

Using Biomedical Ontologies for Data Representation and Management in the Mouse Genome Informatics (MGI) System

Li Ni,* Carol J. Bult, Jim A. Kadin, Joel E. Richardson, Martin Ringwald, Janan T. Eppig, Judith A. Blake
The Jackson Laboratory, Bar Harbor, Maine, U.S.A.

ABSTRACT

Structured vocabularies and ontologies are increasingly utilized for managing, annotating, and analyzing complex biological data. The Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>) resource contributes to and utilizes a variety of vocabularies and ontologies to robustly capture and provide biological information about the laboratory mouse and its use as a model for human disease. Here we report on the power and complexity of using biomedical ontologies for data representation and management at MGI. MGI is a highly integrated database and software system integrating mouse genetic, genomic and phenotypic information including data on gene characterization, pathways, protein classifications, sequence, gene expression, alleles, phenotypes, diseases, and comparative gene data for mouse, human, rat and other mammals.

The Gene Ontology (GO, <http://www.geneontology.org>) is the most widely used vocabulary for providing connections between proteins and their roles in the biological organization. MGI is a founding member of the GO Consortium, and actively participates in the ongoing development of GO, as well as applying GO to functional annotation of mouse gene products. MGI provide an automatically generated text report, a tabular form view and a graphical display of GO annotations in the gene function detail page. MGI also provide comparative graphs of GO annotations for mammalian orthologs. In addition, MGI hosts graphical and statistical tools exploiting the hierarchical structure of the GO to aid in visualization of annotations and the analysis of large data sets such as microarray data.

MGI has spearheaded the development of two other major structured vocabularies in support of annotation and analysis of mouse biology: the Mammalian Phenotype (MP) Ontology, a widely adopted ontological model that enables phenotype annotations to background-specific allelic genotypes at varying degrees of granularity, this structured vocabulary

allows consistent annotation of mouse genotypes with standard phenotype MGI is also the originator of the Adult Mouse Anatomy (MA), it allows us to navigate through the extensive dictionary hierarchies for the different developmental stages, to locate specific anatomical structures within those hierarchies, and to annotate an obtain the expression results associated with those structures.

MGI also incorporates Online Mendelian Inheritance in Man (OMIM) terms to make similarity assertions between mouse models and human disease; incorporates the Mouse Anatomical Dictionary for describing expression data during mouse embryonic development.

Additionally, for mouse gene products functional annotation, MGI start to use the Cell Ontology, the Mouse Embryonic and Adult Anatomy Ontologies, the Evidence Code Ontology, and PSI-mod to add additional information to a Gene Ontology annotation. These combined ontology annotations will be loaded into the database in the near future.

The structured vocabulary-based annotations assist in robust and accurate data mining when posing such complex questions in both computational and individual formats at MGI. Though the use of multiple ontologies, MGI is able to robustly represent many components of knowledge about the mouse model system. While these ontologies are independently developed, the concurrent use of them within the MGI system illuminates some challenges in the intersection of ontologies that MGI ontology developers and curators work to address. For example, the MP and the GO terminologies incorporate mouse anatomical terms from the MA. Testing and updating the MP and the GO to accurately reflect the canonical anatomy organization in the MA requires resources and attention on a regular basis.

MGI is a consortium of database resources with funding obtained from NHGRI (HG00330 and HG02273), NIH/NICHD (HD062499), and NCI (CA89713).

* To whom correspondence should be addressed: li.ni@jax.org