Exercises:
Running Gene Lists

Version 2017-07

Licence

This manual is © 2016-17, Simon Andrews, Boo Virk.

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

In this practical we will use free, web-based resources to analyse gene lists. In doing so, we will determine which functions are significantly overrepresented in the gene lists, and whether the tools we are using show similar results.

Exercise 1: Functional analysis of Gene Lists in PANTHER

You should find a file called "Gene List 174 genes.xlsx" in the Data folder. This list contains 174 differentially expressed genes from human. We are going to run this gene list to obtain functional information.

- Navigate to PANTHER in a web browser (http://www.pantherdb.org/, or search for 'pantherdb').
- Copy and paste the Ensembl ID's for the genes in our gene list into the search box.
- Select 'ID List' for List type.
- Select the appropriate organism (our data is from humans).
- Ensure that the 'Functional classification viewed in gene list' is the selected analysis before submitting.
- Above the table that appears, see how many of the ID's that we searched for were recognised by PANTHER.
- Try searching again using the Gene Symbols from our gene list.
- Continue further steps with the set of ID's that had the highest number of genes recognised.
- Use the Customize Gene list option to add a column showing the PANTHER GO slim biological process annotation into the results table.
- Click the pie chart option at the top of the page to show a pie chart for biological processes (BP pie chart). What are the largest segments?
- Go back to the PANTHER homepage and search as before, this time perform a Statistical overrepresentation test analysis. Don't take the defaults – uncheck the box which says use defaults.
- Try looking at different list subsets from both the GO Slims (cut down GO) and full Gene Ontology.
- What are the most significant processes for this gene list?

Exercise 2: Functional analysis of Gene Lists in DAVID

- Navigate to DAVID in a web browser (https://david.ncifcrf.gov/, or search for 'david bioinformatics').
- Search for the Ensembl ID's from our gene list ("Gene List 174 genes.xlsx"), using the functional annotation tool.
- Select the appropriate options for Identifier and List type before submitting the search.
- Notice that DAVID recognises most of the Ensembl ID's imported.
- Use the Pathway annotation source to include the Reactome pathway annotations into the analysis.
- Look at the Functional annotation table. This table shows us all the annotations for specific genes.
- In DAVID use the Functional annotation clustering tool to determine the most significant clusters and the annotation terms within them.
- Compare these annotation terms to the top functions PANTHER gave us. How similar are they?
- Change the Classification stringency to 'Highest' and rerun the functional analysis clustering (Rerun using options).
- Have the most significant clusters changed?

Exercise 3: Functional analysis of Gene Lists in GOrilla

- Navigate to GOrilla in a web browser (http://cbl-gorilla.cs.technion.ac.il/, or search for 'gorilla gene enrichment')
- Select 'two unranked lists of genes' as the running mode.
- Copy and paste the list of gene symbols from our gene list ("Gene List 174 genes.xlsx") into the target set.
- In the course data folder, you should find a file called 'Human background list', select this file for the background set.
- Ensure that the ontology selected is 'process'.
- Click 'Search for enriched terms' to submit the analysis.
- Browse the results – how similar are they to those results obtained by DAVID and PANTHER?

Exercise 4: Functional analysis of Gene Lists in g:GOST (tool in g:Profiler)

- Navigate to g:Profiler in a web browser (http://biit.cs.ut.ee/gprofiler/, or search for 'g profiler').
- Copy and paste the list of gene symbols from our gene list ("Gene List 174 genes.xlsx") into the query box.
- Tick the option for 'no electronic GO annotations', then submit by clicking the 'g:Profile!' button.
- Browse the resulting enriched terms, do they concur with the other tools we have used?
- Save and open the output as an Excel file.

Exercise 5: Functional analysis in EnrichR

- Navigate to http://amp.pharm.mssm.edu/Enrichr/
- Paste the gene list into the search box and press submit
- Look through the different categories to see what the tool finds as significant. Try out some of the different views which are available for different categories of hits.
- Can you spot any common biological themes coming out between different sections and the previous results you saw in the other searches?