Supplementary material for: Rodriguez-Mena, A., Tarragó-Claramunt, X., Castellani, G., Méndez-Viera, J., & Monleón-Getino, A. (2025). Gango + BioFunctional: A Computational tool for efficient functional gene analysis. *The European chemistry and biotechnology journal, 4*, 69-80. https://doi.org/10.62063/ecb-63. Corresponce: Antonio Monleón-Getino, Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Barcelona, Spain. E-mail: amonleong@ub.edu (A.M.G.); Ph.: ++34-678329864.

Appendix section

Tutorial: GANGO + BioFunctional

This tutorial provides a step-by-step guide on using GANGO and BioFunctional in R with a real example.

The RNA-Seq data used were retrieved from the study “Dysregulated transcriptional responses to SARS-CoV-2 in the periphery” (McClain et al, 2020), which analyzed samples from subjects affected by SARS-CoV-2 infection, seasonal coronavirus, influenza, bacterial pneumonia, as well as healthy controls.

The RNA Sequencing dataset and associated sample metadata were obtained from the public repository GEO (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE161731.

To identify genes differentially expressed between Bacterial and Healthy cohorts, RNA-seq count data were processed using the **DESeq2** package (Love et al., 2014).

To focus the analysis on protein-coding genes, gene biotypes were retrieved using the **biomaRt** package. Only genes annotated as protein_coding in the Ensembl database (GRCh38, hsapiens_gene_ensembl) were retained.

Normalization and differential expression analysis were performed using the DESeq() function. Genes with an adjusted p-value (padj) < 0.1 were considered statistically significant. These genes were further stratified by their log2 fold-change direction: genes upregulated in the Bacterial group (log2FC > 0) and those upregulated in the Healthy group (log2FC < 0) were saved separately.

The final output was a table containing significant genes with their associated Ensembl IDs, affiliation group (Bacterial or Healthy), and organism taxonomy (*Homo sapiens*).

This file was used as input for GANGO.

The GANGO + BioFuncional demo showcases a user-friendly, R-based application designed to streamline the functional analysis of gene expression data. It's important to note that the demo is derived from data presented in a study on transcriptional responses, and focuses on the functional analysis between a group of healthy individuals and a group with bacterial infection. The source of this data is McClain et al. (2020; https://pubmed.ncbi.nlm.nih.gov/32743603/).

For this demo, all necessary files for the analysis are provided in the correct format and can be download in https://github.com/amonleong/Biofunctional.

The demo illustrates how the application processes gene lists or taxonomic classifications to elucidate Gene Ontology (GO) term and KEGG pathway enrichment. It highlights the two main modules:

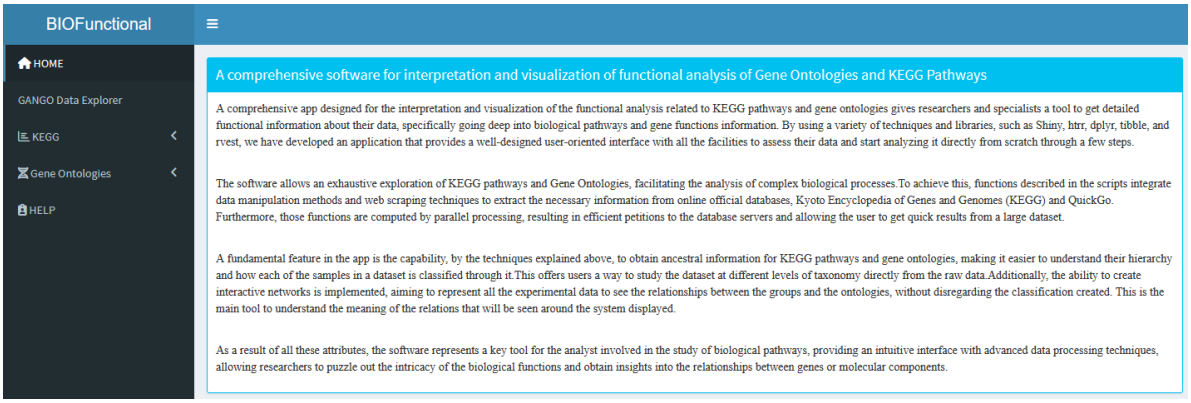
- **GANGO:** Efficiently maps input data (genes/taxa/groups) to GO terms and KEGG pathways.
- **BioFuncional:** Interprets the output from GANGO, incorporating hierarchical information, generating interactive networks, and producing bar plot visualizations.

1. First Steps

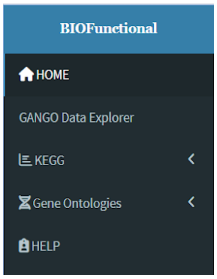
- **Download example files for BioFunctional:** Download sample files from the repository https://github.com/amonleong/Biofunctional .

2. Run the BioFunctional Application

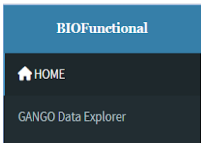
Execute the BioFunctional application in https://alexub.shinyapps.io/BioFunctional/. This will launch the application and display its main menu.



This screen contains the menu:

