The Rat Genome Database (RGD) facilitates genomic and phenotypic data integration across multiple species for biomedical research

M. L. Kaldunski[1] · J. R. Smith[1] · G. T. Hayman[1] · K. Brodie[3] · J. L. De Pons[1] · W. M. Demos[1] · A. C. Gibson[1] · M. L. Hill[1] · M. J. Hoffman[1,2] · L. Lamers[1] · S. J. F. Laulederkind[1] · H. S. Nalabolu[1] · K. Thorat[1] · J. Thota[1] · M. Tutaj[1] · M. A. Tutaj[1] · M. Vedi[1] · S. J. Wang[1] · S. Zacher[4] · M. R. Dwinell[1,2] · A. E. Kwitek[1,2]

Abstract

Model organism research is essential for discovering the mechanisms of human diseases by defining biologically meaningful gene to disease relationships. The Rat Genome Database (RGD, (https://rgd.mcw.edu)) is a cross-species knowledgebase and the premier online resource for rat genetic and physiologic data. This rich resource is enhanced by the inclusion and integration of comparative data for human and mouse, as well as other human disease models including chinchilla, dog, bonobo, pig, 13-lined ground squirrel, green monkey, and naked mole-rat. Functional information has been added to records via the assignment of annotations based on sequence similarity to human, rat, and mouse genes. RGD has also imported well-supported cross-species data from external resources. To enable use of these data, RGD has developed a robust infrastructure of standardized ontologies, data formats, and disease- and species-centric portals, complemented with a suite of innovative tools for discovery and analysis. Using examples of single-gene and polygenic human diseases, we illustrate how data from multiple species can help to identify or confirm a gene as involved in a disease and to identify model organisms that can be studied to understand the pathophysiology of a gene or pathway. The ultimate aim of this report is to demonstrate the utility of RGD not only as the core resource for the rat research community but also as a source of bioinformatic tools to support a wider audience, empowering the search for appropriate models for human afflictions.

Introduction

A major challenge for preclinical research is finding, or establishing, a good model for the human disease of interest—one that best recapitulates the phenotypic and genomic profile of that disease in the human system. The incorporation of additional mammalian species will allow researchers to leverage rich datasets across multiple species to find the best model for their needs. Pathogenesis is a complex process that may look different in various species but often involves common pathways, mechanisms, or genetic risk factors. Many animal models are studied to understand human biology and disease. Integrating multidimensional data from multiple species may aid in understanding human disease.

RGD (https://www.rgd.mcw.edu, (Smith et al. 2020)) is recognized as a comprehensive data resource for laboratory rat (*Rattus norvegicus*) as a model for the study of human disease (Laulederkind et al. 2019). Recognizing that the best model system to study human disease depends on the disease etiology and the specific research question, the wide-ranging bioinformatics platform for rat at RGD is enhanced by integration of other mammalian species' genomic and phenotypic data, particularly human, making it ideally suited for facilitating translational research at a single point of access. Researchers frequently assemble information from more than one model organism when investigating gene–disease relationships. The platforms and tools at RGD support this type of data exploration in a consistent way. RGD has been the front-runner in this endeavor, even before the first public

✉ A. E. Kwitek
akwitek@mcw.edu

[1] Department of Biomedical Engineering, The Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI, USA

[2] Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA

[3] Clinical and Translational Science Institute, Medical College of Wisconsin, Milwaukee, WI, USA

[4] Information Services, Medical College of Wisconsin, Milwaukee, WI, USA

release of the rat reference genome in 2004 (Gibbs et al. 2004).

The ability to probe mechanisms of human disease by scrutinizing the available genomic and phenotypic information in rat and mouse has always been a primary objective at RGD. The rat has been studied as a model for human physiology and disease for well over 160 years (Smith et al. 2019). Decades of targeted inbreeding and more recent developments in genome editing in the rat have provided researchers with the ability to produce strains that develop abnormal conditions which more or less mimic human diseases (Aitman et al. 2008; Szpirer 2020, 2021; Chenouard et al. 2021). Combined with a tractable size, social and generally docile behavior, and a relatively short reproductive cycle, the rat is a model of choice for many studies. However, rat is not always the best model, and researchers need to be able to determine and utilize whichever models most closely mimic the human profile for their disease of interest. In addition, rat physiologic and genetic research often incorporate or build upon studies in species such as mouse to best answer research questions. The goal of RGD is to enable exploration of genes that are functionally related in human and model organisms by providing a highly integrated platform that leverages the extensive body of genetic, genomic, and biologic data available. The addition of mammalian species

