_rows(in.df$MF.result %>% mutate(GO = "MF")) %>% bind_rows
```

```
  filter(Significant > Expected) %>%
```

```
  mutate(gene_count = paste0(Significant, "/", round(Expected))) %>%
```

```
  slice_min(fisher, n = 10) %>%
```

```
  mutate(Term = fct_rev(fct_reorder(Term, fisher))) %>%
```

```
ggplot(aes(x = Term, y = -log10(fisher), fill = GO, label = gene_count)) +
```

```
  geom_bar(stat = "identity", color = "black") +
```

```
  geom_text(hjust = -0.05, aes(y = 0))+
```

```
  scale_fill_brewer(palette = "Greys", direction = -1) +
```

```
  scale_x_discrete(labels = function(x) str_wrap(x, width = 25)) +
```

```
  coord_flip() +
```

```

theme_bw() +
theme(axis.text.x=element_text(colour="black"),
      axis.text.y=element_text(colour="black")) +
ggtitle(paste("data:", deparse(substitute(in.df)), sep = " "))
}

```

## What terms are ELT-2 activated target genes enriched for regardless of developmental stage or intestine expression?

```

all_stages_chip <- read_csv(file = "../05_elt2_target_analysis/03_output/all_stages_chip.csv")

## Rows: 59985 Columns: 5
## -- Column specification -----
## Delimiter: ","
## chr (5): WBGeneID, promoter_status, intestine_expression, elt2_ko, stage
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
all_activated_targets_GO <- runGO(foreground_genes = all_stages_chip %>% filter(elt2_ko == "activated",
                                   background_genes = all_stages_chip %>% pull(WBGeneID) %>% unique(),
                                   WORMGO = WORMGO,
                                   topNodes = 20)

##
## Building most specific GOs .....
## ( 3830 GO terms found. )
##
## Build GO DAG topology .....
## ( 6395 GO terms and 14056 relations. )
##
## Annotating nodes .....
## ( 9654 genes annotated to the GO terms. )
##
## Building most specific GOs .....
## ( 2228 GO terms found. )
##
## Build GO DAG topology .....
## ( 2711 GO terms and 3506 relations. )
##
## Annotating nodes .....
## ( 9770 genes annotated to the GO terms. )
##
## Building most specific GOs .....
## ( 961 GO terms found. )
##
## Build GO DAG topology .....
## ( 1194 GO terms and 2089 relations. )
##
## Annotating nodes .....
## ( 12282 genes annotated to the GO terms. )
##
## -- Classic Algorithm --

```

