Mapping of EFO terms to multiple ontologies at RGD

Stanley JF Laulederkind, GT Hayman, SJ Wang, ML Kaldunski, M Vedi, CA Park, M Tutaj, L Lamers, JR Smith, MA Tutaj,
JL De Pons, MR Dwinell, AE Kwitek

Rat Genome Database, Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, 53226, USA

Abstract

The laboratory rat, *Rattus norvegicus*, is an important model of human health and disease, and experimental findings in the rat have relevance to human physiology and disease. The Rat Genome Database (RGD, https://rgd.mcw.edu) is a model organism database that provides access to a wide variety of curated rat data including disease associations, phenotypes, pathways, molecular functions, biological processes, cellular components, and chemical interactions for genes, quantitative trait loci (QTLs), and strains. A major role of a model organism database is to provide data from laboratory animals in comparison to human data.

To be able to compare data between rat and human it is necessary to use common vocabularies. When data from different sources use different vocabularies, there must be some standard way to compare the data.

Recently RGD imported human genome-wide association study (GWAS) SNPs (designated as QTLs) from the NHGRI-EBI Catalog of human GWAS Data (https://www.ebi.ac.uk/gwas/). This dataset necessitated ontology term mapping because the GWAS SNPs are annotated with EFO (Experimental Factor Ontology), an ontology previously unused at RGD. The biocurators at RGD have mapped a human GWAS-specific subset of terms from EFO to multiple ontologies used at RGD. The term mappings have been used to transfer the EFO annotations to RDO/DO (RGD Disease Ontology/Disease Ontology), HPO (Human Phenotype Ontology), and VT (Vertebrate Trait Ontology) annotations for the human GWAS SNP QTLs.

RGD has also made available the ontological mappings via SSSOM files on the RGD download site (https://download.rgd.mcw.edu/ontology/mappings/).

Data type at RGD	Ontology used at RGD
rat phenotype (qualitative)	Mammalian Phenotype ontology (MP)
rat phenotype (quantitative)	Vertebrate Trait ontology (VT)
	Clinical Measurement Ontology (CMO)
	Measurement Method Ontology (MMO)
	Experimental Condition Ontology (XCO)
	Uberon ontology (UBERON)
	Cell ontology (CL)
human phenotype (qualitative)	Human Phenotype Ontology (HPO)
disease (qualitative)	RGD Disease Ontology (RDO) (axiomatized and extended version of the human Disease Ontology (DO)

Table 1. Ontologies used to curate phenotype and disease at the Rat Genome Database

EFO term subsets	Subset Numbers as of March 2025
total # of unique EFO terms associated with imported GWAS records	7601
total # of unique EFO/HP/MONDO hybrid terms associated with imported GWAS records	2009
# of EFO terms mapped to RDO/DO	5938
# of EFO terms mapped to CMO	472
# of EFO terms mapped to HPO	514
# of EFO terms mapped to MP	209
# of EFO terms mapped to VT	900

Table 2. Number of EFO terms from GWAS records mapped to ontologies at RGD

Figure 1. Multiple browser views showing connections among EFO, HP, RDO/DO, and VT terms. Red arrows show term-synonym relationships and term ID-XREF relationships. Dark Blue circles indicate that the selected terms have annotations (A) in RGD.