that are models for human disease and the provision of tools designed for researchers further support discovery about health and disease (Fig. 1). Multiple facets of data have been expertly curated and organized by controlled vocabularies for querying and browsing by biomedical researchers. Data involving genes, strains, and QTL annotated to disease, gene ontology (GO), molecular pathway, gene–chemical interactions, and qualitative and quantitative phenotype are available for downloading or manipulating within RGD's toolbox. Gene–disease relationships are annotated based on published papers, particularly for rat, human, and mouse. Primary annotations are made to genes of whichever species is utilized in the published report. Annotations are then propagated to genes in other species based on orthology and denoted with an "Inferred from Sequence Orthology" (ISO) evidence code. Additional data are imported from well-recognized resources such as the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) (NCBI Resource Coordinators 2018) and its ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) (Landrum et al. 2020), Mouse Genome Informatics (MGI; http://www.informatics.jax.org/) (Bult et al. 2019), Ensembl (https://www.ensembl.org/index.html) (Howe et al. 2021), Online Mendelian Inheritance in Man (OMIM; https://www.omim.org/) (Amberger et al. 2015), Online Mendelian Inheritance
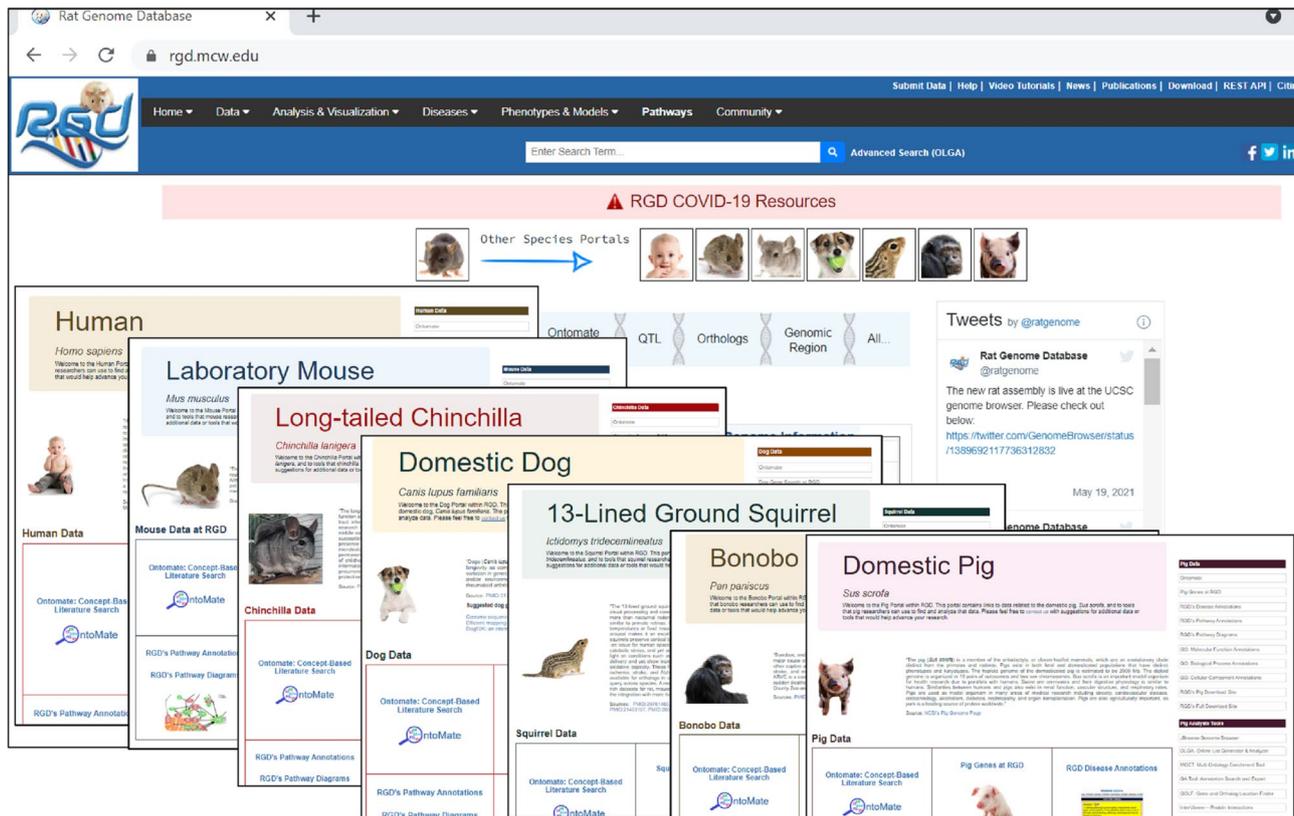

