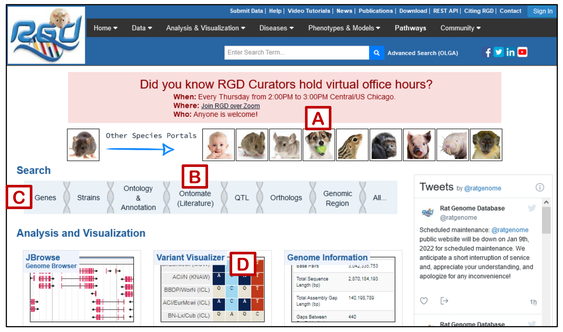
RGD: An Integrated Resource Supporting Data Discovery and Analysis for Dog

Jennifer R Smith[1], Wendy M Demos[1], G Thomas Hayman[1], Morgan Lee Hill[1], Mary L Kaldunski[1], Stanley JF Laulederkind[1], Monika Tutaj[1], Mahima Vedi[1], Shur-Jen Wang[1], Adam C Gibson[1], Logan Lamers[1], Harika S Nalabolu[1], Ketaki Thorat[1], Jyothi Thota[1], Marek A Tutaj[1], Kent Brodie[1, 2], Stacy Zacher[1, 3], Jeffrey L De Pons[1], Melinda R Dwinell[1, 4], Anne E Kwitek[1, 4]

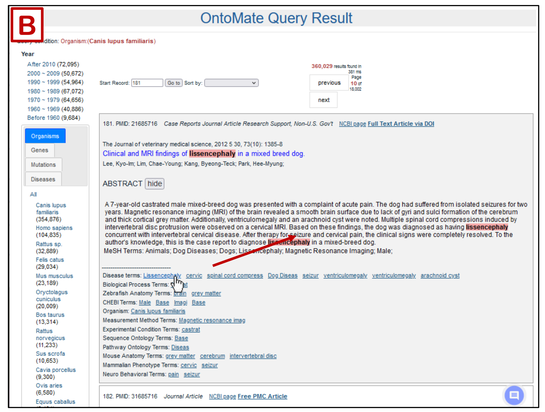
[1] Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI, USA; [2] Clinical and Translational Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA; [3] Information Services, Medical College of Wisconsin, Milwaukee, WI, USA; [4] Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA

Abstract:

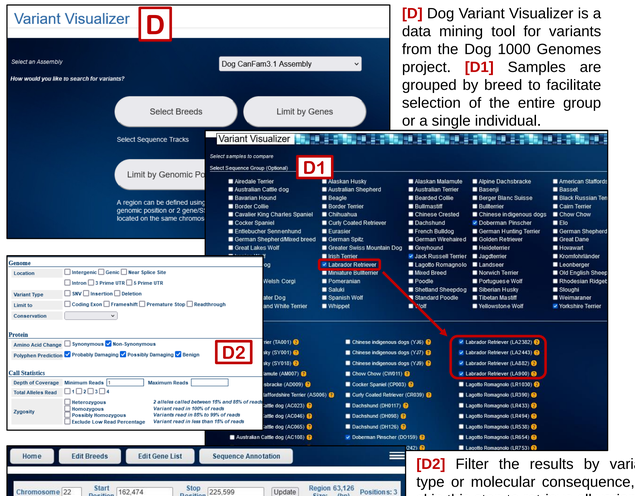
Although originally created as a resource for rat genomic, genetic, phenotypic, and disease data, RGD (Rat Genome Database, https://rgd.mcw.edu) is now a multispecies knowledgebase integrating data for 9 additional mammalian species, including dog, human, mouse, chinchilla, bonobo, pig, 13-lined ground squirrel, green monkey and naked mole-rat. Recognizing that dog (*Canis lupus familiaris*) is an important research organism, both as a model for human disease and in its own right as a companion and working animal, RGD has created the Dog Portal (https://rgd.mcw.edu/wg/species/dog/) to provide easy access to dog-centric data and RGD's suite of analysis tools to explore that data. Data for dog include gene records and orthology data relative to all other RGD species. Annotations for Disease, Pathway and Gene Ontologies are provided both from dog-specific curation, e.g., from the Online Mendelian Inheritance in Animals database, and from assignments made via orthology from the rich resource of human, rat and other species' data available at RGD. In addition, RGD's Variant Visualizer tool houses variant data across 71 breeds, as well as mixed breeds and wolf samples, from the Dog 1000 Genomes Project. Going forward, RGD is interested in partnering with dog researchers who would be willing to submit functional, disease, phenotype or variant data to RGD to further enhance the resource both for dog researchers and for those interested in comparative and translational studies.



