# Hands-on: Gene set enrichment analysis and overrepresentation analysis

1. **EnrichR**
2. **g:Profiler**



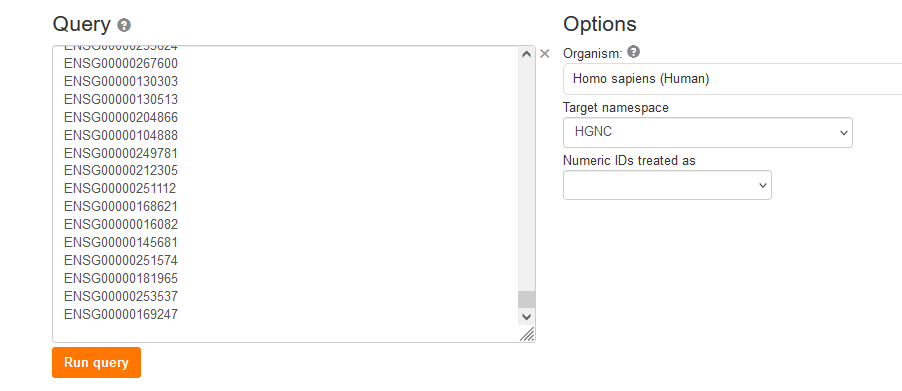
EnrichR is a webtool that allows functional analysis of gene lists. It offers a huge variety of gene sets for functional analysis ranging from Ontologies (Gene Ontology, GO), pathways (WP, REACTOME, KEGG), diseases, drugs, cell types, transcriptional regulation (ENCODE) and more.

1. Go to the **Enrichr** website: <https://amp.pharm.mssm.edu/Enrichr/>
2. To explore available gene set libraries, go to the **Libraries** tab to browse.
3. You can also browse libraries by term, for example “cancer”. From the results, you can download individual gene sets in gmt format.

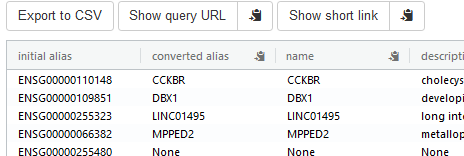
**Q1: Are there gene-sets you already recognize?**

### Gene Ontology analysis

1. Open the file in Excel with upregulated genes, **SigUP.csv**, and copy the full list of Ensembl gene identifiers (without header!).
2. Unfortunately, EnrichR only takes HGNC symbols – so we need to do a quick-and-dirty identifier mapping using g:Convert webtool.
   1. Open g:Convert website
   2. Paste the Ensembl IDs in the query field
   3. Select HGNC in the Target namespace

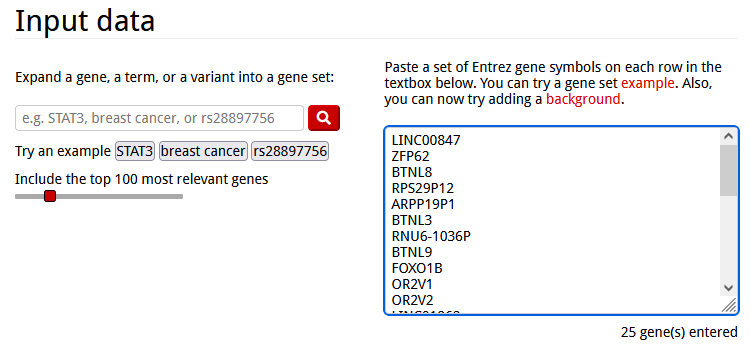


* 1. Hit “Query”
  2. Click the small “copy column to clipboard” symbol of the converted aliases.