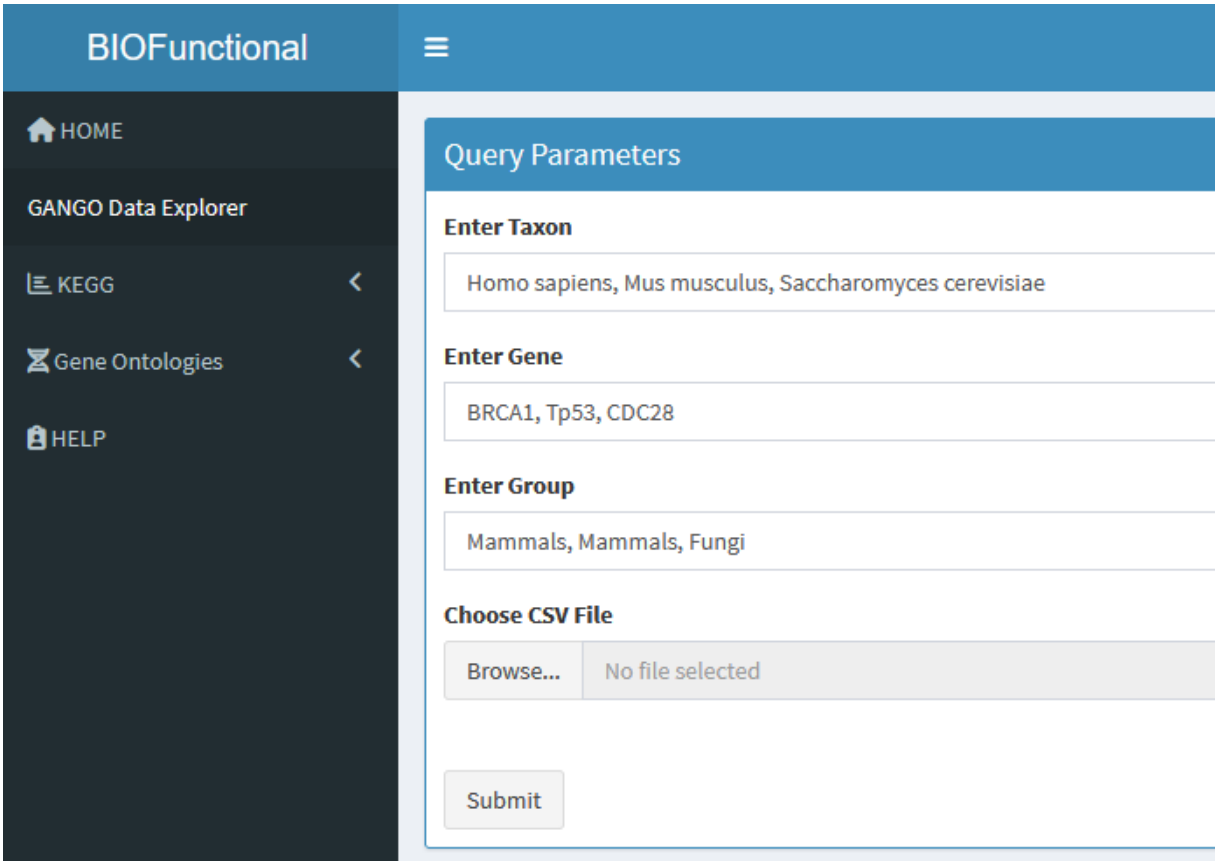


GANGO



5. GANGO: Gene/Taxon/Group to Gene Ontologies and KEGG

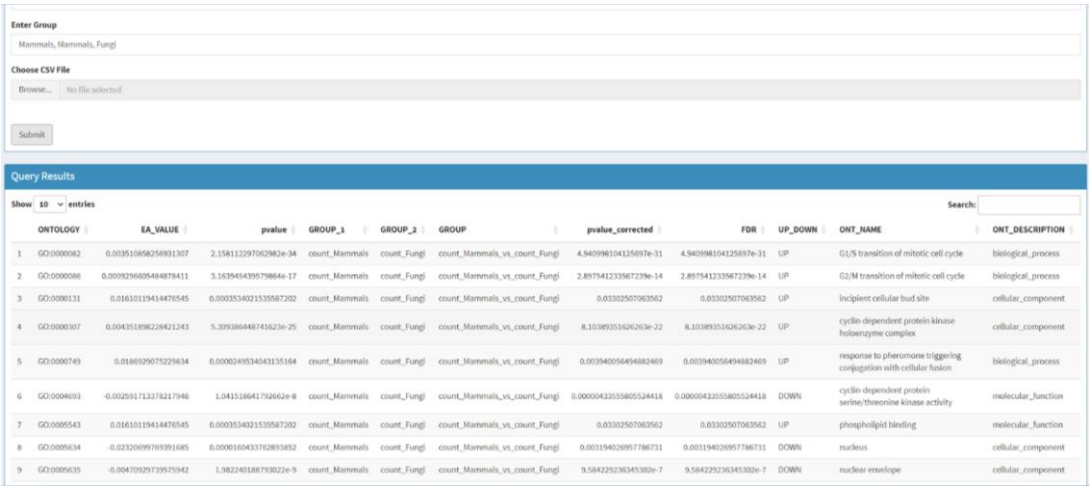
- **Initial Screen:** GANGO's interface allows you to upload your data.
- **Input:** GANGO takes as input a file containing your taxon, gene and group identifiers.
- **Using the example:** Load the example dataset provided with BioFunctional to familiarize yourself with the input format: “GANGO_BACTERIA_HEALTHY_2.csv”
- **Submit:** Run GANGO to map your identifiers to Gene Ontology (GO) terms and KEGG pathways.
  IMPORTANT: The process can take 2-3 hours, resulting in file “GANGO_significant_results_2025-05-09.csv”. (You can skip this step and check the result file directly if preferred.) Move on to the next step.
- **Output:** GANGO will generate a file containing the mapping of your input identifiers to GO terms and KEGG pathways: “GANGO_significant_results_2025-05-09.csv”
- **Download results:** Download the results file. This file serves as the input for BioFunctional.

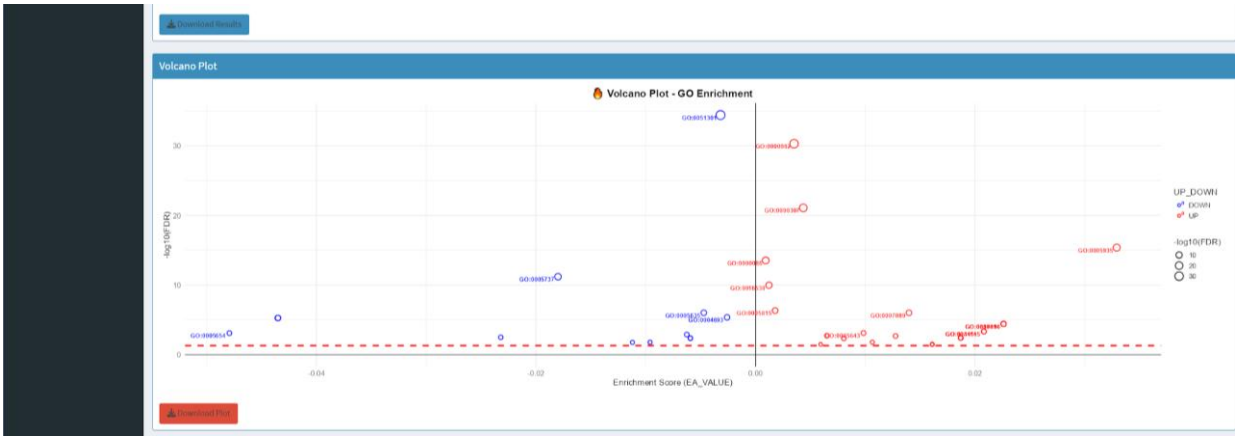


GANGO screen with defect example

Upload the required data in the correct format (eg. GANGO_BACTERIA_HEALTHY_2.csv) and click Submit

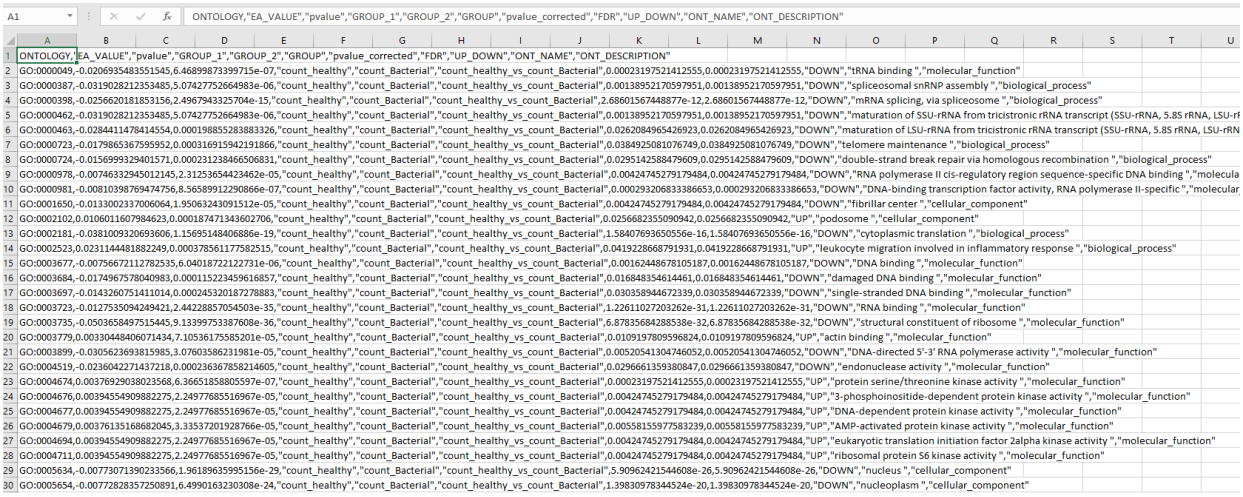
Screen with the results of the user example (Bacteria vs Health) after 2h processing:





The computational time can be high in this process (ex: for 9000 genes use 2.50 hours). The results also present a volcano plot (A volcano plot is a graph that combines the magnitude of change (X-axis) and statistical significance (Y-axis) to quickly visualize which elements (like genes) show large and statistically important differences in a data analysis.)

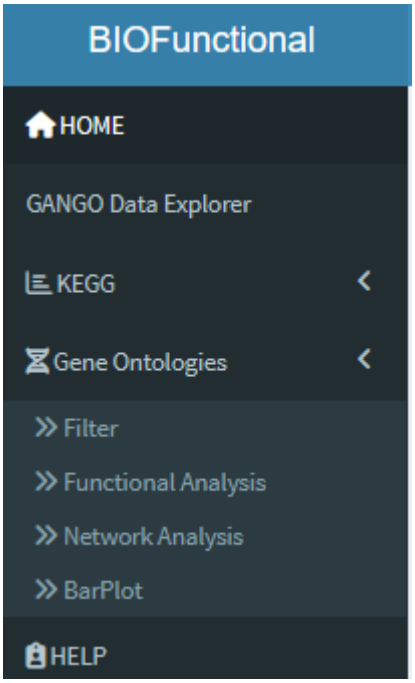
Download FILE RESULTS: "GANGO_significant_results_2025-05-09.csv"



