

# Introduction to Pathway Analysis and Visualization

## Exercise 1 instructions - Functional Enrichment Analysis

The example data files represent lists of differentially expressed genes for ORA and GSEA analysis:

### Example Data

**lung.expr.up.txt:**  $\log_2FC \Rightarrow 2$  AND  $\text{adj.P.Value} \leq 0.05$

**lung.expr.dn.txt:**  $\log_2FC \leq -2$  AND  $\text{adj.P.Value} \leq 0.05$

**lung.expr.entrez.gsea.txt:** All genes and associated rank

The data is a lung cancer dataset from [TCGA](#), comparing lung cancer biopsies versus normal tissue.

### Enrichr

1. Go to the **Enrichr** website: <https://amp.pharm.mssm.edu/Enrichr/>
2. To explore available 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.

### Gene Ontology

1. Open the file with upregulated genes, **lung.expr.up.txt**, and copy the full list of symbols.
2. On the **Analyze** tab, paste the list of gene symbols in the input box.
3. *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.
4. Click **Submit**.
5. When the results load, go to **Ontologies -> GO Biological Process**.
6. 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.
7. To save the bar chart, click either of the **SVG / PNG / JPG** buttons to the right just above the bar graph.
8. It is also possible to change the color scheme by clicking the **Settings** icon at the top right.
9. Switch to the **Table** view at the top. To export the table, click the **Export entries to table** link at the bottom.

10. Open the downloaded table. Each row contains statistics for one GO term, and a list of changed genes for that term.

### WikiPathways

1. At the top of the page at Enrichr, click **Pathways** to see results for pathway sets.
2. Click the heading for **WikiPathway 2021 Human**.
3. Again, we can export the bar graph and table.
4. To share or archive the results, a permanent link is available by clicking the link icon at the top of the page.
5. *Optional*: Repeat for the down-regulated list, **lung.expr.dn.txt**.

### WebGestalt

#### GSEA Pathways

1. Go to the **WebGestalt** website: <http://www.webgestalt.org/>
  2. Select **Homo sapiens** as species, **GSEA** and for Functional Database select **pathway** and then **WikiPathways**.
  3. Under Gene List, select **Gene symbol**.
  4. Open the **lung.expr.entrez.gsea.txt** file and copy the contents.
  5. In the WebGestalt **Gene List** box, paste the gsea example data (**lung.expr.entrez.gsea.txt**).
  6. Skip the **Advanced Parameters** for now and click **Submit**. The results may take several minutes to load.
  7. The top of the results page includes a job summary and a link to download the results (top right).
  8. Results are displayed as a bar chart by default. Right-clicking on the bar chart lets you download in either PNG or SVG format. Clicking on the bars updates the pathway-specific display at the bottom of the page. The **Table** and **Volcano plot** views have similar interactivity.
- 

### Extra exercise

#### g:Profiler

1. Go to the g:Profiler website: <https://biit.cs.ut.ee/gprofiler/gost>. By default, the **Functional Profiling** tab will be selected.
2. Paste the list of upregulated genes in the input box on the left.
3. Make sure the **Homo sapiens** is selected under **Organism**.

4. Expand the Data sources panel to view the default options, and uncheck any databases you don't want.
5. Click **Run query**. Results will load below the analysis interface.
6. If a gene is ambiguously mapped to Ensembl you will get a yellow box asking you to choose which one to use. There is also an option to **Select the Ensembl ID with the most GO annotations**, both for each individual case, and a button to set this as the default for any ambiguous mapping. Once you fix the ambiguous mappings you can rerun the query.
7. The default results display is a dot plot, with results for different resources represented in different colors. Clicking on a dot displays the details for that term/pathway.
8. To see the full results table, click **Detailed results**. This gives you an interactive display for each database where you can filter the results by term size and keyword.
9. The colors used in the matrix are described under **Legend**. For example, for the GO results, each gene association with a GO term is color-coded based on the evidence type from the GO ontology.
10. Results can be downloaded in PNG and table format using the buttons under the filtering interface.