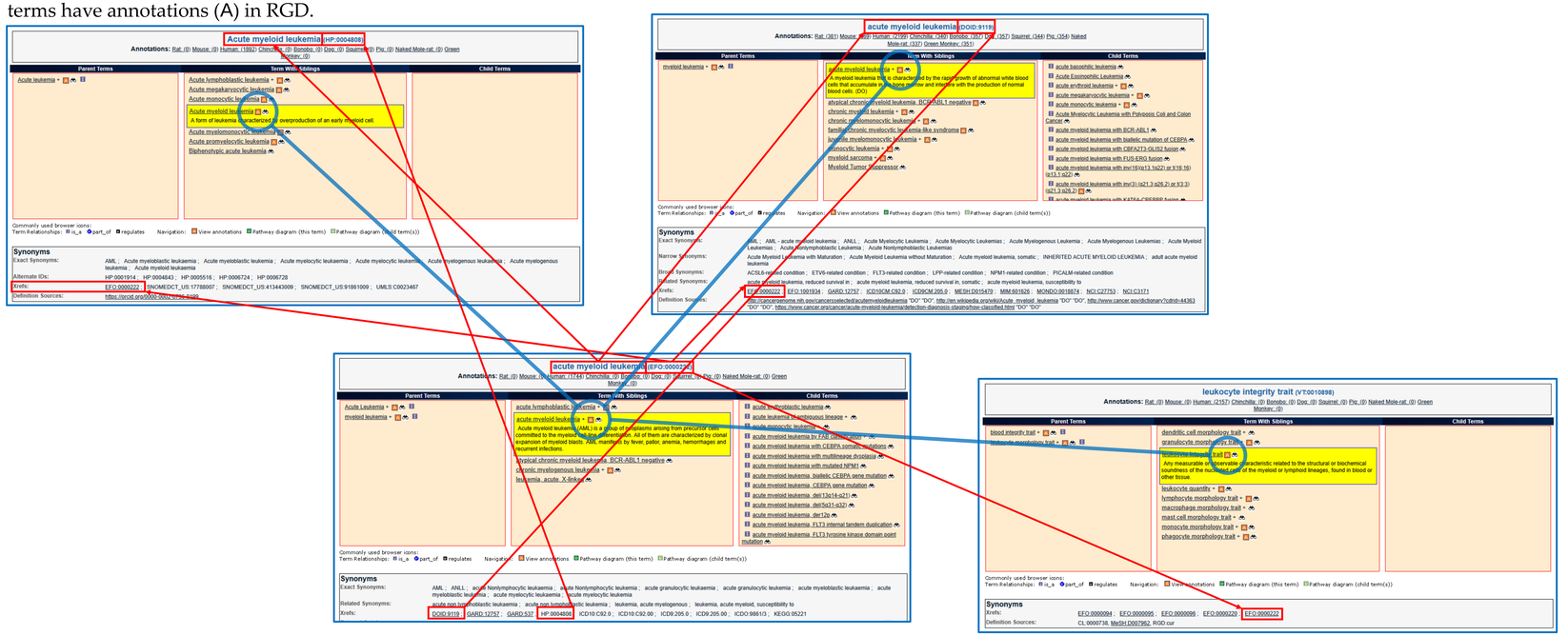
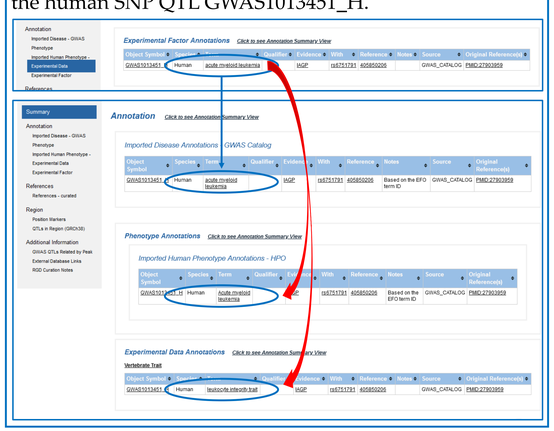


Figure 2. Based on EFO/RDO, EFO/HPO, and EFO/VT mapping, RDO, HPO, and VT annotations are assigned to the human SNP QTL GWAS1013451_H.



Summary

By mapping the Experimental Factor Ontology (EFO) to the RGD Disease Ontology (RDO), the Human Phenotype Ontology (HPO), and the Vertebrate Trait Ontology (VT), RGD has expanded the ability of database users to compare rat and other model organism data to human data and the reverse. The mapping of EFO terms to these various other ontologies will be continuously updated as more human GWAS data is imported into RGD. The ontology mappings will be a useful and necessary aide in translational research moving forward.

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