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	ONTOLOGY,"EA_VALUE","pvalue","GROUP_1","GROUP_2","GROUP","pvalue_corrected","FDR","UP_DOWN","ONT_NAME","ONT_DESCRIPTION"																				
2	GO:0000049,-0.0206935483551545,6.468998739971 5e-07,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.00023197521412555,0.00023197521412555,"DOWN","tRNA binding ","molecular_function"																				
3	GO:0000387,-0.0319028212353485,5.0742775266498 3e-06,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.00138952170597951,0.00138952170597951,"DOWN","spliceosomal snRNP assembly ","biological_process"																				
4	GO:0000398,-0.0256620181853156,2.496794332570 4e-15,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",2.686015674488 77e-12,2.686015674488 77e-12,"DOWN","mRNA splicing, via spliceosome ","biological_process"																				
5	GO:0000462,-0.0319028212353485,5.0742775266498 3e-06,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.00138952170597951,0.00138952170597951,"DOWN","maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rR																				
6	GO:0000463,-0.0284411478414554,0.000198855283883326,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0262084965426923,0.0262084965426923,"DOWN","maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRN																				
7	GO:0000723,-0.0179865367595952,0.000316915942191866,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0384925081076749,0.0384925081076749,"DOWN","telomere maintenance ","biological_process"																				
8	GO:0000724,-0.0156999329401571,0.000231238466506831,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0295142588479609,0.0295142588479609,"DOWN","double-strand break repair via homologous recombination ","biological_process"																				
9	GO:0000978,-0.0074633294501214 5,2.312536544234 62e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"DOWN","RNA polymerase II cis-regulatory region sequence-specific DNA binding ","molecula																				
10	GO:0000981,-0.0081039876947 4756,8.565899122908 66e-07,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0002932068333386653,0.0002932068333386653,"DOWN","DNA-binding transcription factor activity, RNA polymerase II-specific ","molecular																				
11	GO:0001650,-0.0133002337006064,1.950632430915 12e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"DOWN","fibrillar center ","cellular_component"																				
12	GO:0002102,0.0106011607984623,0.000187471343602706,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0256682355090942,0.0256682355090942,"UP","podosome ","cellular_component"																				
13	GO:0002181,-0.0381009320693606,1.156951484068 86e-19,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",1.5840769365055 6e-16,1.5840769365055 6e-16,"DOWN","cytoplasmic translation ","biological_process"																				
14	GO:0002523,0.0231144481882249,0.000378561177582515,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0419228668791931,0.0419228668791931,"UP","leukocyte migration involved in inflammatory response ","biological_process"																				
15	GO:0003677,-0.0075667211278 2535,6.040187221227 31e-06,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0016244867810 5187,0.0016244867810 5187,"DOWN","DNA binding ","molecular_function"																				
16	GO:0003684,-0.0174967578040983,0.000115223459616857,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0168483546144 61,0.0168483546144 61,"DOWN","damaged DNA binding ","molecular_function"																				
17	GO:0003697,-0.0143260751411014,0.000245320187278883,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0303589446723 39,0.0303589446723 39,"DOWN","single-stranded DNA binding ","molecular_function"																				
18	GO:0003723,-0.0127553094249421,2.442288570545 03e-35,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",1.2261102720326 2e-31,1.2261102720326 2e-31,"DOWN","RNA binding ","molecular_function"																				
19	GO:0003735,-0.0503658497515445,9.133997533876 08e-36,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",6.878356842885 38e-32,6.878356842885 38e-32,"DOWN","structural constituent of ribosome ","molecular_function"																				
20	GO:0003779,0.0033044840607 1434,7.105361755852 01e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0109197809596824,0.0109197809596824,"UP","actin binding ","molecular_function"																				
21	GO:0003899,-0.0305623693815985,3.076035862319 81e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0052054130474 6052,0.0052054130474 6052,"DOWN","DNA-directed 5'-3' RNA polymerase activity ","molecular_function"																				
22	GO:0004519,-0.0236042271437218,0.000236378582 14605,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0296661359380847,0.0296661359380847,"DOWN","endonuclease activity ","molecular_function"																				
23	GO:0004674,0.0037692903802 3568,6.366518588055 97e-07,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0002319752141 2555,0.0002319752141 2555,"UP","protein serine/threonine kinase activity ","molecular_function"																				
24	GO:0004676,0.0039455490988 2275,2.249776855169 67e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"UP","3-phosphoinositide-dependent protein kinase activity ","molecular_function"																				
25	GO:0004677,0.0039455490988 2275,2.249776855169 67e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"UP","DNA-dependent protein kinase activity ","molecular_function"																				
26	GO:0004679,0.0037613516868 2045,3.335372019287 66e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0055815597758 3239,0.0055815597758 3239,"UP","AMP-activated protein kinase activity ","molecular_function"																				
27	GO:0004694,0.0039455490988 2275,2.249776855169 67e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"UP","eukaryotic translation initiation factor 2alpha kinase activity ","molecular_function"																				
28	GO:0004711,0.0039455490988 2275,2.249776855169 67e-05,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",0.0042474527917 9484,0.0042474527917 9484,"UP","ribosomal protein S6 kinase activity ","molecular_function"																				
29	GO:0005634,-0.0077282835725089 1,6.499016323030 8e-24,"count_healthy","count_Bacterial","count_healthy_vs_count_Bacterial",1.3983097834452 4e-20,1.3983097834452 4e-20,"DOWN","nucleoplasm ","cellular_component"																				

"GANGO_significant_results_2025-05-09.csv" file

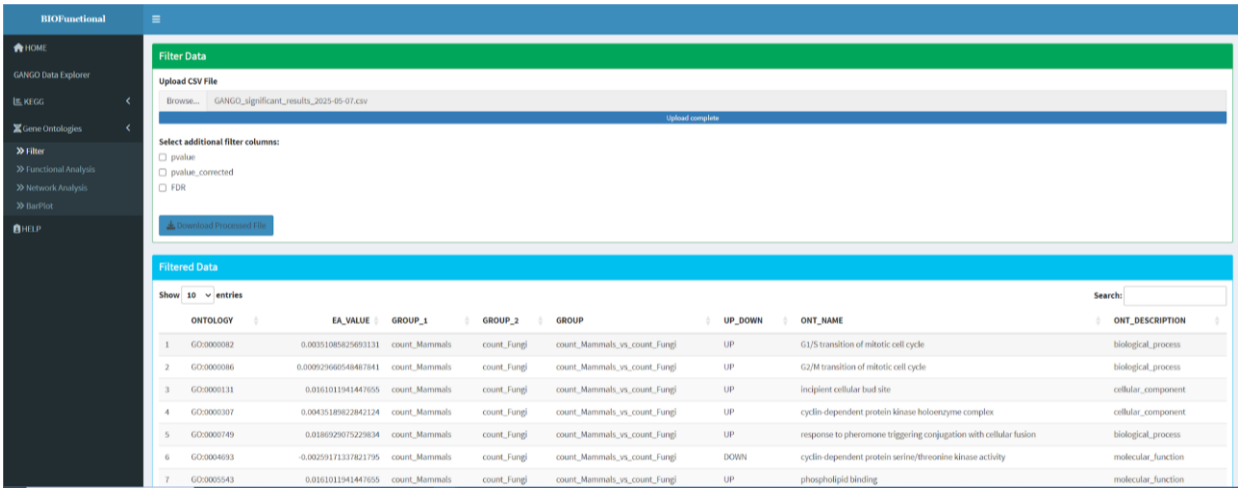
BIOFUNCTIONAL



The program is also capable of analyzing KEGG, but in this example we will focus on Gene Ontologies.