```

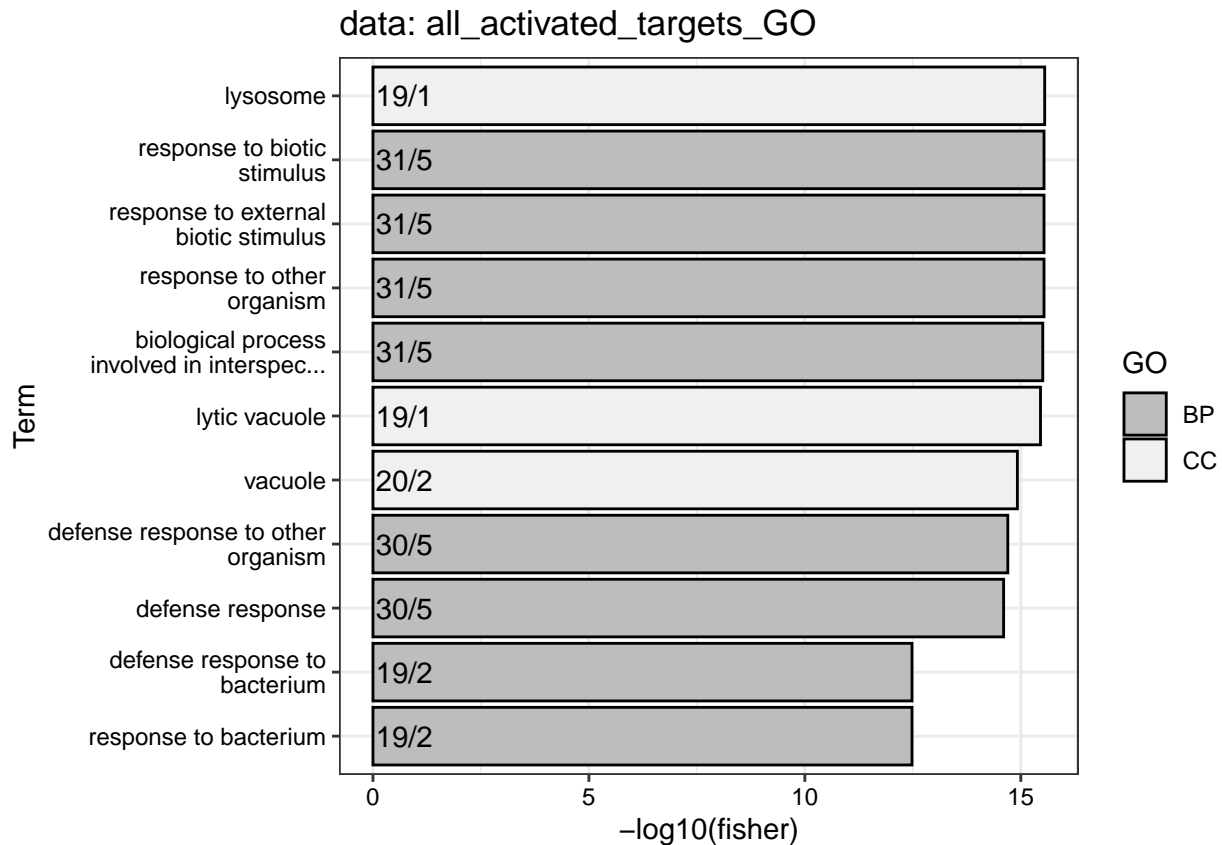
##
## the algorithm is scoring 645 nontrivial nodes
## parameters:
## test statistic: fisher
##
## -- Weight01 Algorithm --
##
## the algorithm is scoring 645 nontrivial nodes
## parameters:
## test statistic: fisher
##
## Level 14: 2 nodes to be scored (0 eliminated genes)
##
## Level 13: 5 nodes to be scored (0 eliminated genes)
##
## Level 12: 10 nodes to be scored (7 eliminated genes)
##
## Level 11: 16 nodes to be scored (300 eliminated genes)
##
## Level 10: 31 nodes to be scored (905 eliminated genes)
##
## Level 9: 54 nodes to be scored (1391 eliminated genes)
##
## Level 8: 63 nodes to be scored (1956 eliminated genes)
##
## Level 7: 95 nodes to be scored (2608 eliminated genes)
##
## Level 6: 120 nodes to be scored (4017 eliminated genes)
##
## Level 5: 116 nodes to be scored (6114 eliminated genes)
##
## Level 4: 77 nodes to be scored (8056 eliminated genes)
##
## Level 3: 41 nodes to be scored (8917 eliminated genes)
##
## Level 2: 14 nodes to be scored (9233 eliminated genes)
##
## Level 1: 1 nodes to be scored (9548 eliminated genes)
##
## -- Classic Algorithm --
##
## the algorithm is scoring 253 nontrivial nodes
## parameters:
## test statistic: fisher
##
## -- Weight01 Algorithm --
##
## the algorithm is scoring 253 nontrivial nodes
## parameters:
## test statistic: fisher
##
## Level 11: 1 nodes to be scored (0 eliminated genes)
##
## Level 10: 5 nodes to be scored (0 eliminated genes)

```