1. Go back to EnrichR website
2. On the **Analyze** tab, paste the list of gene symbols in the input box.

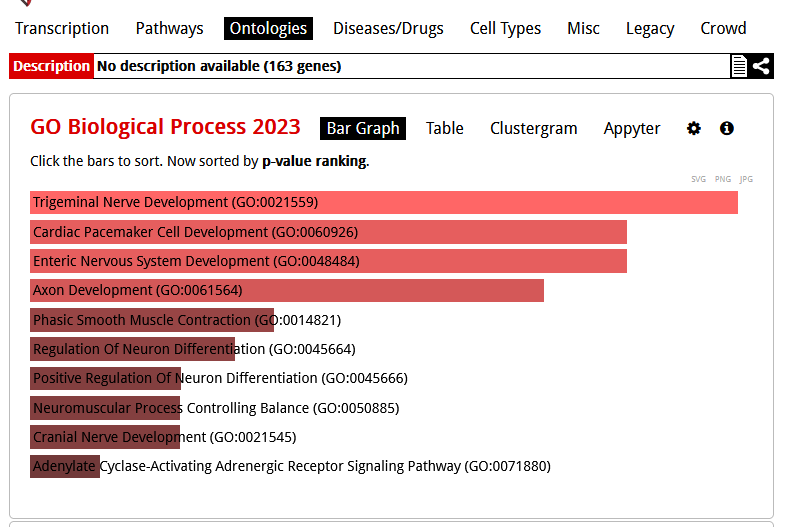
*Note:* There is an option to contribute your gene list, which makes it available to others. If you select this option, you will be prompted to assign a name and provide a brief description of the list.



1. Click **Submit.**
2. When the results load, go to **Ontologies -> GO Biological Process**.



1. By default, the terms are sorted by p-value ranking. Clicking on any of the bars representing terms re-sorts the bar graph by the different score.



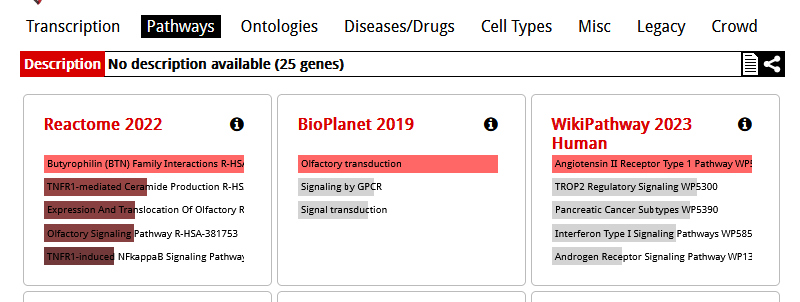
1. To save the bar chart, click either of the **SVG** / **PNG** / **JPG** buttons to the right just above the bar graph.
2. It is also possible to change the color scheme by clicking the **Settings** icon at the top right.
3. Switch to the **Table** view at the top. To export the table, click the **Export entries to table** link at the bottom.
4. Open the downloaded table. Each row contains statistics for one GO term, and a list of changed genes for that term.

**Q2: What are the top10 GO biological processes enriched with upregulated genes?**

**Q3: Do you expect these processes to be altered in your dataset?**

### WikiPathways

1. At the top of the page at Enrichr, click **Pathways** to see results for pathway sets.



1. Click the heading for **WikiPathway 2021 Human.**
2. Again, export the bar graph and table.
3. To share or archive the results, a permanent link is available by clicking the link icon at the top of the page.

**Q4: What are the top10 WikiPathways pathways enriched with upregulated genes?**

**Q5: Do you expect these pathways to be altered if you compare up-regulated genes in FB neurons from COS vs. control?**

Repeat the Gene Ontology and WikiPathway enrichment analysis with the downregulated genes.