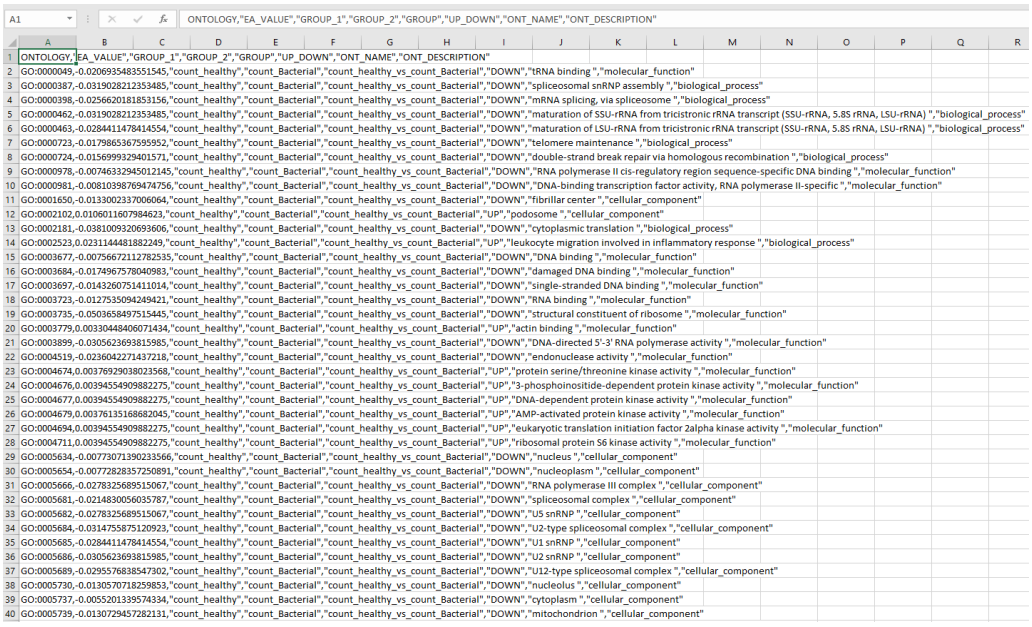
3. BioFunctional: Gene Ontology Analysis

- **Gene Ontologies (Input):** BioFunctional uses the output file from GANGO (or a manually created file) containing gene ontologies: "GANGO_significant_results_2025-05-09.csv"
- **Column Selection:** BioFunctional automatically selects essential columns from the input file to guarantee the proper operation of the different algorithms. You can select additional columns of your interest for filtering and visualization in subsequent steps.
- **Download Preprocessed Files (OUTPUT):** Download the preprocessed data. This is an intermediate step to prepare the data for functional análisis: "filtered_data_2025-05-13.csv"



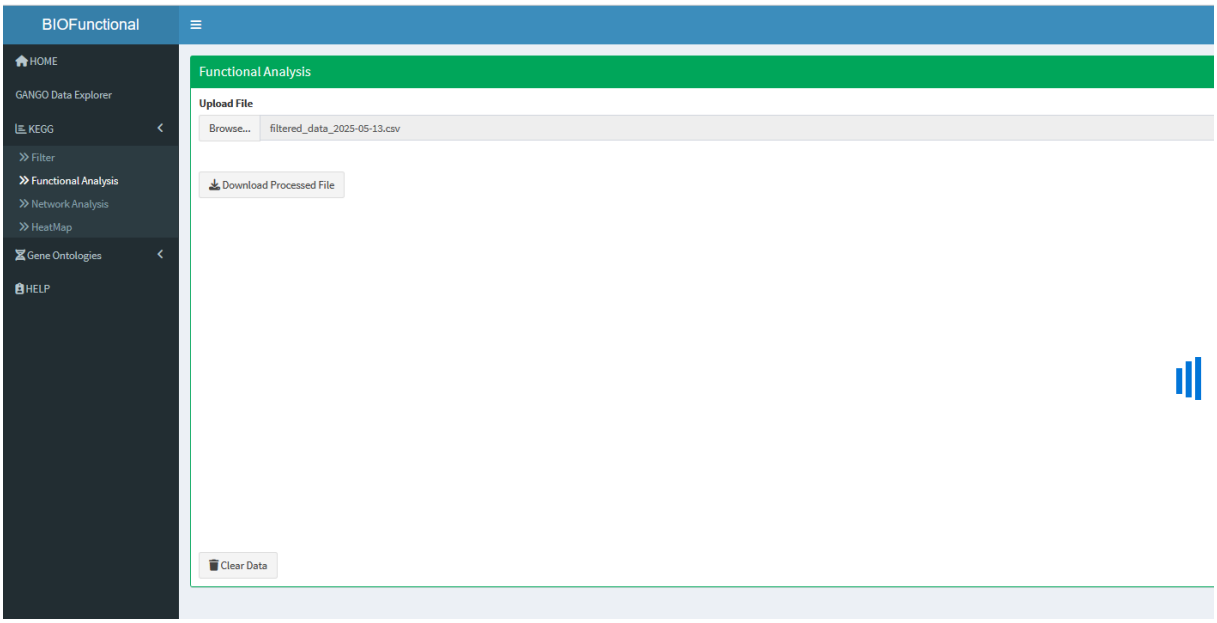
BioFunctional pre-selects essential columns from the input file. You can select additional columns for filtering and visualization in subsequent steps.

DOWNLOAD PREPROCESS FILES: "filtered_data_2025-05-13.csv"

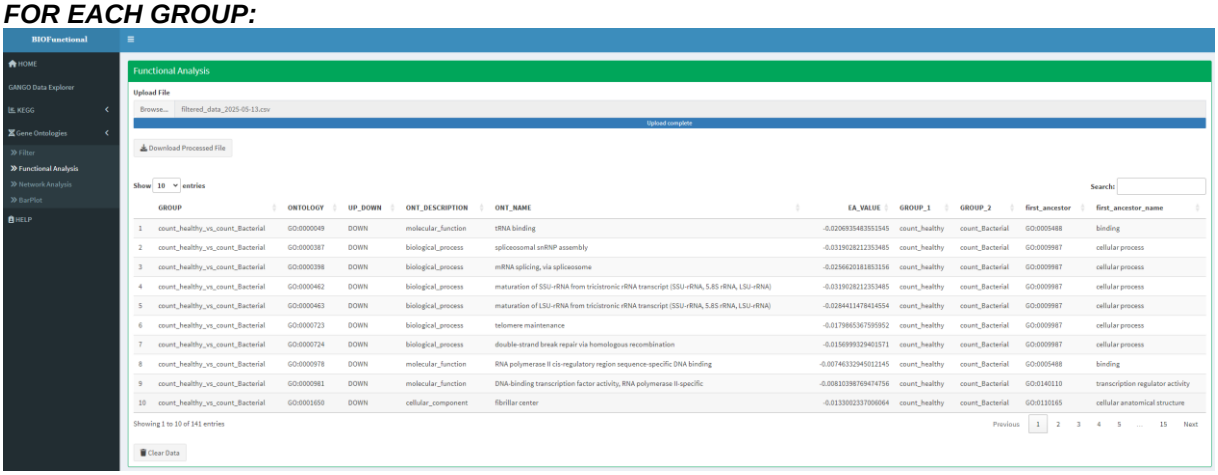


4. Functional Analysis

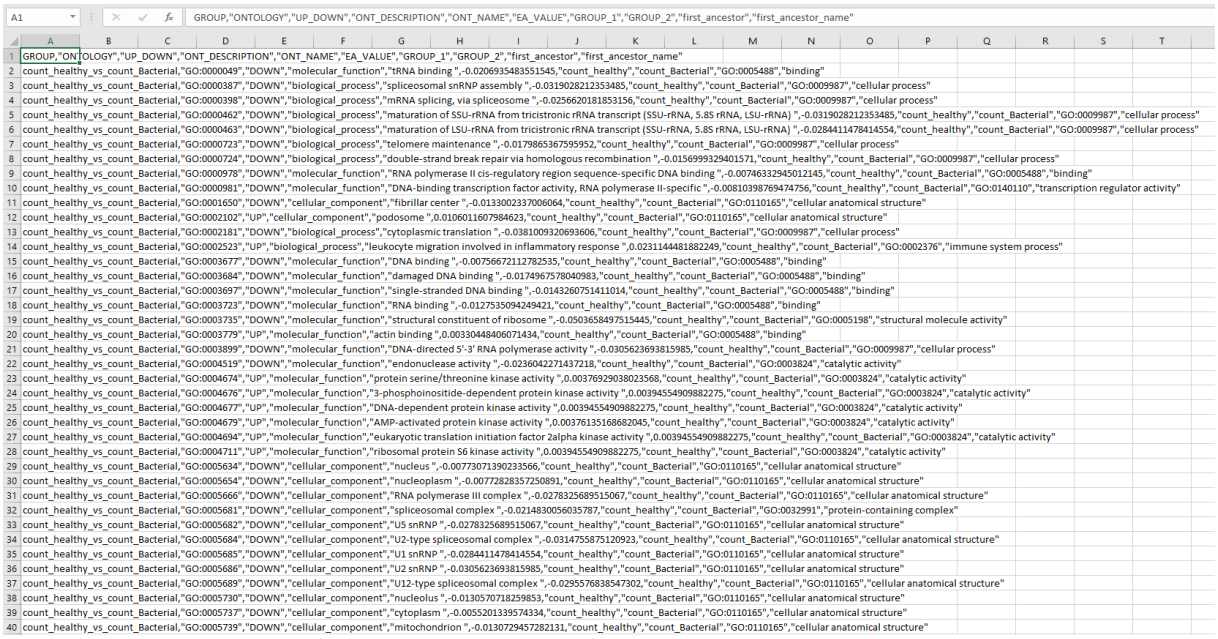
- **Input:** Use the preprocessed file obtained in the previous step: "filtered_data_2025-05-13.csv"
- **Process:** BioFunctional analyzes the Gene Ontology terms and determines their hierarchies for each group in your data.
- **Output:** The tool adds columns to the table, indicating the GO term hierarchies:" go.csv"
- **Download Processed File:** Download the processed file, which now includes GO term hierarchy information.



