Gene Ontology (GO) Enrichment

In a previous exercise you identified all gene in *Cryptosporidium* species that contain a with the $YXX\Phi$ 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?

| | Genes f egy: Pro | | | somal/lys | sosomal sorti | ng signal Y | 'XXphi a | nd TM domain | Add 110 Genes to Basket Download 110 Genes | | | | |
|-----------|---------------------|---------------------|------------|---------------------------------|---------------------|-------------|-----------|-------------------------|---|--|--|--|--|
| = 🝸 (| Click on a i | number in | this table | to limit/filte | er your results | | | | | | | | |
| | | | Ар | icomplexa | 1 | Chr | romerida | | | | | | |
| All | Ortholog | | yptosporio | lium | Gregarina | Chromera | Vitre | lla | | | | | |
| Results | s Groups | C.homini | | C.parvum | G.niphandrodes | | V.brassic | aformis | | | | | |
| | | TU502 | RN66 | lowa II | Unknown strain | CCMP2878 | CCMP: | 3155 | | | | | |
| 282 | 205 | 81 | 110 | 91 | 0 | 0 | 0 | | | | | | |
| | Results | Genome 5 Next La | | | Results | | | | Add Columns | | | | |
| | 🗘 Gene | ID | 🗘 Ger | nomic Loc | ation 🥝 | | | Product Description | ption 🎱 🚛 | | | | |
| | CMU_037 | 790 | DS989 | 726: 41,920 | - 43,329 (+) | | | hypothetical protein, c | onserved | | | | |
| | CMU_0380 | 090 | DS989 | 726: 122,94 | 6 - 124,004 (-) | | | hypothetical protein, c | onserved | | | | |
| | CMU_038 | 180 | DS989 | 726: 142,78 | 5 - 144,212 (+) | | | hypothetical protein, | Word Cloud X | | | | |
| | CMU_0382 | 270 | DS989 | 726: 163,29 | 2 - 164,866 (+) | | | transmembrane amir | Word Cloud Data | | | | |
| | CMU_0383 | 340 | DS989 | 726: 184,85 | 9 - 189,688 (-) | | | ABC transporter familia | Filter words by rank: 1 to 50 | | | | |
| | CMU_0390 | 090 | DS989 | 726: 401,35 | 2 - 406,928 (-) | | | hypothetical protein | Sort by: Rank A-Z Use slider or enter numbers to adjust filter | | | | |
| | CMU_039 | 120 | DS989 | 726: 411,042 | 2 - 412,847 (-) | | | hypothetical protein, | Mouse over a word to see its occurence in the column | | | | |
| | CMU_039 | 130 | DS989 | 726: 413,14 | 3 - 416,793 (-) | | | hypothetical protein, | | | | | |
| | CMU_039 | 540 | DS989 | DS989726: 506,535 - 506,903 (+) | | | | hypothetical protein | dehydrogenase othic diptophale draj domalin endornembrane ethanolarinephosphotransferase eukaryosis family teronecin finger glycosyl hand | | | | |
| | CMU_039 | 560 | DS989 | DS989726: 512,239 - 513,519 (+) | | | | glycosyl transferase | | | | | |
| | CMU_0413 | 340 | DS989 | 726: 923,37 | 5 - 926,695 (-) | | | hypothetical protein, | ITYPOUTICUIOUI ii maintenance mannosyltransferase mpv17 oxidoreductase oxo patched peptidase phosphaldidite phosphaldidirositol | | | | |
| | CMU_0416 | 610 | DS989 | 726: 983,29 | 2 - 984,672 (+) | | | oxidoreductase, shor | pamit pro2 protein putative subuit transferate transmethrane transporter type zinc | | | | |
| | CMU_0417 | 760 | DS989 | 726: 1,005,5 | 531 - 1,006,922 (-) |) | | hypothetical protein, | plasmid pmp22 PIOLGIII putative subunit transferase transmembrane transport transporter type zinc | | | | |
| | CMU_0418 | 310 | DS989 | 726: 1,014,5 | 553 - 1,015,645 (-) |) | | hypothetical protein, | | | | | |
| | CMU_0422 | 220 | DS989 | 726: 1,102,2 | 277 - 1,103,155 (+ | ·) | | Mpv17 / PMP22 family | y protein | | | | |
| Display a | menu 0422 | 250 | DS9897 | 726: 1,109,4 | 77 - 1,112,080 (+ |) | | zinc finger, C3HC4 typ | pe domain-containing protein | | | | |

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:

| Analyze your Gene | ne View New Analysis Analyze Res | ults | |
|-------------------|----------------------------------|------------------------------|---|
| | → GO | | kinase phosphatase exported membrane |
| | Gene Ontology Enrichment | Metabolic Pathway Enrichment | Word Enrichment |

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

| Gene Results Genome View Gene Ontology Enrichment S Analyze Results | BETA |
|---|--|
| | [Rename This Analysis Copy These Parameter Values |
| Gene Ont | ology Enrichment |
| Find Gene Ontology terms that are enriched in your gene result. Read More | |
| • Parameters | |
| Organism 🕗 | Cryptosporidium muris RN66 |
| Ontology 🚱 | Cellular Component Molecular Function Biological Process |
| GO Association Sources | Select all Clear all InterPro predictions Anotation Center |
| P-Value Cutoff (0 - 1.0 | 0.05 |
| | Submit |
| | |

Examine your results. Do they make sense?

| ene Results Genome View Gene Ontology E | inrichment 🙁 🗛 | nalyze Results | | | anomo Th | | Copy These F | anamatan Value |
|--|---|--|---|--------------------|----------|--------------|---------------------------------|----------------|
| | 6 | ene Ontolog | y Enrichment | [" | tename m | lis Analysis | Copy These P | arameter value |
| | | selle Ontolog | y Emiciment | | | | | |
| Find Gene Ontology terms that are enriched in your ge | ene result. Read More | | | | | | | |
| > Parameters | | | | | | | | |
| Parameters | | | | | | | | |
| | | | | | | | | |
| Analysis Results: | | | | This ana | | may be los | Download An at if you change | your gene resu |
| ot a total of 11 results Filter : | | | | | Тс | save this a | analysis result, p | lease downloa |
| \$ GO ID \$ GO Term | Genes in the the bkgd with this term | Genes in your cresult with this term | Percent of bkgd Genes in your result | Fold enrichment | | P- value | Senjamini | Sonferror |
| 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 |
| | 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/