```

##
## Level 9: 17 nodes to be scored (1 eliminated genes)
##
## Level 8: 16 nodes to be scored (408 eliminated genes)
##
## Level 7: 31 nodes to be scored (1708 eliminated genes)
##
## Level 6: 50 nodes to be scored (2044 eliminated genes)
##
## Level 5: 53 nodes to be scored (2857 eliminated genes)
##
## Level 4: 45 nodes to be scored (3960 eliminated genes)
##
## Level 3: 25 nodes to be scored (6059 eliminated genes)
##
## Level 2: 9 nodes to be scored (7435 eliminated genes)
##
## Level 1: 1 nodes to be scored (9262 eliminated genes)
##
## -- Classic Algorithm --
##
## the algorithm is scoring 84 nontrivial nodes
## parameters:
## test statistic: fisher
##
## -- Weight01 Algorithm --
##
## the algorithm is scoring 84 nontrivial nodes
## parameters:
## test statistic: fisher
##
## Level 10: 1 nodes to be scored (0 eliminated genes)
##
## Level 9: 4 nodes to be scored (0 eliminated genes)
##
## Level 8: 10 nodes to be scored (5 eliminated genes)
##
## Level 7: 12 nodes to be scored (97 eliminated genes)
##
## Level 6: 17 nodes to be scored (541 eliminated genes)
##
## Level 5: 13 nodes to be scored (1003 eliminated genes)
##
## Level 4: 12 nodes to be scored (5381 eliminated genes)
##
## Level 3: 12 nodes to be scored (10219 eliminated genes)
##
## Level 2: 2 nodes to be scored (11695 eliminated genes)
##
## Level 1: 1 nodes to be scored (12136 eliminated genes)
all_activated_targets_G0_plot <- fisherG0plot(all_activated_targets_G0)
all_activated_targets_G0_plot

```



```
ggsave(filename = "../03_output/GO_plots/all_activated_targets_GO_plot.pdf", plot = all_activated_targets_GO)
```

What terms are ELT-2 target repressed genes enriched for regardless of developmental stage or intestine expression?

```
all_repressed_targets_GO <- runGO(foreground_genes = all_stages_chip %>% filter(elt2_ko == "repressed",
  background_genes = all_stages_chip %>% pull(WBGeneID) %>% unique(),
  WORMGO = WORMGO,
  topNodes = 20)
```

```
##
## Building most specific GOs .....
## ( 3830 GO terms found. )
##
## Build GO DAG topology .....
## ( 6395 GO terms and 14056 relations. )
##
## Annotating nodes .....
## ( 9654 genes annotated to the GO terms. )
##
## Building most specific GOs .....
## ( 2228 GO terms found. )
```

```

##
## Build GO DAG topology .....
## ( 2711 GO terms and 3506 relations. )
##
## Annotating nodes .....
## ( 9770 genes annotated to the GO terms. )
##
## Building most specific GOs .....
## ( 961 GO terms found. )
##
## Build GO DAG topology .....
## ( 1194 GO terms and 2089 relations. )
##
## Annotating nodes .....
## ( 12282 genes annotated to the GO terms. )
##
##      -- Classic Algorithm --
##
##      the algorithm is scoring 1224 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
##      -- Weight01 Algorithm --
##
##      the algorithm is scoring 1224 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
## Level 17:  2 nodes to be scored    (0 eliminated genes)
##
## Level 16:  7 nodes to be scored    (0 eliminated genes)
##
## Level 15:  9 nodes to be scored    (7 eliminated genes)
##
## Level 14:  9 nodes to be scored    (40 eliminated genes)
##
## Level 13: 18 nodes to be scored    (96 eliminated genes)
##
## Level 12: 29 nodes to be scored    (226 eliminated genes)
##
## Level 11: 48 nodes to be scored    (581 eliminated genes)
##
## Level 10: 84 nodes to be scored    (1211 eliminated genes)

```