GET THE HIERARCHIES OF THE GENE ONTOLOGIES AND INDICATE THEM IN THE TABLE FOR EACH GROUP:



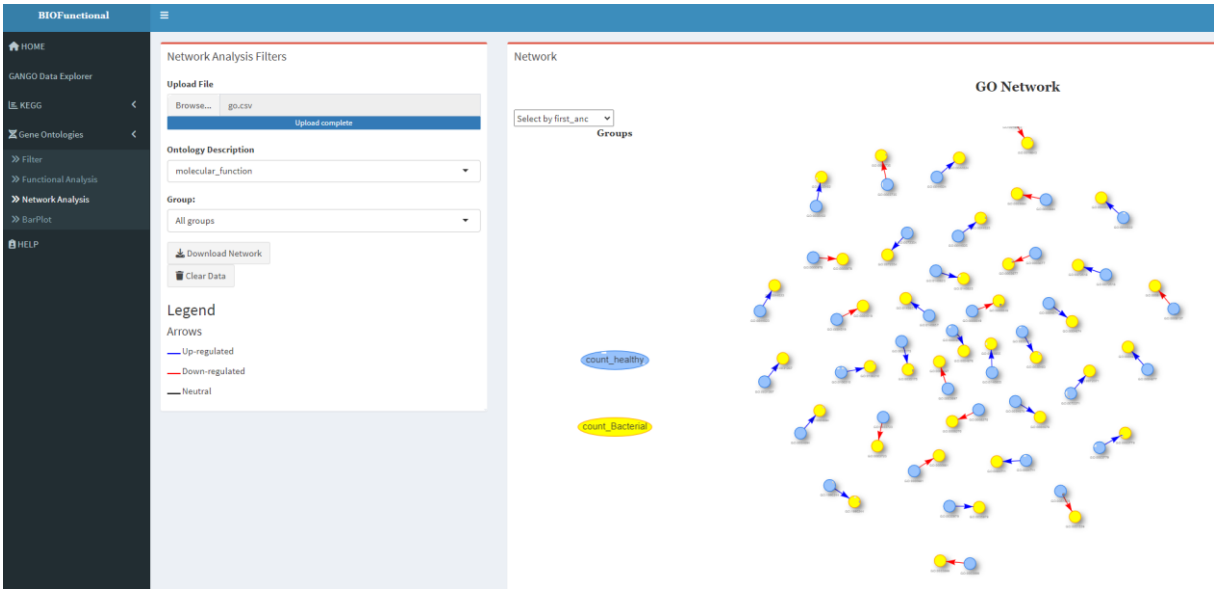
DOWNLOAD PROCESSED FILE: go.csv

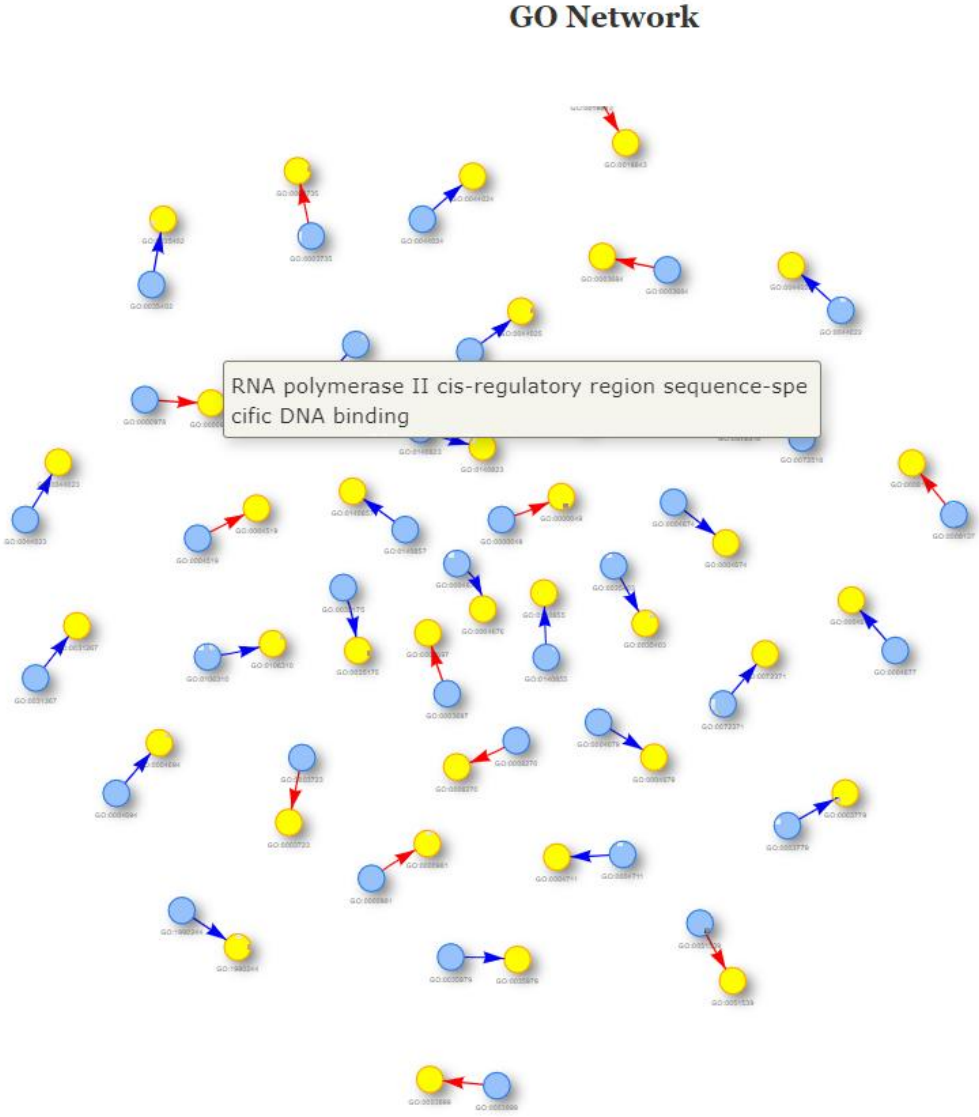


GROUP	ONTOLOGY	UP_DOWN	ONT_DESCRIPTION	ONT_NAME	EA_VALUE	GROUP_1	GROUP_2	first_ancestor	first_ancestor_name
count_healthy_vs_count_Bacterial	GO:0000049	DOWN	molecular_function	tRNA binding	-0.0206935483551545	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0000387	DOWN	biological_process	spliceosomal snRNP assembly	-0.0319028212353485	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000398	DOWN	biological_process	mRNA splicing, via spliceosome	-0.0256620181853156	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000462	DOWN	biological_process	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	-0.0319028212353485	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000463	DOWN	biological_process	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	-0.0284411478414554	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000723	DOWN	biological_process	telomere maintenance	-0.0179865367595952	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000724	DOWN	biological_process	double-strand break repair via homologous recombination	-0.0156999329401571	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0000978	DOWN	molecular_function	RNA polymerase II cis-regulatory region sequence-specific DNA binding	-0.0074633294501214	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0000981	DOWN	molecular_function	DNA-binding transcription factor activity, RNA polymerase II-specific	-0.0081039876947456	count_healthy	count_Bacterial	GO:0140110	transcription regulator activity
count_healthy_vs_count_Bacterial	GO:0001650	DOWN	cellular_component	fibrillar center	-0.0133002337006064	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0002102	UP	cellular_component	podosome	0.0106011607984623	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0002181	DOWN	biological_process	cytoplasmic translation	-0.038100932693606	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0002523	UP	biological_process	leukocyte migration involved in inflammatory response	0.0231144481882249	count_healthy	count_Bacterial	GO:0002376	immune system process
count_healthy_vs_count_Bacterial	GO:0003677	DOWN	molecular_function	DNA binding	-0.00756672112782535	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0003684	DOWN	molecular_function	damaged DNA binding	-0.0174967578040983	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0003697	DOWN	molecular_function	single-stranded DNA binding	-0.0143260751411014	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0003723	DOWN	molecular_function	RNA binding	-0.0127535094249421	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0003735	DOWN	molecular_function	structural constituent of ribosome	-0.0503658497515445	count_healthy	count_Bacterial	GO:0005198	structural molecule activity
count_healthy_vs_count_Bacterial	GO:0003779	UP	molecular_function	actin binding	0.00330448406071434	count_healthy	count_Bacterial	GO:0005488	binding
count_healthy_vs_count_Bacterial	GO:0003899	DOWN	molecular_function	DNA-directed 5'-3' RNA polymerase activity	-0.0305623693815985	count_healthy	count_Bacterial	GO:0009987	cellular process
