

The GO Consortium

The following organisations are actively involved in the GO project.

Berkeley Bioinformatics and Ontology Project (BBOP)

British Heart Foundation – University College London (BHF-UCL)

DictyBase (*Dictyostelium discoideum*)

EcoliWiki (*Escherichia coli*)

FlyBase

GeneDB (including databases for several protozoan parasites)

Gramene (grains)

Institute for Genome Sciences (IGS)

J Craig Venter Institute (bacterial species)

Mouse Genome Informatics (MGI)

Plant-Associated Microbe Gene Ontology (PAMGO) consortium

PomBase (fission yeast)

Rat Genome Database (RGD)

Reactome at Cold Spring Harbor Laboratory and EMBL-EBI

Saccharomyces Genome Database (SGD)

The Arabidopsis Information Resource (TAIR)

UniProt and InterPro

WormBase (*Caenorhabditis elegans*)

Zebrafish Information Network (ZFIN)

Need Help?

www.geneontology.org

Documentation: www.geneontology.org/GO.doc.shtml

Support: gohelp@genome.stanford.edu

Open Biomedical Ontologies: <http://obofoundry.org>



Services | Research | Training | Industry

The Gene Ontology project

www.geneontology.org

To non-scientists, the language used by biologists can seem confusing; strewn with acronyms, jargon, and common words repurposed for scientific usage. Within the scientific community, this confusion is often compounded as terms are given organism-specific meanings – for example, ‘budding’ means one thing for plants and another for yeast – and different phrases are employed to refer to the same phenomena. Even subtle differences in the usage of terms and phrases can hamper communication between scientists, and can make automated data exchange prohibitively difficult.

The Gene Ontology (GO) project is a major initiative to produce an unambiguous, universal language to describe biological phenomena across all species, from *Escherichia coli* to elephants. It is used to capture information about the function and localisation of gene products.

The aims of the GO project are:

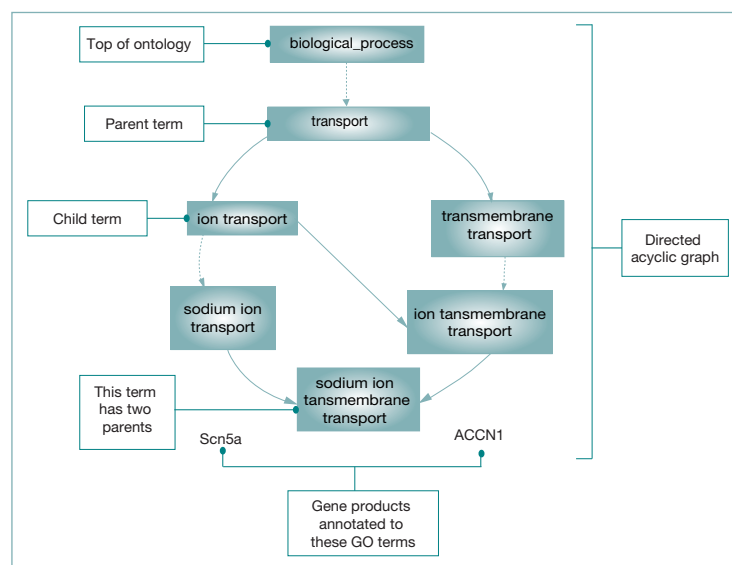
- to maintain and further develop the GO ontology, the vocabulary of terms for representing gene product attributes;
- to annotate genes and gene products, and to assimilate and disseminate annotation data;
- to provide tools to facilitate access to all aspects of the data provided by the GO project.

At EMBL-EBI we are involved in two aspects of the GO project: developing the ontology in collaboration with other members of the GO Consortium, and annotating several of the EBI’s databases with GO terms. This factsheet describes what the ontology is and how you can access and use ontology data. For more information on GO annotation, see the GOA factsheet.

The content of GO

The GO ontology is a structured, controlled vocabulary; a set of defined terms (words or phrases) representing biological phenomena and entities, with relationships between the terms. The ontology is split into three sub-ontologies that cover different gene product properties.

- Molecular function – defines the activities performed by individual gene products at the molecular level; for example, the catalysis of reactions or binding to certain compounds or protein complexes.



GO terms are organised in directed acyclic graphs (DAGs) – hierarchical structures in which any ‘child’ (more specialised term) can have many ‘parents’ (less specialised terms). For example, the biological process term ‘sodium ion transmembrane transport’ has two parents. Any gene product that is annotated to this term is automatically annotated to both ‘sodium ion transport’ and ‘ion transmembrane transport’. Some terms and relationships have been omitted for clarity.

- Biological process – terms represent a series of molecular events or functions with a defined beginning and end. Processes span all levels of granularity, covering subcellular metabolic pathways, tissue- and organ-level chains of events, all the way up to organism-level processes such as locomotion and interactions with other organisms.
- Cellular component – describes the parts of the cell – subcellular structures, locations and macromolecular complexes – and its extracellular environment in which a gene product may be localised. Examples include cytoplasm, ribosome and translation release factor complex.

GO terms do not describe gene products themselves, rather the activities they perform and the locations where they are found. A description of a gene product is built from the combinations of components, processes and functions with which a gene product is annotated. For example, you won't find 'cytochrome' in the molecular function ontology, but you will find the functions that cytochromes perform, such as electron carrier activity.

What can I do with GO?

GO helps you give biological meaning to your data. It is used for purposes as diverse as:

- integrating proteomic information from different organisms;
- assigning functions to protein domains;
- finding functional similarities in genes that are overexpressed or underexpressed in diseases and as we age;
- predicting the likelihood that a particular gene is involved in causing disease;
- analysing groups of genes that are co-expressed during development;
- developing automated ways of deriving information about gene function from the literature;
- verifying models of genetic, metabolic and product interaction networks.

For references, see www.geneontology.org/cgi-bin/biblio.cgi

How do I use it?

GO project data is freely available in several formats. Users can search, browse and analyse ontology and annotation data using tools produced by the GO Consortium, such as AmiGO (<http://amigo.geneontology.org>) and OBO-Edit (<http://oboedit.org>). A full list of GO- and GO-related tools can be found at www.geneontology.org/GO.tools.shtml. Ontology, annotation and other GO data can be downloaded from the website in a number of configurations and formats: www.geneontology.org/GO.downloads.shtml

The GO Consortium supports the creation of ontologies beyond the scope of the GO project, and numerous consortium members are involved in these efforts. The tools used by GO curators to create and edit the gene ontology are freely available from the GO website and are employed by curators of other ontologies.

Contributions, questions and feedback

We welcome feedback from all users. Questions or comments on all aspects of the project can be directed to the GO helpdesk, gohelp@geneontology.org or submitted online <http://geneontology.org/help.shtml>. We also have a number of ontology, annotation and tool-related trackers on SourceForge, <http://geneontology.sf.net> ●

Further reading

The Gene Ontology Consortium. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.* 38, D31-D335

The Gene Ontology Consortium. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.* 36, D440-D444

The Gene Ontology Consortium. (2001) Creating the Gene Ontology resource: design and implementation. *Genome Res.* 11, 1425-1433

The Gene Ontology Consortium. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25-29

About GO

The Gene Ontology Consortium is supported by a P41 grant from the US National Human Genome Research Institute [HG 002273].