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, 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. To bridge gaps in data description from different sources, the biocurators at the Rat Genome Database (RGD) have mapped a human GWAS-specific subset of terms (The NHGRI-EBI Catalog of human genome-wide association studies/GWAS Catalog Data) (https://www.ebi.ac.uk/gwas/) from EFO (Experimental Factor Ontology) to multiple ontologies used at RGD. RGD has also made available the ontological mappings via SSSOM files on the RGD download site (https://download.rgd.mcw.edu/ontology/mappings/).

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

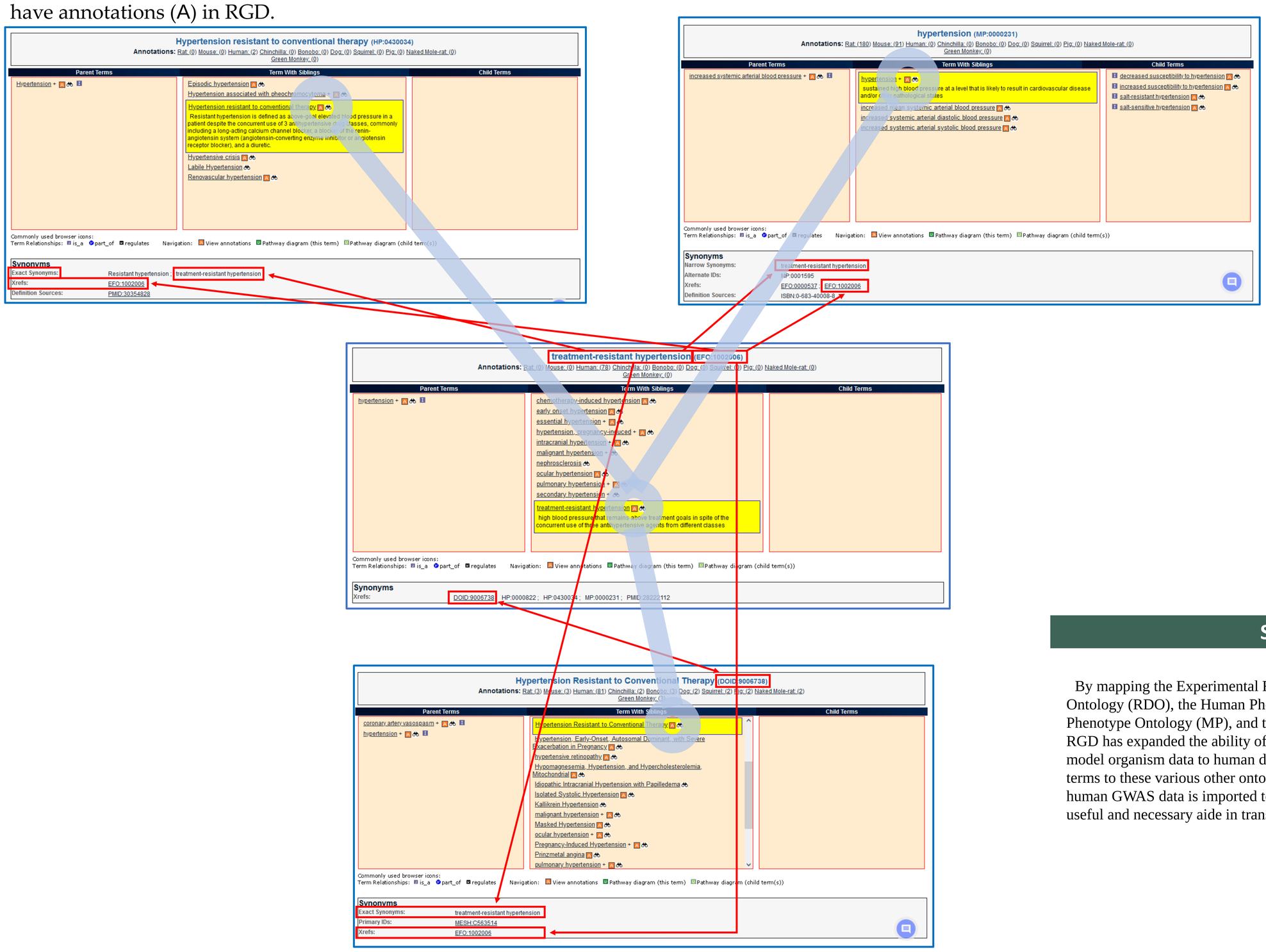
Data type at RGD	Ontology used at RGD
rat phenotype (qualitative)	Mammalian Phenotype ontology (MP)
rat phenotype (quantitative)	Vertebrate Trait ontology (VT)
	Clinical Measurement Ontology (CMO)
	Measurement Method Ontology (MMO)
	Experimental Condition Ontology (XCO)
	Uberon ontology (UBERON)
	Cell 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)

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

EFO term subsets	Subset Numbers as of July 2024
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

Figure 2. EFO terms from GWAS records mapped to ontologies at RGD

Figure 3. Multiple browser views showing connections among EFO, RDO, HP, and MP terms. Red arrows show term-synonym relationships and term ID-XREF relationships. Blue highlights indicate that the selected terms have annotations (A) in RGD.



Summary

By mapping the Experimental Factor Ontology (EFO) to the RGD Disease Ontology (RDO), the Human Phenotype Ontology (HPO), the Mammalian Phenotype Ontology (MP), and the Clinical Measurement Ontology (CMO), 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 to RGD. The ontology mappings will be a useful and necessary aide in translational research moving forward.