



**Jane Lomax**

GO Curation Coordinator  
 PhD in parasite population  
 genetics, University of  
 Cambridge, 2002.  
 At EMBL-EBI since 2002.

## The Gene Ontology Editorial Office

### DESCRIPTION OF SERVICES

The Gene Ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene-product attributes across all species. The aims of the Gene Ontology project are threefold: to maintain and further develop its ontologies of gene and gene product attributes; to annotate genes and gene products, and assimilate and disseminate annotation data; and to provide tools to facilitate access to all aspects of the data provided by the Gene Ontology project. The GO ontologies cover three key biological domains that are shared by all species: the cellular component (the parts of a cell or its extracellular environment); molecular function (the elemental activities of a gene product at the molecular level, e.g. binding or catalysis); and biological process (operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units, e.g. cells, tissues, organs, and organisms).

Groups participating in the GO Consortium include major model organism databases and other bioinformatics resource centres. At EMBL-EBI, the GO Editorial Office plays a key role in managing the distributed task of developing and maintaining the GO vocabularies. We contribute to a number of other GO project efforts, including web presence, software testing, user support and education.

### SUMMARY OF PROGRESS

- Added logical definitions (a.k.a. cross-products) to GO for the first time, enabling the development of a tool for automatic term addition;
- Progressed toward alignment with and creation of cross-products to an external chemical ontology;
- Developed ontology content in the areas of signalling, transcription and kidney development;
- Established relationships between the biological process and molecular function ontologies.

### MAJOR ACHIEVEMENTS

The Gene Ontology is dynamic: existing terms and relationships are augmented, refined, and reorganised as biological knowledge advances. Major improvements have been made over the lifetime of the GO project in several areas of the ontology, usually in consultation with experts in relevant subject areas. Table 1 shows the size (as of November 2010) of each of the four ontologies maintained by the GO Consortium.

Table 1. Status of the GO vocabularies as of November 2010

Total GO Terms	32935
Molecular Function Terms	8893
Cellular Component Terms	2771
Biological Process Terms	19819

Significant changes introduced to GO in 2010 affect both biological and logical aspects of the ontologies: logical definitions (also known as 'cross-products') have been added to GO for the first time, which has allowed for the development of a tool for automatic term addition. We have made much progress toward aligning with and creating cross-products to an external chemical ontology. We have also developed ontology content in the areas of signalling, transcription and kidney development. In addition, relationships have been established between the biological process and molecular function ontologies.

We made considerable progress toward creating cross-products for GO terms (Mungall et al., 2010). These definitions help improve computability and support more sophisticated tool development. Our work has concentrated on 'internal' cross-products, i.e. those that define GO terms by referring to other GO terms. The first set of cross-products – between regulatory processes and regulated processes or functions – were added to the GO file in early 2010. Subsequently, two further sets have been added: biological processes involved in other biological processes, and cellular components that are part of other cellular components.

As a result of these changes, we developed a tool – TermGenie – that allows users to add new GO terms that conform to a cross-product template directly to the ontologies. Terms are automatically placed correctly within the ontology, and textual definitions and synonyms are automatically generated. This tool reduces the workload for ontology editors and helps reduce human error in the ontologies.

We generated cross-products to externally maintained ontologies that intersect with GO. To this end, we are active members of the OBO Foundry (Smith et al., 2007), a collaboration to establish a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. Earlier this year, GO became one of the founder sets of OBO Foundry ontologies.

The biggest effort over the past year went into aligning GO with the Chemical Entities of Biological Interest (ChEBI) ontology, with the aim of generating cross-products between GO and ChEBI. This involved a two-day meeting with the ChEBI ontology developers in September 2010 to reconcile some of the critical differences between the two ontologies. We hope the first ChEBI cross-products will be added to GO early in 2011.

GO has traditionally comprised three orthogonal ontologies, but we have been working to add relationships between these ontologies to enrich the biological representation. In 2010 we have added 'part\_of' relationships between the molecular-function and biological-process ontologies. For example, we have made many transporter functions 'part\_of' their corresponding transport process.

