Gene Ontology (GO) Enrichment

In a previous exercise you identified all gene in Cryptosporidium species that contain a with the YXXΦ receptor signal motif:

http://cryptodb.org/cryptodb/im.do?s=928309b4c1b9ef3f

Look at the results for C. muris. Are you able to identify types of proteins that are enriched in the list just by looking? How confident are you about your conclusion? To help, you can also click on the graph icon next to the product description column to generate a word cloud of all the words in the product descriptions of your results. Is this statistical?

Enrichment analysis allows you to look for overrepresented features in your results. GO terms provide a structured vocabulary for product descriptions. At EuPathDB we get GO terms from the official annotation and by transferring GO terms associated with InterPro domains (using InterPro scan).

Here are some useful links with additional information:

http://geneontology.org/page/documentation

http://geneontology.org/page/go-enrichment-analysis
To run a GO enrichment analysis, click on the “Analyze Results” tab then click on the GO enrichment option:

Next select “the cellular component” option and submit your analysis:

Examine your results. Do they make sense?
Some more info on Fischer’s exact test:
http://udel.edu/~mcdonald/statfishers.html

Some more info about Odds rations:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/

Some more info on false discovery rates and P value correction:
http://brainder.org/2011/09/05/fdr-corrected-fdr-adjusted-p-values/