[A] Access RGD's Dog Portal by clicking the dog icon on RGD's home page. The Dog Portal provides consolidated access to dog-related data, tools for analyzing that data, and both internal and external resources of interest to dog researchers.



[B] OntoMate is an ontology-based literature search tool developed at RGD. The link in the Dog Portal provides one-click access to dog-related abstracts from PubMed, each tagged with terms such as disease, pathway, phenotype and more. Mouse over a term to see it in the abstract. Use the facets in the left column to filter the results by date, organism, gene, mutation or disease.

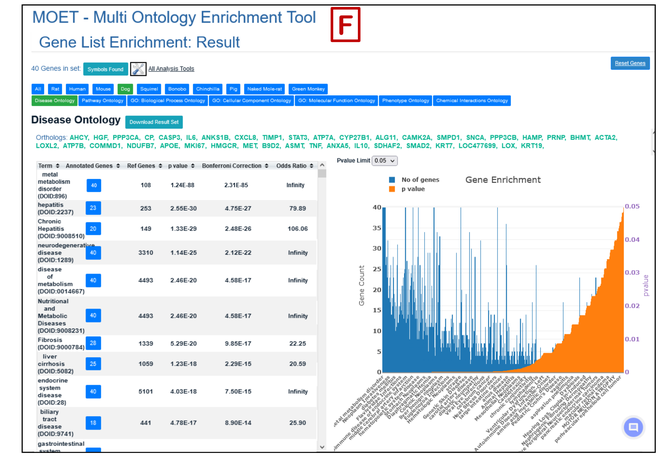


[D] Dog Variant Visualizer is a data mining tool for variants from the Dog 1000 Genomes project. [D1] Samples are grouped by breed to facilitate selection of the entire group or a single individual.

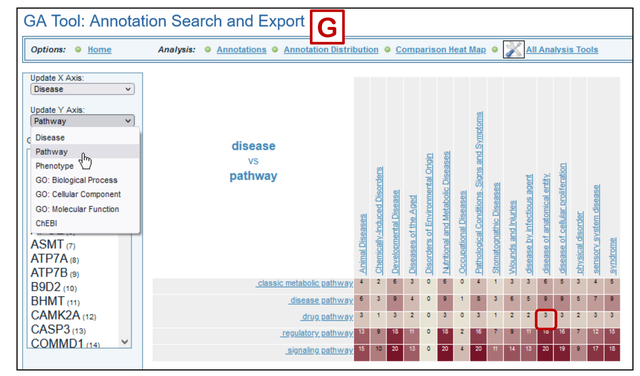
[D2] Filter the results by variant type or molecular consequence, or skip this step to retrieve all variants.

[D3] The variant result display shows the reference nucleotide at the top and the variants in each selected sample. The arrow highlights the position of the WD-associated variant.

[D4] The variant detail display gives data about the variant itself (reference and variant nucleotides, read depth, zygosity, etc) and information about the resulting amino acid change and damage predictions where available. This variant is predicted to be "possibly damaging" with respect to the displayed peptide and "probably damaging" for other ATP7B isoforms.

