

Gene Ontology

Nociones Básicas de Bioinformática y Genómica
(*Máster en Bioinformática, Universidad de Valencia*)

David Montaner

2015-05-18

Ontology in Computer Sciences

Ontology: a **model** for describing the world.

Represents *knowledge* as a *hierarchy* of *concepts* within a *domain*.

Describes a shared vocabulary to denote:

- types of concepts
- properties of the concepts
- interrelationships of the concepts

See [Wikipedia article](#) for further descriptions, and observe that . . .
There is also generally an expectation that the features of the model in an ontology should closely resemble the real world.

Gene Ontology

Started in 1998 for:

- *Drosophila melanogaster*
- *Mus musculus*
- *Saccharomyces cerevisiae*.

Aims to standardize the representation of gene and gene product attributes:

- Develop a controlled vocabulary of gene and gene product attributes.
- Annotate genes and gene products.
- Tools.

Cite: Ashburner M, et al. (May 2000) Nat. Genet

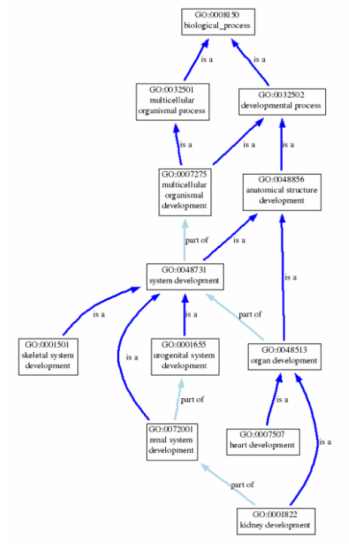
Gene Ontology

The ontology covers three domains¹:

- **cellular component**: parts of a cell or extra cellular environment.
- **molecular function**: activity of genes or gene products at molecular level.
- **biological process**: process of a living organism accomplished by one or more molecular functions (different levels: cells, tissues, organs. . .)

¹formally there are tree ontologies

Structure: directed acyclic graph (DAG)



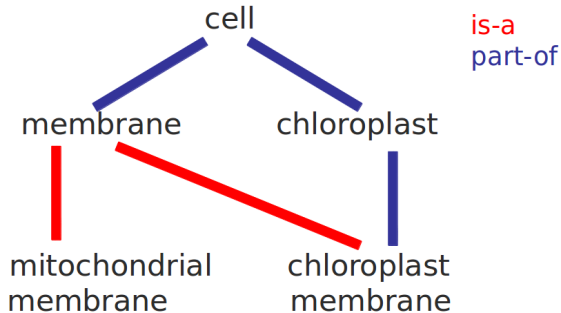
Gene Ontology Relationships

Relationships between terms of the ontology:

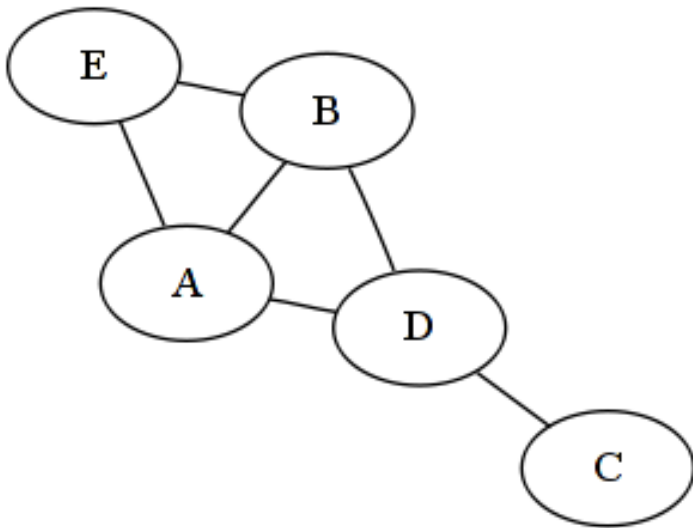
- is a
- is part of
- regulates

See a complete explanation of the [GO relations](#) here.