count_healthy_vs_count_Bacterial	GO:0004519	DOWN	molecular_function	endonuclease activity	-0.0236042271437218	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004674	UP	molecular_function	protein serine/threonine kinase activity	0.00376929038023568	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004676	UP	molecular_function	3-phosphoinositide-dependent protein kinase activity	0.00394554909882275	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004677	UP	molecular_function	DNA-dependent protein kinase activity	0.00394554909882275	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004679	UP	molecular_function	AMP-activated protein kinase activity	0.00376135168682045	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004694	UP	molecular_function	eukaryotic translation initiation factor 2alpha kinase activity	0.00394554909882275	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0004711	UP	molecular_function	ribosomal protein S6 kinase activity	0.00394554909882275	count_healthy	count_Bacterial	GO:0003824	catalytic activity
count_healthy_vs_count_Bacterial	GO:0005634	DOWN	cellular_component	nucleus	-0.00773071390233566	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005654	DOWN	cellular_component	nucleoplasm	-0.00772828357250891	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005666	DOWN	cellular_component	RNA polymerase III complex	-0.0278325689515067	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005681	DOWN	cellular_component	spliceosomal complex	-0.0214830056035787	count_healthy	count_Bacterial	GO:0032991	protein-containing complex
count_healthy_vs_count_Bacterial	GO:0005682	DOWN	cellular_component	U5 snRNP	-0.0278325689515067	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005684	DOWN	cellular_component	U2-type spliceosomal complex	-0.0314755875120923	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005685	DOWN	cellular_component	U1 snRNP	-0.0284411478414554	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005686	DOWN	cellular_component	U2 snRNP	-0.0305623693815985	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005689	DOWN	cellular_component	U12-type spliceosomal complex	-0.0295576838547302	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005730	DOWN	cellular_component	nucleolus	-0.0130570718259853	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005737	DOWN	cellular_component	cytoplasm	-0.0055201339574334	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure
count_healthy_vs_count_Bacterial	GO:0005739	DOWN	cellular_component	mitochondrion	-0.0130729457282131	count_healthy	count_Bacterial	GO:0110165	cellular anatomical structure

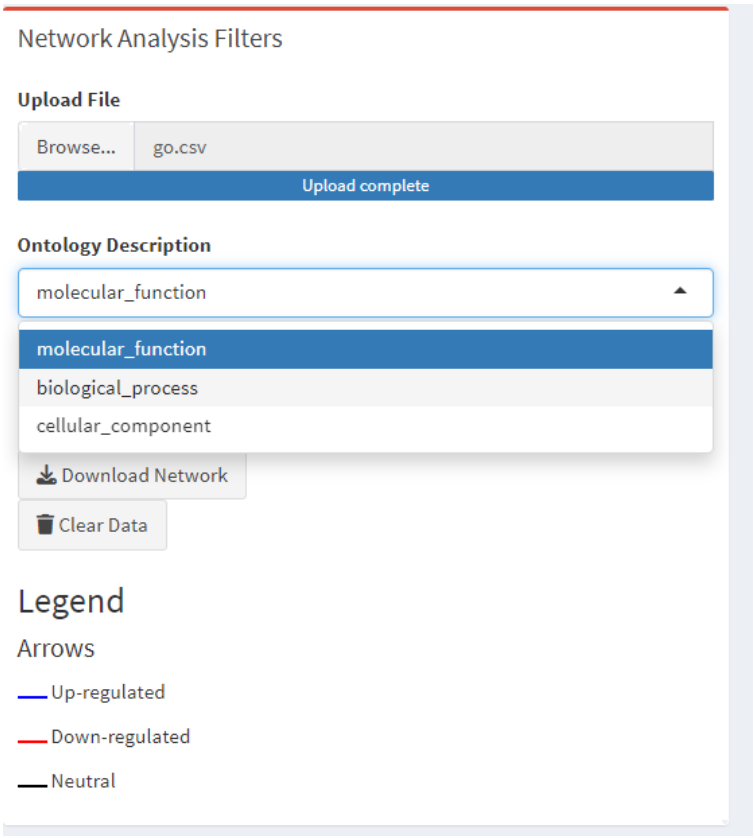
5. Network Analysis

- **Input:** Use the ontology table FILE from the previous step: “go.csv”
- **Selection:** BioFunctional allows you to select GO terms based on:
  - GO category (Biological Process, Cellular Component, Molecular Function).
  - Hierarchy filters.
- **Visualization:** BioFunctional creates a network graph that represents the relationships between chosen GO terms. This allows users to examine how ontology relationships vary across different experimental groups. The tool provides extensive options for filtering and visualizing the network, such as selecting by ontology and reordering groups to highlight specific comparisons.

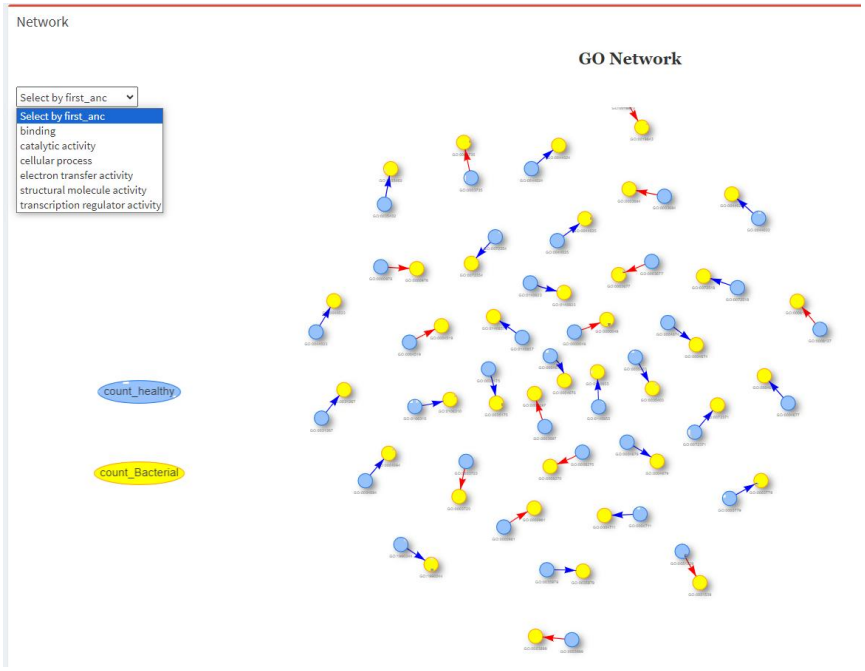




There are different options and drop-down menus to SELECT BY GO BIOLOGICAL, FUNCTION, ETC



