Tutorial 10 – GO Term Enrichment Analysis

MICB405 – BIOINFORMATICS – 2021W-T1

19 NOVEMBER 2021

AXEL HAUDUC

Gene Ontology (GO)

Framework/database of functions of genes

- “Protein synthesis”
- “Nucleic acid metabolism”

Gene ontology *annotations* link specific genes to 1 or more GO terms present in the database

Each GO term will further have connections to other GO terms and other information (such as experimental evidence)

Categories of GO terms

Cellular component

- The *physical part* of the cell linked to the gene

Molecular function

- Elemental *functions* linked to the gene, such as binding of substrates

Biological process

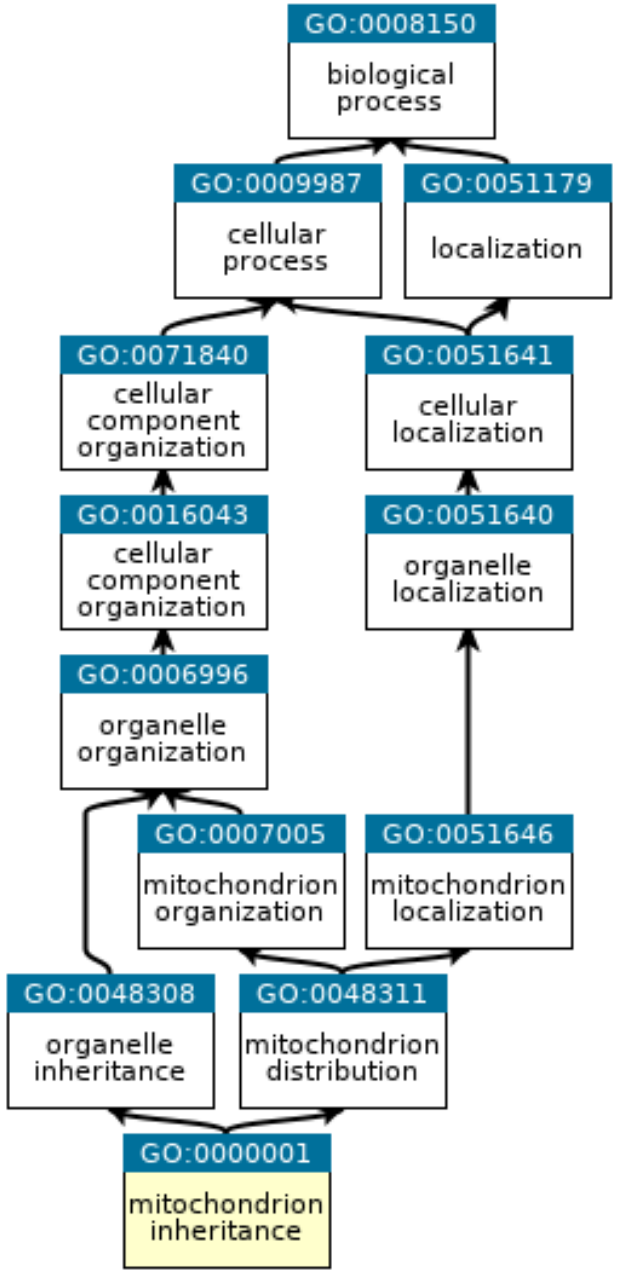
- Sequences of *operations* or *events* linked to the gene, such as cell growth

GO database is extensive

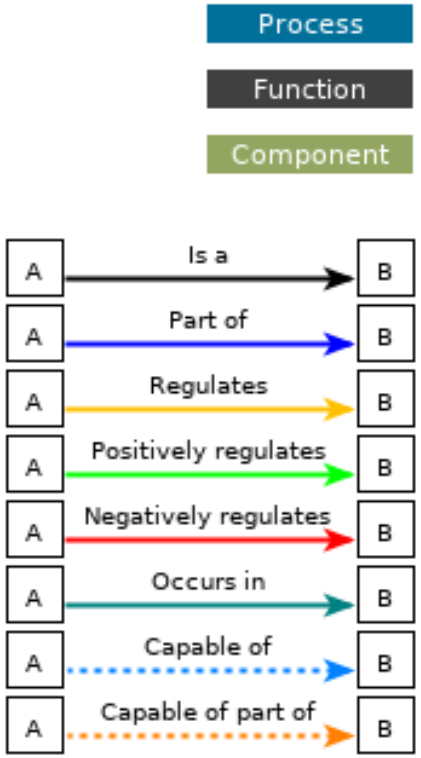
GO contains ~45,000 terms

- 29,698 biological processes
- 11,147 molecular functions
- 4201 cellular components

