

Gene Ontology

1. Go to the Gene Ontology (GO) website: <http://www.geneontology.org/>
From the "Documentation" menu, follow the links to:
"Introduction" and "Ontology".

What is Gene Ontology?

What are the three ontologies?

2. From the "Tools" menu, follow the link to AmiGO 2, then search the GO database for annotations associated with the *white* gene of *Drosophila melanogaster* (also indexed by the FlyBase number FB:FBgn0003996).

How many GO terms are associated with this gene?

What biological processes is it involved in?

What is its molecular function?

What cell components is it found in?

Hint: you can use the "Ontology (aspect)" filter to look only at Process (P), Function (F), or Component(C)

3. Search the GO database for the human (*Homo sapiens*) BRCA1 gene (indexed by the UniProt ID UniProtKB:P38398).

How many GO terms are associated with this gene?

What biological processes is it involved in?

What cell components is it found in?

What is its molecular function?

4. The DAVID website provides several web-based tools for the functional analysis of gene sets: <https://david.ncifcrf.gov/>

Go to the "Gene Functional Classification" tool:

<https://david.ncifcrf.gov/gene2gene.jsp>

At the tab on the left, you can upload a list of gene names and choose a background (reference list). From the course website, you can download a list called "EurOverexpressed.txt". This is a list of *Drosophila melanogaster* genes that were found to have higher expression in European flies than in African flies in a microarray experiment. Upload this list to David (be sure to select that the identifiers are "FLYBASE_GENE_ID" in step 2 and that this is a "Gene List" in step 3). Your list of genes might be present in the genomes of multiple species, thus you should choose *Drosophila melanogaster* for the "list" and "background". Finally, you should choose "highest" classification stringency and re-run the analysis.

Which clusters of genes (that is, genes with common functions) show the greatest enrichment in the list?

What are the common functions of these clusters? (Click on the red "T" symbol for each cluster to see the enriched GO terms and the associated p-values.)

5. A comparison of human, chimpanzee, and mouse orthologs identified genes with an accelerated rate of evolution in the human lineage. These are candidates for genes that have been positively selected during human evolution. A list of these genes can be downloaded from the course website as the file “PosHum.txt”.

Use the DAVID Gene Functional Classification tool to see which clusters show enrichment for positively selected genes in humans. Be sure to use the correct gene list and background. Note that the “PosHum.txt” list provides identifiers that are “OFFICIAL_GENE_SYMBOL”. Re-run the analysis using “highest” classification stringency.