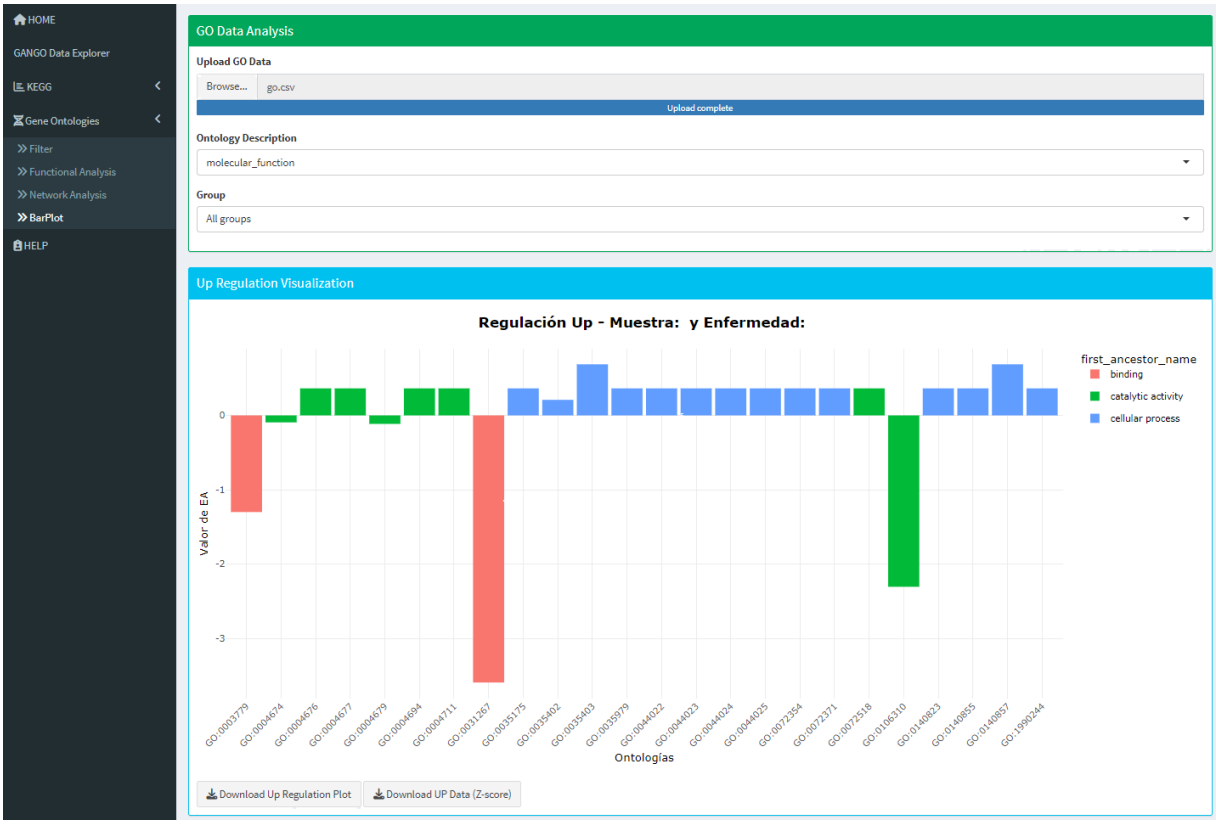
ALSO FOR DIFFERENT HIERARCHY FILTERS

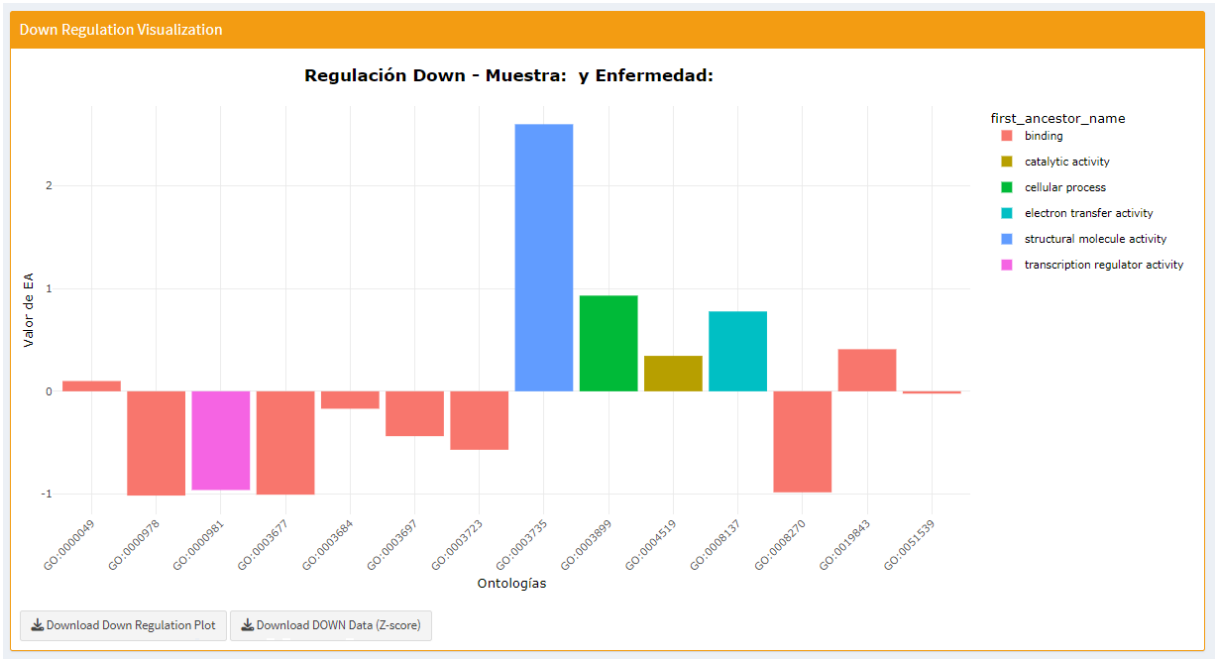


6. Bar Plot

- **Input:** Use the GO term file from the functional análisis: “go.csv”

- **Visualization:** BioFunctional creates a bar plot showing the enrichment of specific GO terms.





DOWNLOAD FILE: "DOWN_data_zscore_2025-05-13.csv"

7.Interpreting Gene Ontology (GO) Data obtained in Biofunctional: Challenges and Manual Interpretation

Gene Ontology (GO) provides a structured vocabulary to describe gene functions across different organisms. It categorizes these functions into three domains:

- **Biological Process:** Describes biological objectives to which the gene contributes. (e.g., "cell cycle")
- **Cellular Component:** Describes where the gene product acts. (e.g., "nucleus")
- **Molecular Function:** Describes the gene product's biochemical activity. (e.g., "DNA binding")

Why is GO Interpretation Difficult?
Interpreting GO data, especially after a GO enrichment analysis (identifying over-represented GO terms in a set of genes), can be challenging for several reasons:

- **Hierarchical Complexity:** GO terms are organized in a complex, hierarchical structure (a Directed Acyclic Graph, or DAG). A gene can be associated with multiple terms at different levels, making it difficult to pinpoint the most relevant functions.
- **Redundancy:** Due to the hierarchical structure, some GO terms may be redundant, with broader terms encompassing more specific ones. This can make it hard to extract precise biological insights.
- **Context Dependence:** The biological significance of a GO term can vary depending on the specific experimental context, organism, or tissue type.
- **Annotation Bias:** Some genes are better studied and annotated than others, leading to bias in GO term representation. Well-studied genes may show up more frequently in GO enrichment results, potentially overshadowing the roles of less-studied but equally important genes.
- **Large Datasets:** High-throughput experiments often generate long lists of significantly enriched GO terms, making it difficult to manually sift through and identify the most meaningful ones.

7.1.How to Perform Manual Interpretation
Despite these challenges, manual interpretation by experts is crucial for drawing accurate and biologically relevant conclusions. Here's how a biologist or clinician might approach it:

