Integration of data from the human GWAS Catalog into the Rat Genome Database

Jennifer R Smith[1], Logan Lamers[1], Stanley JF Laulederkind[1], G Thomas Hayman[1], Shur-Jen Wang[1], Monika Tutaj[1], Mary L Kaldunski[1], Mahima Vedi[1], Wendy M Demos[1], Marek A Tutaj[1], Jyothi Thota[1], Adam C Gibson[1], Akhilanand Kundurthi[1], Varun Reddy Gollapally[1], Kent C Brodie[2], Stacy Zacher[3], Jeffrey L De Pons[1], Melinda R Dwinell[1], Anne E Kwitek[1]

[1]Rat Genome Database, Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA. [2]Clinical and Translational Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA. [3]Finance and Administration, Medical College of Wisconsin, Milwaukee, WI, USA.

The Rat Genome Database (RGD, https://rgd.mcw.edu), the premiere online resource for genetic, genomic, disease and phenotype data for rat, also integrates a substantial corpus of data for human and other mammalian species to facilitate comparative studies. For researchers interested in cross-species studies of disease- and phenotype-related genomic regions, RGD has numerous resources including standardized disease annotations across multiple species; rat, mouse and human QTL associated with disease, phenotype, trait and measurement terms; and human disease- and phenotype-associated variants imported from NCBI's ClinVar database. Now RGD has incorporated human variants and their associated phenotypes from the NHGRI-EBI GWAS Catalog.

All data available at GWAS Catalog for over 676,000 human GWAS results were loaded as is into a "holding" table in the RGD database. These data were filtered based on whether a risk allele was specified. SNPs with a risk allele different than reference at that location were loaded as variant records. In addition, GWAS QTL records with associated p-values for the listed trait(s) were created for all SNPs regardless of risk allele status.

GWAS Catalog records are associated with terms from the Experimental Factor Ontology (EFO) to denote the phenotypic trait studied. To make these data consistent with RGD's existing annotations for genes, QTL, strains, and human variants, EFO terms were matched to terms in the vocabularies already in use and applicable annotations created for both variants and QTL.

These data are discoverable at RGD by searching for terms in the applicable ontologies, using "search by position" for QTLs, and using RGD's Variant Visualizer tool to search for GWAS Catalog SNPs. In addition, lists of overlapping QTL and variants can be found on RGD's human gene, QTL and variant pages.

The incorporation of GWAS Catalog variants has substantially expanded the corpus of human genotype-to-phenotype data at RGD. Future work will include improvements to RGD's website and tools to make the data more findable and usable. Also, to support comparative genomics analyses, RGD is actively looking to integrate comparable GWAS datasets for rat.