# Multiple Ontologies for Prioritizing and Standardizing Biological Information at the Rat Genome Database

Stan Laulederkind, Mary Shimoyama, Brad Taylor, Victoria Petri, Tim Lowry, Tom Hayman, Jennifer Smith, Rajni Nigam, Jeff De Pons, Melinda Dwinell, Simon Twigger, Diane Munzenmaier, Howard Jacob
Rat Genome Database, Bioinformatics Program, Medical College of Wisconsin

## Abstract

The Rat Genome Database (RGD) uses five ontologies to standardize targeted areas of biological information: Gene Ontology (GO), Mammalian Phenotype Ontology (MP), Pathway Ontology (PW), Disease Ontology (DO), and Behavior Ontology (BO). These ontologies are used to annotate information for genes, QTLs, and strains.

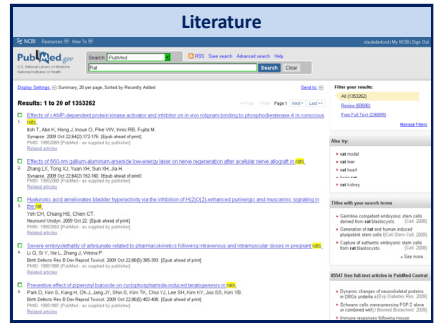
To provide broad functional annotation across the genome, RGD employs both manual and automated processes. Identifying and prioritizing literature for manual curation is a daunting process since there are more than 1.25 million rat papers in PubMed. A focus on the major areas of information represented by the ontologies (molecular function, biological process, cellular component, phenotype and pathways) is one way to limit the papers suitable for manual curation. All rat MP and BO annotations are manually curated from the rat literature while PW and DO annotations are curated from rat, mouse, and human literature.

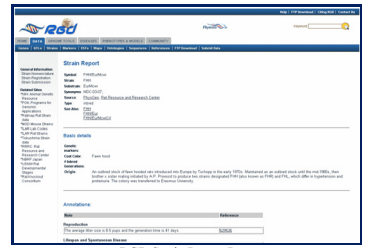
A further approach to limiting the scope of manual curation has been to target specific disease areas for in-depth curation. A gene list is developed from information at databases such as GeneCards, Phenopedia, GenAtlas, and individual disease databases. The genes are ranked according to frequency of citation and source of information. The curators review all rat literature published for each of the genes on the list and make manual annotations for GO, MP, PW, and DO. In addition, a subset of mouse and human literature providing evidence of association of the gene to the targeted diseases is also reviewed and disease annotations based on orthology are created. This process provides comprehensive biological information on the targeted genes.

Automated processes are also in place to provide broader coverage of functional information across the genome. Leveraging the manual curation efforts of other members of the Gene Ontology Consortium, RGD has created an automated pipeline to bring in mouse and human GO annotations based on experimental evidence and to assign these to rat genes based on orthology. A second pipeline imports computationally determined rat annotations from the Gene Ontology Annotation Database (GOA) at the European Bioinformatics Institute (EBI). These automated processes allow RGD to provide greater functional coverage of the genome than manual curation alone would provide.

By focusing on the areas of information represented by the ontologies, the curators can easily eliminate unrelated literature, reducing the curation burden. The ontologies also provide a standardized method for representing this information and allow RGD to leverage the annotations manually curated by other groups in order to provide a greater breadth of information to its users.

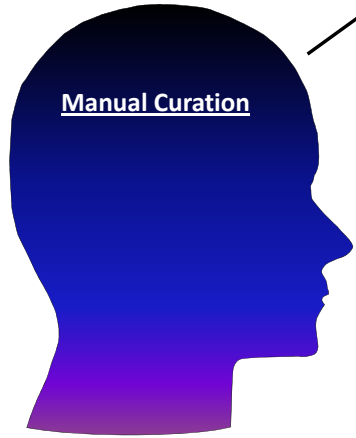
The Rat Genome Database is funded by grant HL64541 from the National Heart, Lung, and Blood Institute on behalf of the NIH.



RGD Strain Report Page

RGD QTL Report Page



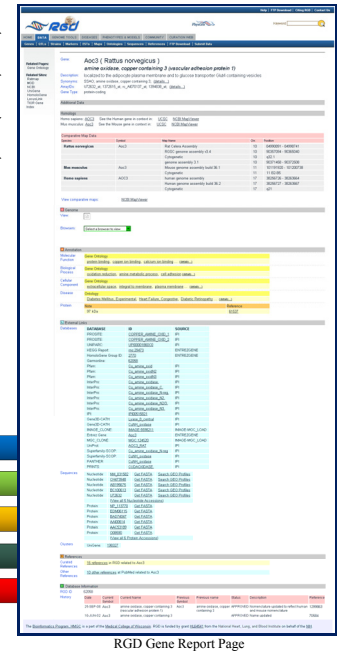

Behavior Ontology (MeSH)

Disease Ontology (MeSH)

Gene Ontology (GO)

Pathway Ontology (PW)

Phenotype Ontology (MP)



RGD Gene Report Page

**Automated Data Pipelines**

Entrez Gene (gene records)

GOA (annotations)

IPI (protein IDs)

Nomenclature (MGI, HGNC)

MGI (orthology)