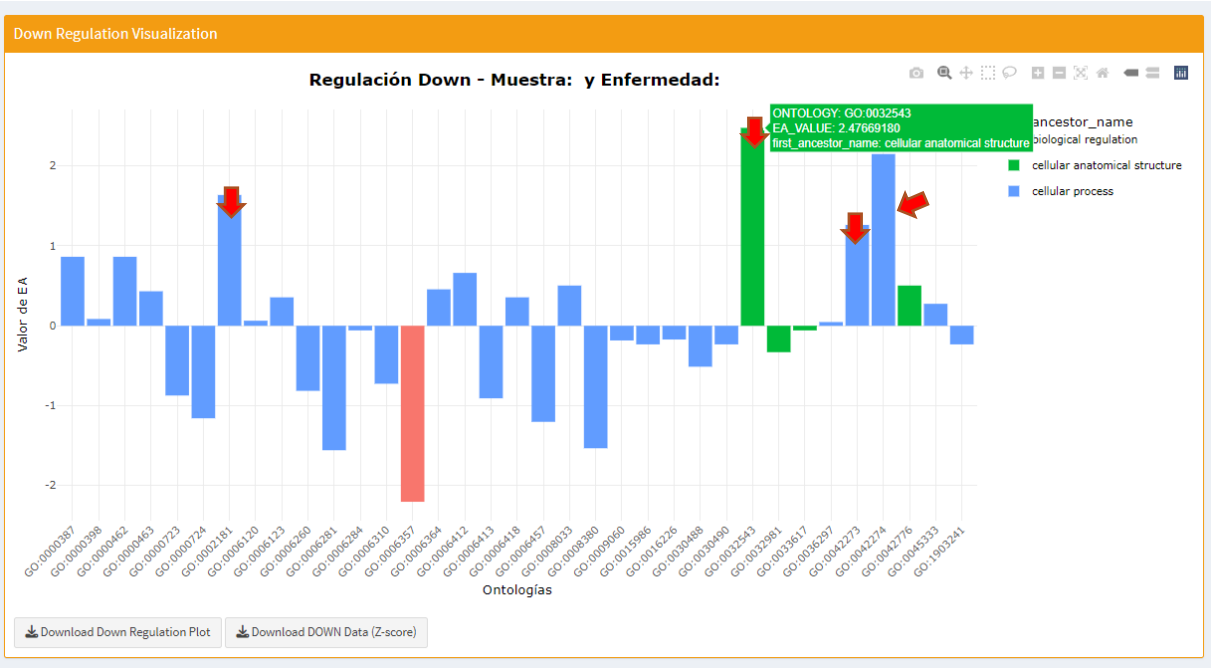
1. **Prioritize Relevant GO Terms:**
    - Focus on the most significantly enriched GO terms (based on statistical measures like p-values or Z-values in Biofunctional using the barplots).
    - Consider the specific biological question or experimental context to narrow down the list.
2. **Navigate the GO Hierarchy:**
    - Use GO browsers (like AmiGO) to explore the relationships between GO terms.
    - Identify parent and child terms to understand the broader and more specific functions associated with the gene set.
    - Look for the "sweet spot" in the hierarchy: terms that are specific enough to be informative but not so granular that they become overwhelming.
3. **Literature Review (Bibliography):**
    - Consult scientific literature (PubMed, etc.) to understand the biological functions associated with the identified GO terms.
    - Investigate the genes annotated with those terms and their known roles in the relevant biological processes or disease conditions.
    - Look for supporting evidence that connects the GO terms to the experimental findings.
4. **Cross-referencing and Validation (Checking Results):**
    - Compare the GO analysis results with other relevant data, such as:
        - Other functional enrichment analyses (e.g., KEGG pathway analysis).
        - Gene expression data.
        - Protein-protein interaction networks.
        - Phenotype data.
    - Validate the findings by checking if the identified GO terms align with the expected biological functions based on prior knowledge.
    - Look for consistency and convergence of evidence from multiple sources.

By combining computational analysis with expert knowledge and thorough literature review, biologists and clinicians can effectively interpret GO data and gain valuable insights into gene function and their roles in biological systems.

***7.2*-GO Term Interpretation with CURIE Technology**

The GO terms of interest can also be interpreted using a novel technology developed by the BIOST3 group at the University of Barcelona, called CURIE. This technology is currently in an experimental phase and pending patent or registration. While highly efficient and requiring very few cases to be effectively utilized, it remains experimental and may be subject to potential errors or inconsistencies

The GO of interest in the plot. In this example we are interested in the GO enriched significatively in the bacteria group for biological process (Negatives, or regulation down) can be downloaded for study using option "Download DOWN DATA (Z-SCORE)" button. Obtain file: "DOWN_data_zscore_2025-05-13.csv"



This gene ontologies can be extracted manually (expert criteria) or using an special R script (SEE "go selection.R"):

In the regulation down plot (related to bacterial infection and biological process), the displayed ontologies serve as an illustrative example of the many that were detected. It's crucial to note that users must determine the most relevant ontologies for their specific research (RED ARROWS). For instance, this could involve selecting all detected ontologies, or focusing solely on those with the most positive or negative Z-score values. The preceding graph highlights the most positively enriched ontologies, and as a further example, CURIE has interpreted one of these ontologies. It's important to understand that the actual interpretation work is far more extensive than what's shown in this brief example.

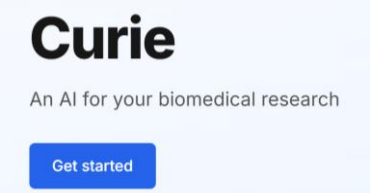
Ontologies in the regulation down plot (related to bacterial infection and biological process):

ONTOLOGY	ONT_DESCRIPTION	ONT_NAME
12 GO:0002181	biological_process	cytoplasmic translation
55 GO:0006357	biological_process	regulation of transcription by RNA polymerase II
89 GO:0032543	biological_process	mitochondrial translation
106 GO:0042274	biological_process	ribosomal small subunit biogenesis

GO Term Codes: These GO terms can be converted into a comma-separated list of description codes using an R script. This is crucial for interpreting the results and use an artificial intelligence:

“cytoplasmic translation ,regulation of transcription by RNA polymerase II ,mitochondrial translation ,ribosomal small subunit biogénesis”

EXAMPLE INTERPRETATION USING AN EXTERNAL ARTIFICIAL INTELLIGENCE (CURIE AI) FOR GO TERMS***:*** “cytoplasmic translation and bacterial infection”



CURIE AI
09:42

The study of cytoplasmic translation is crucial in understanding bacterial infection mechanisms. It involves the process by which ribosomes translate messenger RNA (mRNA) into proteins, which are essential for bacterial growth, survival, and pathogenicity. During bacterial infection, pathogens employ various strategies to manipulate the host's cellular machinery, including hijacking the translation machinery for their own benefit. This can involve the production of effector proteins that modulate host cell functions or the use of virulence factors that directly damage the host. Recent studies have highlighted the role of specific bacterial proteins in interacting with host ribosomes to enhance translation efficiency. For example, the jumbophage protein interacts with both the ribosome and the phage nucleus, facilitating efficient propagation of the phage within the host cell. Additionally, the GTPase BipA plays a dual role in ribosome assembly and surface structure biogenesis in Salmonella enterica serovar Typhimurium, demonstrating the importance of translation regulation in bacterial virulence. Understanding the molecular mechanisms of cytoplasmic translation during bacterial infection provides opportunities for developing targeted therapies. Interrupting protein synthesis pathways or enhancing the immune response against bacterial pathogens could lead to novel treatment strategies. Future research should focus on elucidating the precise mechanisms by which bacteria manipulate the host's translation machinery. This knowledge will contribute to the development of new antibiotics and therapeutic interventions against bacterial infections.

References:
- A ribosome-interacting jumbophage protein associates with the phage nucleus to facilitate efficient propagation. (PMID: 39992933)
- The dual functions of the GTPase BipA in ribosome assembly and surface structure biogenesis in Salmonella enterica serovar Typhimurium. (PMID: 40203049)
- Molecular basis for the phosphorylation of bacterial tyrosine kinase Wzc. (PMID: 40210632)
- Analysis of cytosolic mtDNA release during Staphylococcus aureus infection. (PMID: 40058964)
- Deficiency in peptidoglycan recycling promotes β-lactam sensitivity in Caulobacter crescentus. (PMID: 40066998)

Was this response helpful?

Referenced Articles

A ribosome-interacting jumbophage protein associates with the phage nucleus to facilitate efficient propagation.

PMID: 39992933DOI: 10.1371/journal.ppat.1012936
View Full Text
Show Abstract

The dual functions of the GTPase BipA in ribosome assembly and surface structure biogenesis in Salmonella enterica serovar Typhimurium.

PMID: 40203049DOI: 10.1371/journal.ppat.1013047
View Full Text
Show Abstract

Molecular basis for the phosphorylation of bacterial tyrosine kinase Wzc.

PMID: 40210632DOI: 10.1038/s41467-025-58693-7
View Full Text
Show Abstract

References

McClain, M. T., Constantine, F. J., Henao, R., Liu, Y., Tsalik, E. L., Burke, T. W., Steinbrink, J. M., Petzold, E., Nicholson, B. P., Rolfe, R., Kraft, B. D., Kelly, M. S., Saban, D. R., Yu, C., Shen, X., Ko, E. M., Sempowski, G. D., Denny, T. N., Ginsburg, G. S., & Woods, C. W. (2021). Dysregulated transcriptional responses to SARS-CoV-2 in the periphery. *Nat Commun* **12**, 1079. https://doi.org/10.1038/s41467-021-21289-y

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*, 550. https://doi.org/10.1186/s13059-014-0550-8