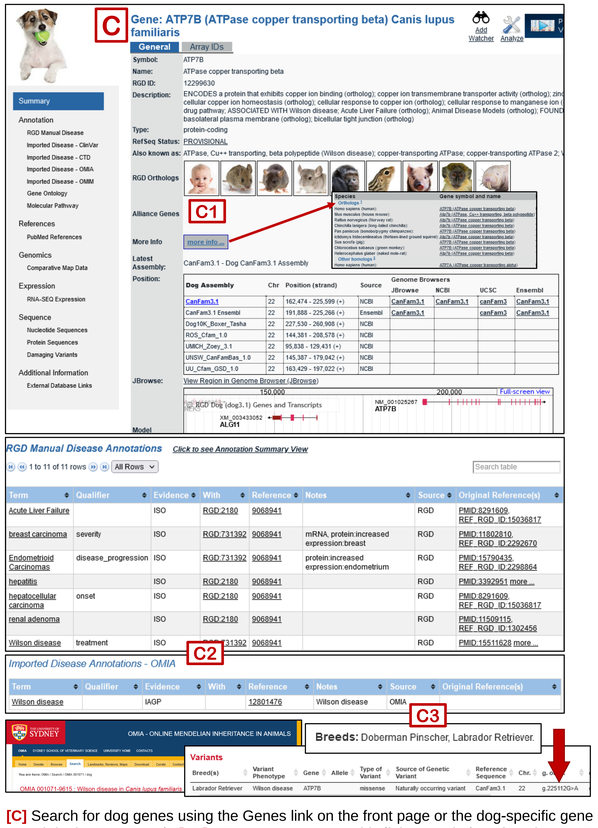


[F] MOET, the Multi-Ontology Enrichment Tool, does ontology enrichment analysis for any vocabulary used for gene annotation at RGD and for any RGD species. The results list the number of genes from the input and reference lists annotated to each over-represented term with the uncorrected and Bonferroni corrected p-values and odds ratios. Click the "Annotated Genes" number to see the list of input genes annotated to the term and resubmit that subset for analysis. The results shown here are for the 40 dog genes associated with Wilson disease.

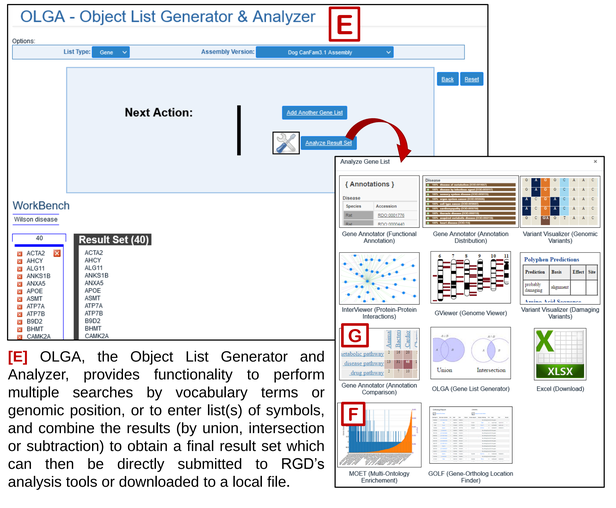


[G] RGD's Gene Annotator (GA) Tool returns all annotations for an input list of genes and their orthologs. The "Comparison Heat Map" function of the tool allows users to find genes annotated to terms from two different ontologies or two different branches of the same ontology. Clicking on highlighted square brings up the list of three genes annotated to any "disease of anatomical entity" and "drug pathway", including ATP7B.

[H] Genome Information Pages give detailed stats for the reference sequence and its genome annotations. Chromosome pages list the numbers of genes by type, the number of disease-related genes that map to that chromosome and more.



[C] Search for dog genes using the Genes link on the front page or the dog-specific gene search in the Dog Portal. [C1] RGD gene pages provide links to orthologs for other RGD species, including human, via the species icons or the list accessed by clicking "More info…" [C2] ATP7B has been shown to be associated with Wilson disease (WD) and its related phenotypes in human, rat and mouse. Similarly, an association was also seen between WD and a variant in Doberman Pinschers and Labrador Retrievers (arrow) as reported in the Online Mendelian Inheritance in Animals database and imported into RGD [C3]. The corresponding variant can be found in several Doberman and Labrador samples in RGD's Variant Visualizer tool (see below).



[E] OLGA, the Object List Generator and Analyzer, provides functionality to perform multiple searches by vocabulary terms or genomic position, or to enter list(s) of symbols, and combine the results (by union, intersection or subtraction) to obtain a final result set which can then be directly submitted to RGD's analysis tools or downloaded to a local file.

RGD's data and tools are freely and publicly available. For your convenience, the original version of this poster is also available at https://doi.org/10.6084/m9.figshare.17435090

RGD is supported by grant R01HL064541 from the National Heart, Lung, and Blood Institute of the National Institutes of Health.