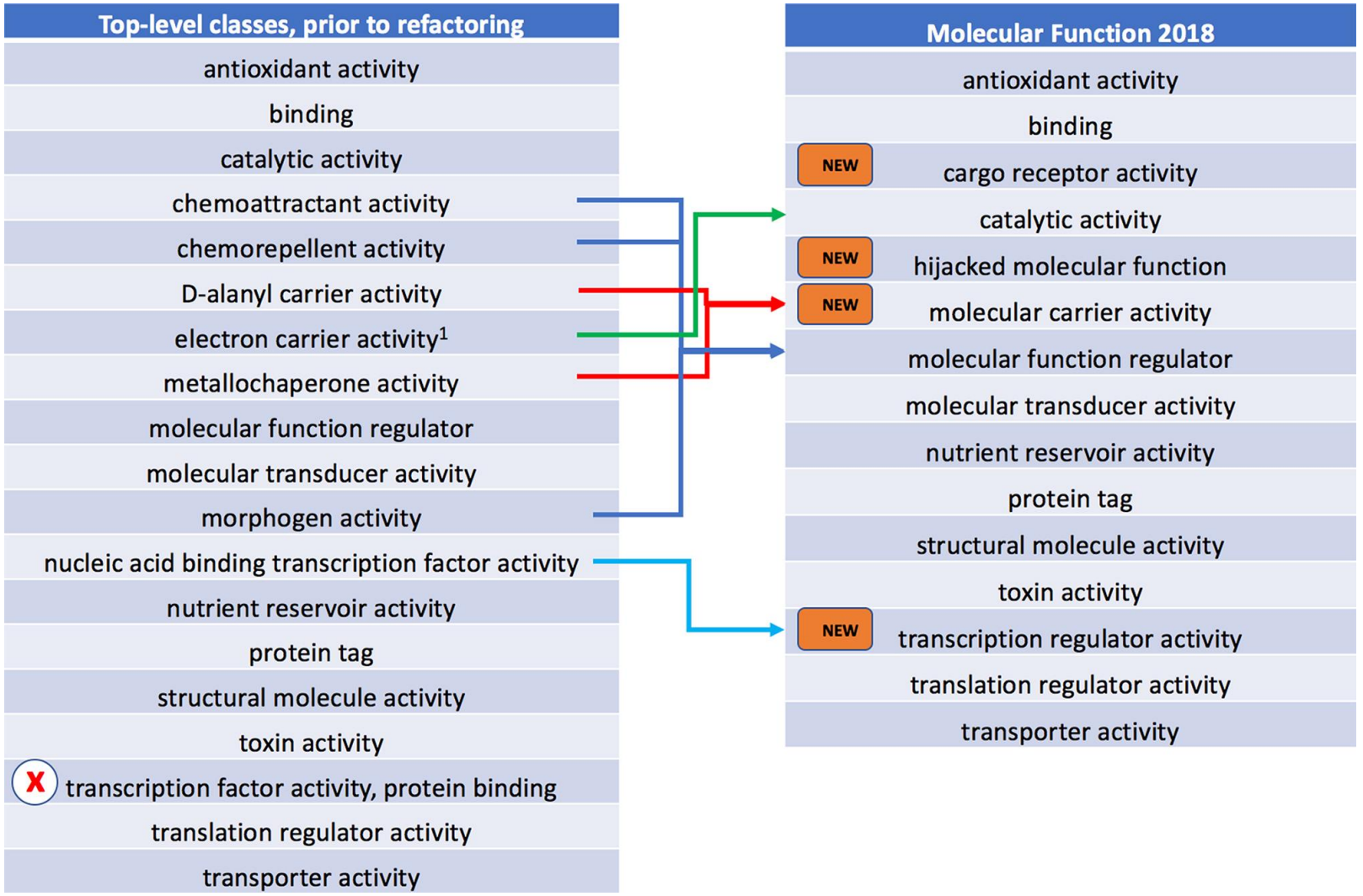
Linked by ~134,000 relationships



QuickGO - https://www.ebi.ac.uk/QuickGO



GO terms as a graph



GO terms get constantly revised as new data emerges

Potential applications of GO term enrichment analysis are numerous

Query information linked to a single gene

- "What functionality is *Oct4* associated with?"

Get lists of genes associated with a function/process/component

- "What genes are linked to glucose metabolism?"

Get overrepresented functionalities linked to a gene list

- "Overall, what do my DESeq2 differentially expressed genes do?"
- This is what today's tutorial will focus on