

Exercises: Presenting Results

Licence

This manual is © 2016-17, Laura Biggins, Boo Virk, Christel Krueger

This manual is distributed under the creative commons Attribution-Non-Commercial-Share Alike 2.0 licence. This means that you are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution. You must give the original author credit.
- Non-Commercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

Please note that:

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Full details of this licence can be found at

<http://creativecommons.org/licenses/by-nc-sa/2.0/uk/legalcode>

Exercise 1: Checking for Artefacts

In this exercise we will check for functional enrichment using one of the web-based tools that we saw in the previous section. We will then check for any obvious artefacts using the `gene_screen` shiny app.

There should be a file named “`genelist1.txt`” in your data folder. This contains an unordered list of mouse genes. The identifiers are official gene symbols.

- Open a web browser (preferably Firefox) and navigate to one of the web-based functional enrichment tools (e.g. DAVID, PANTHER, GOrilla).
- Use the file “`genelist1.txt`” as the query genes, either by uploading the file directly to the website or by opening the file and pasting in the list of genes.

If you use GOrilla, in Step 2 of the input options where it asks you to choose the running mode, you'll have to select the 2nd option, “two unranked lists of genes” (target and background lists).

Use the file “`background_mouse_genes_seqmonk.txt`” for the background list. Upload the file rather than pasting in the genes as GOrilla copes better with the number of genes like this (you can still paste in the target genes).

- In Firefox, use the URL below to navigate to the `gene_screen` app – googling will not work, you'll have to type it in.
Internet Explorer does not support the app, but other modern web browsers should be fine.
http://www.bioinformatics.babraham.ac.uk/shiny/gene_screen/.
- Paste the list of genes from `genelist1.txt` into the query genes section.
- Scroll down through the graphs and see whether any look suspicious.

Exercise 2: Plotting Results

Open the file `table1.xlsx` in Excel. This is the result of a functional enrichment test from DAVID.

- Clean up the dataset ready for plotting – choose how to remove any categories that appear to be redundant. Remember that there are not strict rules for how to do this, but make sure you are consistent.
- Using the tidied dataset, produce 3 separate plots (using Excel or any other tool of choice) showing the following:
 - No of genes
 - Enrichment score
 - P-value

Exercise 3: Reducing Redundancy

The file `genelist2.txt` contains another unordered list of mouse genes. If this list of genes is run through GOrilla or DAVID's functional annotation chart, we see a lot of redundant categories in the results.

Instead of attempting to condense the results manually this time, try some different tools for reducing redundancy. The 3 methods below each use a different approach, try all of them and see whether you find similar results.

- DAVID functional annotation clustering
This clusters categories based on the number of genes that overlap between the categories.
- GOrilla and then REVIGO (see Exercise 1a for notes on running GOrilla and remember to select the option in the advanced parameters that allows the results to be exported directly to REVIGO)
REVIGO uses a semantic similarity based algorithm to remove redundant terms.
- Panther Statistical overrepresentation test using GOslim
GOslims use a subset of gene ontology terms. Panther uses GOslims by default.