```

##
## Level 9: 142 nodes to be scored (2010 eliminated genes)
##
## Level 8: 146 nodes to be scored (3184 eliminated genes)
##
## Level 7: 167 nodes to be scored (4485 eliminated genes)
##
## Level 6: 194 nodes to be scored (5870 eliminated genes)
##
## Level 5: 186 nodes to be scored (6829 eliminated genes)
##
## Level 4: 108 nodes to be scored (8487 eliminated genes)
##
## Level 3: 58 nodes to be scored (9058 eliminated genes)
##
## Level 2: 16 nodes to be scored (9431 eliminated genes)
##
## Level 1: 1 nodes to be scored (9616 eliminated genes)
##
##      -- Classic Algorithm --
##
##      the algorithm is scoring 385 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
##      -- Weight01 Algorithm --
##
##      the algorithm is scoring 385 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
## Level 12: 1 nodes to be scored (0 eliminated genes)
##
## Level 11: 4 nodes to be scored (0 eliminated genes)
##
## Level 10: 11 nodes to be scored (24 eliminated genes)
##
## Level 9: 25 nodes to be scored (49 eliminated genes)
##
## Level 8: 37 nodes to be scored (498 eliminated genes)
##
## Level 7: 50 nodes to be scored (1809 eliminated genes)
##
## Level 6: 75 nodes to be scored (2153 eliminated genes)

```

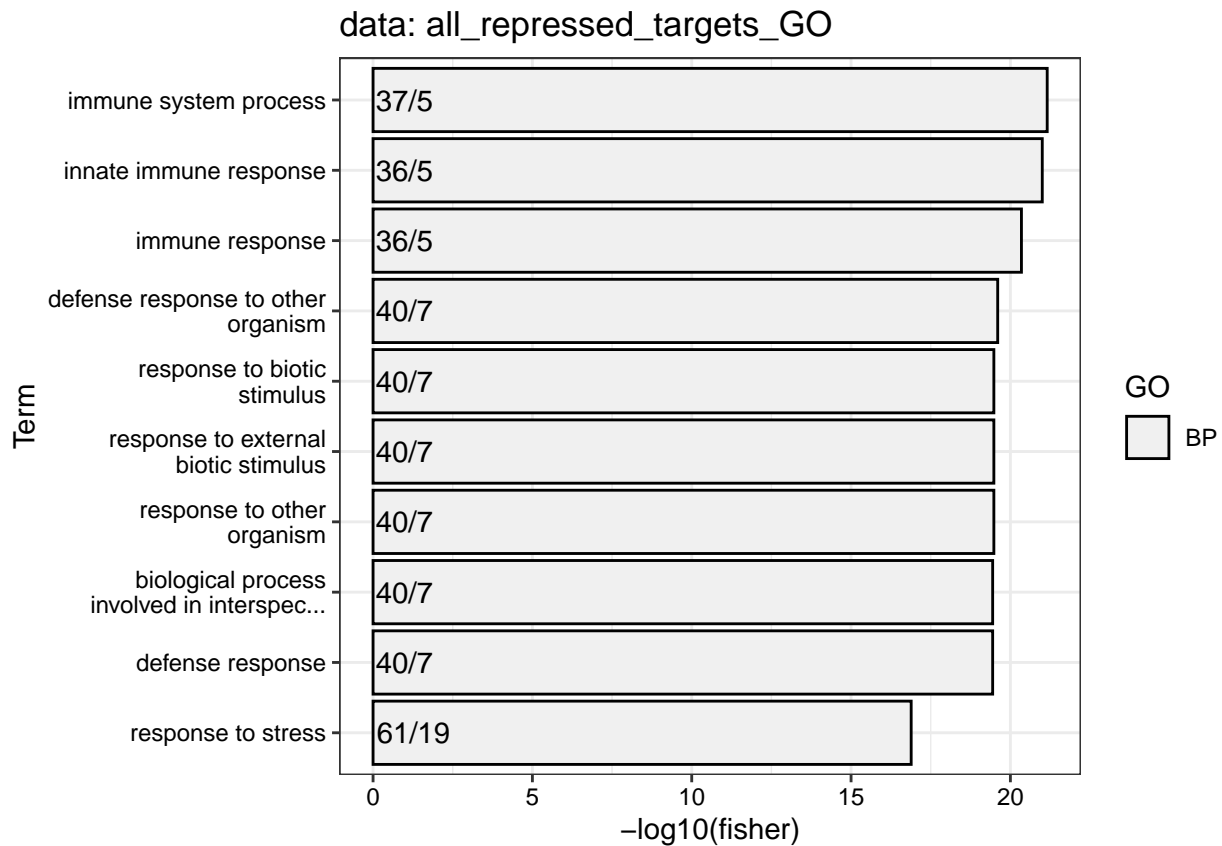
```