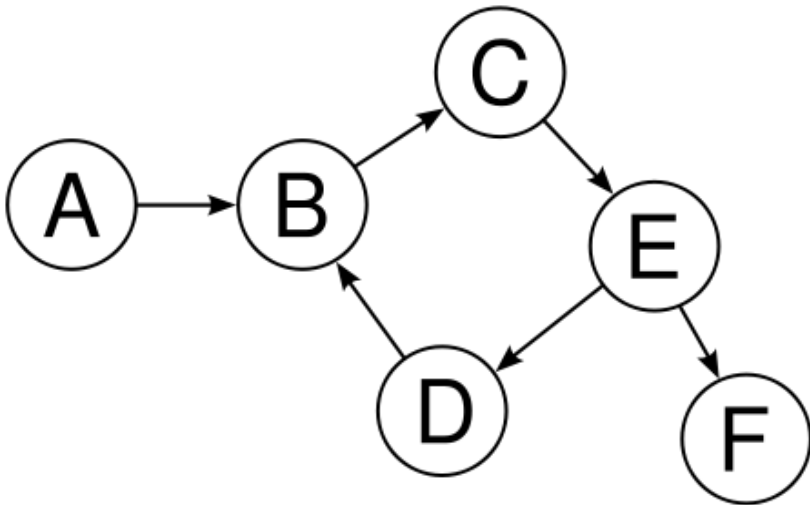
Gene Ontology Relationships



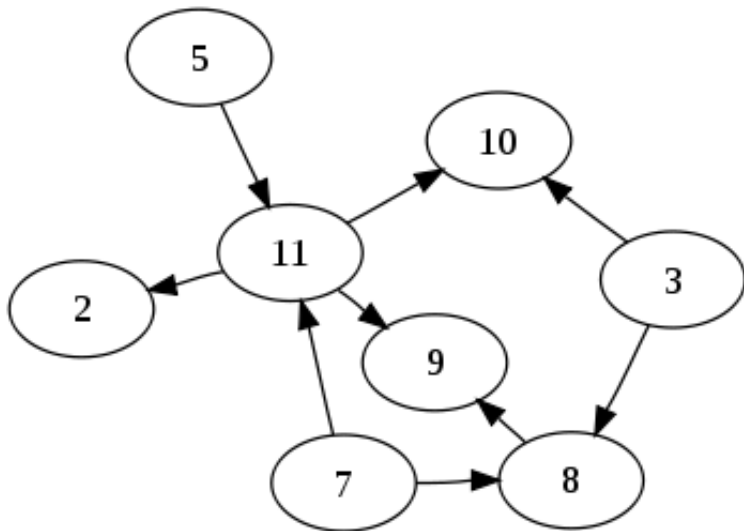
Undirected graph



Cyclic graph (*with cycles*)



DAG



Gene Annotation

When genes are annotated to a GO term an *Evidence Code* is assigned to the *membership relationship*:

- Experimental Evidence Codes
- Computational Analysis Evidence Codes
- Author Statement Evidence Codes
- Curator Statement Evidence Codes
- Automatically-assigned Evidence Codes
- Obsolete Evidence Codes

See full evidence code description here:

<http://www.geneontology.org/GO.evidence.shtml>

Some Remarks

- Gene Ontology terms are defined *a priori* to gene membership.
 - A gene may be annotated under many GO terms.
 - A gene *inherits* all *parent* annotations
 - Genes may belong to GO terms *via* many different *evidences*
 - Just normal (sane) functions are described. Functions which are unique to mutants or diseases are not (e.g. oncogenesis)
 - The GO vocabulary is designed to be species-neutral and applicable to prokaryotes, eukaryotes, single and multicellular organisms.
-
- Gene - GO association data may be downloaded straight from the GO consortium. . .
 - but usually is easier to get that information from the different *genomic projects pages* or from Ensembl or other collection webs

Links

- Gene Ontology Consortium: web of the consortium
- GO documentation: An Introduction to the Gene Ontology
- GO data: An Introduction to the Gene Ontology
- *Official* search tool: <http://amigo.geneontology.org>
query, browse and visualize ontologies and gene product
annotation data.