

FROM RESOURCE TO RESEARCH: MGI AND GO

Mary E. Dolan, Joel E. Richardson, Janan T. Eppig, Martin Ringwald,
Carol J. Bult, James A. Kadin, Judith A. Blake¹

Keywords: Gene Ontology, computational genomics

The Mouse Genome Informatics (MGI) (<http://www.informatics.jax.org>) system provides a comprehensive public resource about the laboratory mouse that integrates information on sequences, maps, genes and gene families, expression, alleles and variants, strains and mutant phenotypes. The integration of such diverse data depends upon quality determinations of object identities and relationships and upon the use of defined, structured vocabularies (ontologies).

The Gene Ontology (GO) Project provides structured, controlled vocabularies in the domain of molecular biology, which have been incorporated in various bioinformatics resources to aid biological annotations. The GO project model has been extended for the development of other ontologies and has fostered standardization among model organism database systems. MGI also incorporates several different classification schemes that enable a variety of query capabilities for our users. The Mouse Anatomical Dictionary and Phenotype Classifications provide the mechanism for annotation of aspects of gene expression, QTL analysis, and incorporation of information about experimental mouse mutants.

The challenge for ontology developers is to construct fully documented, easily maintained ontologies that are accessible to the larger scientific community. Moreover, an important goal is to develop a semantic framework that not only adds meaning to the data but also, with the hierarchical structure of the vocabulary, can be used to create new information through inference. For example, the GO structure is a directed acyclic graph (DAG) in which each annotation node can have one or more child nodes and must have one or more parent nodes. Gene products are annotated at varying levels of detail based on experimental and computational evidence. Due to inference from the structure of the GO vocabulary, a gene product annotated to a finer level of detail is also annotated to any ancestor (coarser) level as well.

We have developed and adapted a number of software tools to facilitate the use of MGI resources in a GO context and take advantage of the GO structure to use this valuable annotation resource as a research tool. A MGI-GO browser provides tree views and links to annotated data sets. Thus experimentally annotated sets of genes can be analyzed by function, process or component via their GO representations. We also provide several MGI GO tools, which permit a user to explore what any set of annotated genes, for example, a cluster of up-regulated genes, have in common. The MGI GO TermFinder tool compares annotations of the gene set with overall MGI gene annotations to find statistically significant overrepresentations of GO classifications. Similarly, the MGI GO_Slim Chart tool compares the distribution of the gene set in a predefined classification scheme with the overall MGI gene distribution.

MGI is funded by grants from NIH/NHGRI, NIH/NICH and NCI. The GO project is funded by NIH/NHGRI and by the European Union RTD program.

¹ Mouse Genome Informatics, The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 USA.
E-mail: mdolan@informatics.jax.org, jer@informatics.jax.org,
jte@informatics.jax.org, ringwald@informatics.jax.org,
cjb@informatics.jax.org, jak@informatics.jax.org, jblake@informatics.jax.org