##
## Level 5: 75 nodes to be scored (3095 eliminated genes)
##
## Level 4: 65 nodes to be scored (4545 eliminated genes)
##
## Level 3: 31 nodes to be scored (6962 eliminated genes)
##
## Level 2: 10 nodes to be scored (7816 eliminated genes)
##
## Level 1: 1 nodes to be scored (9317 eliminated genes)
##
##      -- Classic Algorithm --
##
##      the algorithm is scoring 201 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
##      -- Weight01 Algorithm --
##
##      the algorithm is scoring 201 nontrivial nodes
##      parameters:
##          test statistic: fisher
##
## Level 11: 5 nodes to be scored (0 eliminated genes)
##
## Level 10: 12 nodes to be scored (0 eliminated genes)
##
## Level 9: 19 nodes to be scored (59 eliminated genes)
##
## Level 8: 29 nodes to be scored (315 eliminated genes)
##
## Level 7: 26 nodes to be scored (633 eliminated genes)
##
## Level 6: 27 nodes to be scored (1940 eliminated genes)
##
## Level 5: 33 nodes to be scored (2849 eliminated genes)
##
## Level 4: 25 nodes to be scored (6100 eliminated genes)
##
## Level 3: 22 nodes to be scored (11001 eliminated genes)
##
## Level 2: 2 nodes to be scored (11819 eliminated genes)
##
## Level 1: 1 nodes to be scored (12190 eliminated genes)

```



```
all_repressed_targets_GO_plot <- fisherGOplot(all_repressed_targets_GO)
all_repressed_targets_GO_plot
```



```
ggsave(filename = "../03_output/GO_plots/all_repressed_targets_GO_plot.pdf", plot = all_repressed_targets_GO_plot)
```

## Session info

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats4 parallel stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] forcats_0.5.1 stringr_1.4.0 dplyr_1.0.8
```

```

## [4] purrr_0.3.4      readr_2.1.2      tidyr_1.2.0
## [7] tibble_3.1.6     ggplot2_3.3.5    tidyverse_1.3.1
## [10] topGO_2.44.0     SparseM_1.81     GO.db_3.13.0
## [13] AnnotationDbi_1.54.1 IRanges_2.26.0   S4Vectors_0.30.2
## [16] Biobase_2.52.0   graph_1.70.0     BiocGenerics_0.38.0
##
## loaded via a namespace (and not attached):
## [1] bitops_1.0-7      matrixStats_0.61.0 fs_1.5.2
## [4] lubridate_1.8.0   bit64_4.0.5       RColorBrewer_1.1-3
## [7] httr_1.4.2        GenomeInfoDb_1.28.4 tools_4.1.0
## [10] backports_1.4.1   utf8_1.2.2        R6_2.5.1
## [13] DBI_1.1.2         colorspace_2.0-3   withr_2.5.0
## [16] tidyselect_1.1.2  bit_4.0.4          compiler_4.1.0
## [19] cli_3.2.0         rvest_1.0.2        xml2_1.3.3
## [22] labeling_0.4.2    scales_1.2.0       digest_0.6.29
## [25] rmarkdown_2.13    XVector_0.32.0     pkgconfig_2.0.3
## [28] htmltools_0.5.2   highr_0.9          dbplyr_2.1.1
## [31] fastmap_1.1.0     rlang_1.0.2        readxl_1.4.0
## [34] rstudioapi_0.13   RSQlite_2.2.12     farver_2.1.0
## [37] generics_0.1.2    jsonlite_1.8.0     vroom_1.5.7
## [40] RCurl_1.98-1.6    magrittr_2.0.3     GenomeInfoDbData_1.2.6
## [43] Rcpp_1.0.8.3      munsell_0.5.0      fansi_1.0.3
## [46] lifecycle_1.0.1   stringi_1.7.6      yaml_2.3.5
## [49] zlibbioc_1.38.0   grid_4.1.0         blob_1.2.3
## [52] crayon_1.5.1      lattice_0.20-45     Biostings_2.60.2
## [55] haven_2.4.3       hms_1.1.1          KEGGREST_1.32.0
## [58] knitr_1.38        pillar_1.7.0       reprex_2.0.1
## [61] glue_1.6.2        evaluate_0.15      modelr_0.1.8
## [64] png_0.1-7         vctrs_0.4.0        tzdb_0.3.0
## [67] cellranger_1.1.0  gtable_0.3.0       assertthat_0.2.1
## [70] cachem_1.0.6      xfun_0.30          broom_0.8.0
## [73] memoise_2.0.1     ellipsis_0.3.2

```