2010 has seen major improvements to the biological content of several areas of the ontologies; transcription and transcription factors; signalling; and kidney development. The changes in these areas were developed in collaboration with biological experts, often culminating in a face-to-face meeting such as the kidney development meeting, held at EMBL-EBI in January 2010.

## FUTURE PLANS

The GO Editorial Office will continue to work closely with the rest of the GO Consortium and with biological experts to ensure that the ontologies are comprehensive, logically rigorous and biologically accurate. Improvements begun or continued in 2010 on signalling, kidney development and other topics will therefore continue, and we intend to start developing terms in the area of neurobiology. We will continue adding further sets of cross-products to GO, allowing us to improve TermGenie so that more routine term addition can be done automatically. This will free up editing time for more complex, biologically detailed work. These cross-product sets will include links to ChEBI, the first set of external cross-products to be added to GO. We also hope to start making cross-products to the Cell Ontology on 2011. Additional links between the biological process and molecular function ontologies will be created using new process-specific function terms.

## SELECTED REFERENCES

Mungall, C.J., et al. (2010) Cross-product extensions of the Gene Ontology. *J. Biomed. Inform.* (in press). Published online 10 February; DOI: 10.1016/j.jbi.2010.02.002.

Smith, B., et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251-1255.

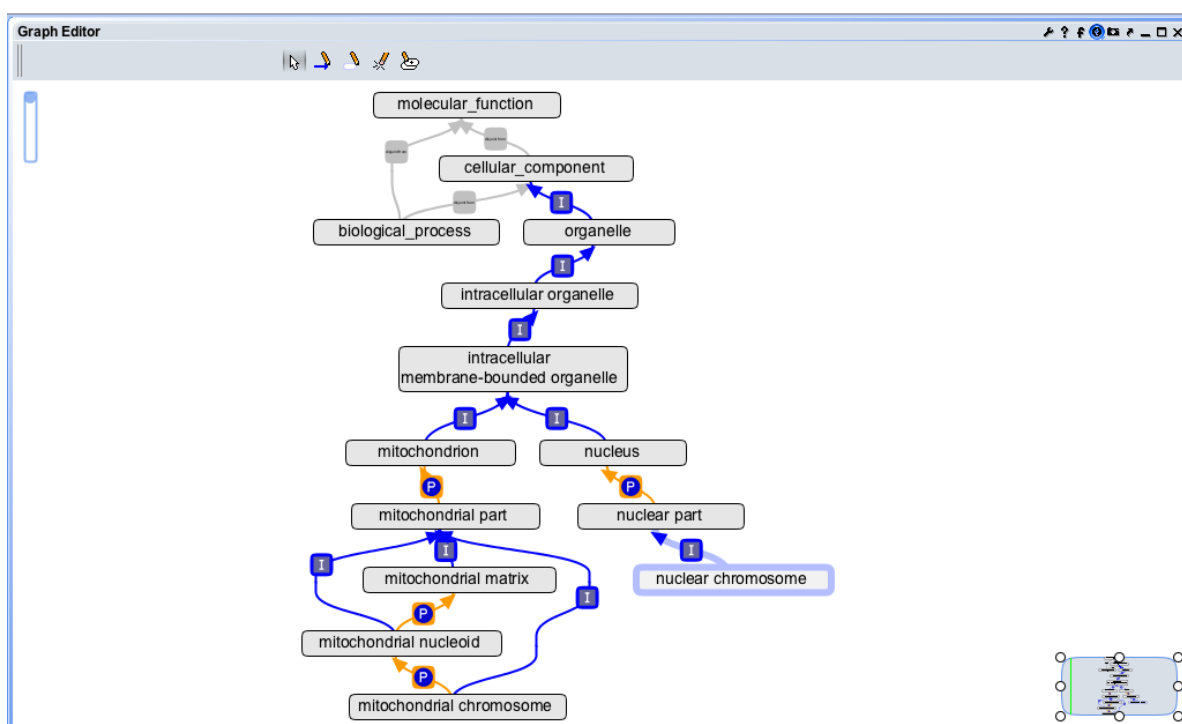


Figure. Structure of the Gene Ontology, shown in OBO Edit.