1. Repeat for the down-regulated list, **SigDOWN.csv**. And answer the questions accordingly

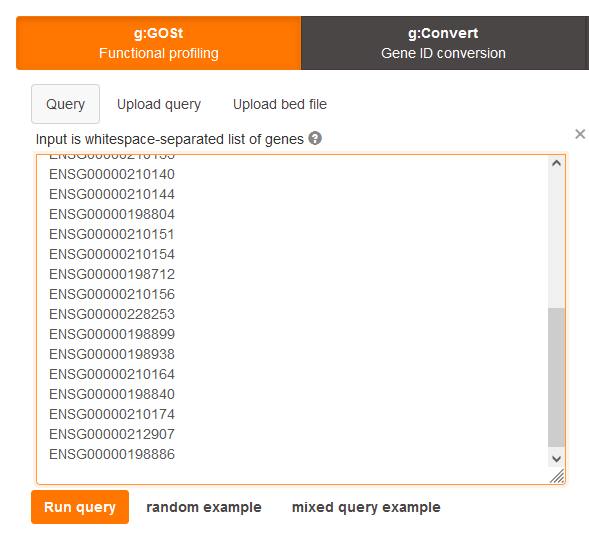
### Literature:

1. Documentation and help <https://maayanlab.cloud/Enrichr/help#basics>
2. Chen EY, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 128(14). doi: 10.1186/1471-2105-14-128.
3. Day-Richter J, et al. Gene Ontology OBO-Edit Working Group, Lewis S. OBO-Edit–an ontology editor for biologists. Bioinformatics. Aug 2007;23(16):2198-200. doi: 10.1093/bioinformatics/btm112

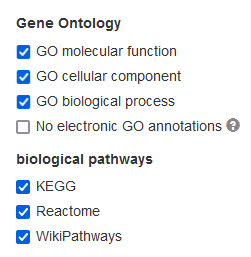


g:Profiler is a public, free webtool for analysis and manipulation of gene lists. It offers different functions that include gene set enrichment analysis (g:GOSt) but also identifier conversion (g:Convert), orthology search (g:Orth) and SNP ID to gene name conversion (g:SNPense). In this tutorial we are using the g:GOSt function to analyse the function of our differentially expressed genes.

1. Go to the website <https://biit.cs.ut.ee/gprofiler/gost>
2. Open your file SigUP.csv with the significantly upregulated genes. Copy the column (without header!) with the Ensembl identifiers the query field. In contrast to EnrichR, g:Profiler accepts different gene identifiers.



1. Options and advanced options: automatic settings should be ok for this example. Check if the species is correct (Homo sapiens). If using numeric IDs make sure to select the correct ID system.
2. Data sources: Select here which gene sets you would like to see in the results. For functional analysis I would recommend Gene Ontology and the biological pathways.



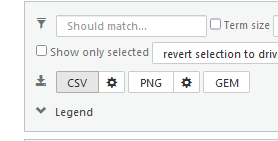
1. Press “Run query” (you will have to scroll down to see the results)
2. Interpretation of the results





The overview will give you a figure of the distribution of significant terms for the different gene sets (looks nice, but actually not very helpful). Below are the top significant overrepresented terms (more helpful).

Go to “Detailed results” – there you can see for each gene set, which terms (GO) or pathways (biological pathways) are overrepresented. You can export the file to CSV using the download function.



**Q6: What are the top10 GO biological processes enriched with upregulated genes?**

**Q7: Are they different to the results of EnrichR?**

1. Repeat with the list of downregulated genes

**Q8: What are the top10 GO biological processes enriched with downregulated genes?**

**Q9: Are they different to the results of EnrichR?**

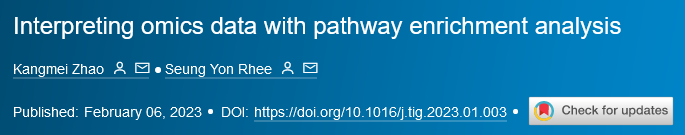
## R package gprofiler2

There is a R package available for g:Profiler (gprofiler2) that allows to perform functional analysis (and also the other functions like ID conversion) in R.

## Literature

* Liis Kolberg, Uku Raudvere, Ivan Kuzmin, Priit Adler, Jaak Vilo, Hedi Peterson: g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update) Nucleic Acids Research, May 2023; doi:10.1093/nar/gkad347
* Documentation on website: <https://biit.cs.ut.ee/gprofiler/page/docs>

## About comparability of the results of different analysis tools



## How to compare different GO analysis results



<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04828-2>