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).
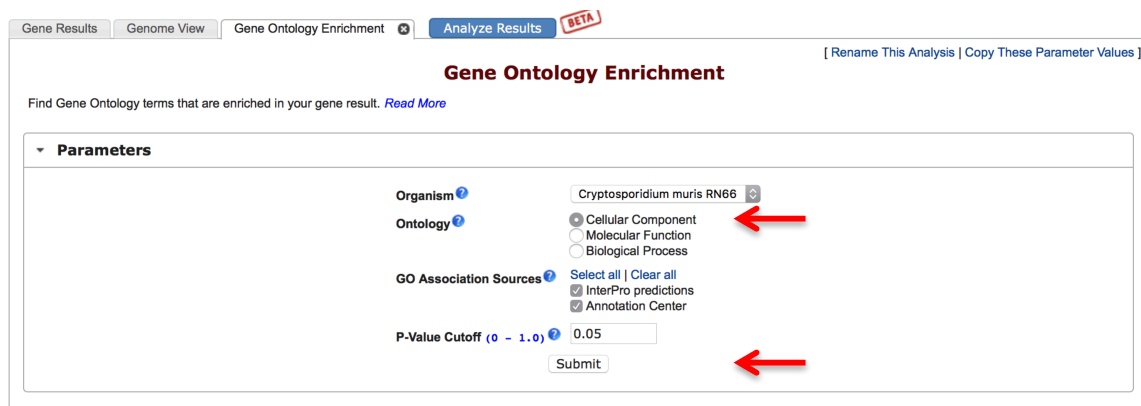Here are some useful links with additional information:

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:
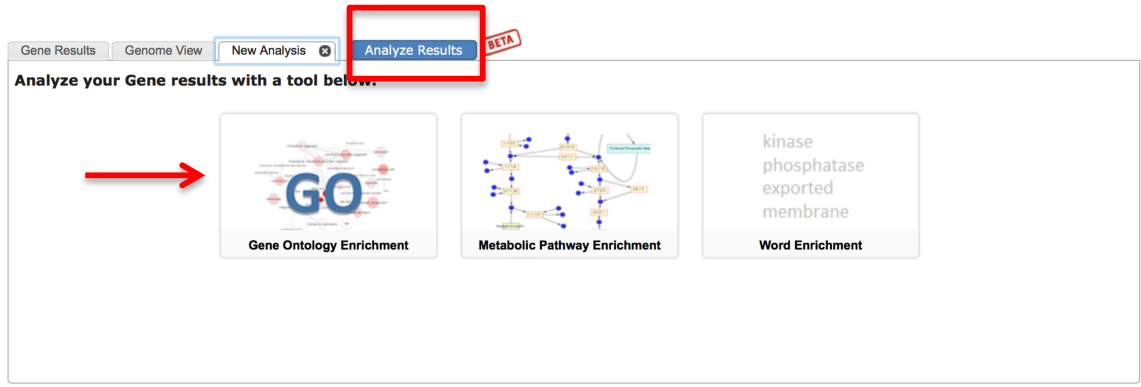
Next select "the cellular component" option and submit your analysis:

Examine your results. Do they make sense?

Analysis Results:

Got a total of 11 results   Filter :

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd Genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0016020	membrane	271	25	9.2	4.37	8.8	1.76e-11	1.94e-10	1.94e-10
GO:0031224	intrinsic to membrane	121	18	14.9	7.05	11.23	3.82e-11	2.10e-10	4.20e-10
GO:0016021	integral to membrane	119	17	14.3	6.77	10.4	2.90e-10	1.06e-9	3.19e-9
GO:0044425	membrane part	177	18	10.2	4.82	7.46	1.18e-8	3.25e-8	1.30e-7
GO:0005575	cellular_component	870	30	3.4	1.63	2.99	4.14e-4	9.10e-4	4.55e-3
GO:0005789	endoplasmic reticulum membrane	13	3	23.1	10.93	11.66	3.83e-3	6.02e-3	4.21e-2
GO:0042175	nuclear outer membrane-endoplasmic reticulum membrane network	13	3	23.1	10.93	11.66	3.83e-3	6.02e-3	4.21e-2
GO:0044432	endoplasmic reticulum part	14	3	21.4	10.15	10.82	4.58e-3	6.30e-3	5.04e-2
GO:0005783	endoplasmic reticulum	18	3	16.7	7.89	8.4	8.46e-3	1.03e-2	9.31e-2
GO:0012505	endomembrane system	33	3	9.1	4.31	4.55	3.66e-2	4.03e-2	4.03e-1
GO:0031090	organelle membrane	37	3	8.1	3.84	4.05	4.79e-2	4.79e-2	5.27e-1

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/