Fig. 1 Species-specific landing pages: the RGD homepage offers links to species-centric portals, which provide consolidated access to data and tools

in Animals (OMIA; https://www.omia.org/home/) (Nicholas 2021), Gene Ontology Consortium (GOC, http://geneontology.org/) (Ashburner et al. 2000; Gene Ontology Consortium 2021), UniProtKB (https://www.uniprot.org/help/uniprotkb) (UniProt Consortium 2019), miRGate (http://mirgate.bioinfo.cnio.es/miRGate/) (Andres-Leon et al. 2015). More recently, RGD expanded to incorporate RNA-Seq expression values with the associated detailed, standardized metadata from GEO (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al. 2013) and Expression Atlas (https://www.ebi.ac.uk/gxa/home) (Kapushesky et al. 2010) for rat and human, and from GTEx (https://gtexportal.org/home/) (GTEx Consortium 2013) for human. RGD gene pages also link to external gene expression analysis resources such as PhenoGen for rat and mouse (Bhave et al. 2007).

To enable the use of these data, RGD has developed a robust infrastructure with a suite of innovative tools and portals for data discovery and analysis. These include the Object List Generator and Analyzer tool (OLGA), the Gene Annotator (GA) tool for exploring functional annotations for a list of genes, interactive Pathway Portal diagrams, InterViewer for visualizing protein–protein interactions, species-specific JBrowse genome browsers, and RGD's Disease Portals which present consolidated data for disease categories. Variant Visualizer was developed to investigate rat genetic variants but has been expanded to include human variants from ClinVar and variants in numerous sequenced dog breeds (Laulederkind et al. 2019; Smith et al. 2020).

Manual curation is vital to the incorporation and standardization of data in RGD, as it is for all model organism databases. These data are propagated to several external databases such as NCBI and Ensembl and provide key functional information identified in the laboratory rat and other species. Curators at RGD identify relevant literature found in PubMed to assign ontology terms to genome objects using OntoMate, a text-mining tool that automatically tags PubMed abstracts with gene, species, and ontology information (Liu et al. 2015). The ultimate aim of this curation is to enhance the utility of RGD not only for the core community of rat researchers but also to empower researchers with a cross-species knowledgebase to drive discoveries that have the potential to improve human health. This expanded offering of data for multiple species and the analysis tools to easily and efficiently leverage these data, give researchers an excellent resource for discovering precision models for their diseases and/or genes of interest.

Creating a cross-species platform for human disease model organisms

The study of genomics and disease is greatly benefitted by the inclusion of multiple species for comparative analyses (Prentice and Webster 2004; Howe et al. 2018; Alliance of Genome Resources Consortium 2020; Baldridge et al. 2021). This concept is widely accepted and has led to the development of a series of comparative tools (for example, Harris et al. 2004; Buels et al. 2016; Mungall et al. 2017; Tholey et al. 2017; Dunn et al. 2019; Ruiz-Arenas et al. 2020; Smith et al. 2020; Foley et al. 2021)). From its inception, RGD has championed comparative genomics, initially providing comparative maps and genomic and phenotypic data for rat, mouse, and human (Table 1) (Kwitek et al. 2001; Twigger et al. 2002, 2004) with the goal of better understanding human disease and pathophysiology by integrating data from rat, the major historical physiological and pharmacological model, and mouse, the major genetic model. The importance of centralizing data from multiple species was also recognized with the recent creation of the Alliance of Genome Resources (the Alliance) (Alliance of Genome Resources Consortium 2020), a centralized database with harmonized data across six established databases for model organisms (yeast, *C. elegans, Drosophila*, zebrafish, mouse, rat) and the Gene Ontology Consortium (GOC). RGD is a founding member of the Alliance and a member of the GOC.

RGD also recognized there is utility in leveraging disease models in species less often considered a traditional model organism and, beginning with chinchilla in 2014, has expanded to include genomic and phenotypic data from eight mammalian species, with the release of two additional species imminent (Table 2), the selection of which has been primarily based on user requests and prioritization of models of human disease with an emphasis on diseases of the heart, lung, and blood. Extensive breed diversification makes *Canis lupus familiaris* (dog) an excellent research model for a wide variety of disorders such as heart disease, cancer, and diabetes (Karlsson and Lindblad-Toh 2008; Davis and Ostrander 2014), and the species has an array of breeds that have been sequenced, which provides comprehensive genomic variation information (Plassais et al. 2019). *Pan*

Table 1 Data incorporated into RGD in 2000 when the database was established

Species	Rat	Human	Mouse
Genes	1987	14	857
Markers (SSLP, EST)	19,562		
Strains	76		
Maps/Assemblies	5	1	

Table 2 Counts of data objects by species

Species	Rat	Human	Mouse	Chinchilla	Bonobo	Dog	13-lined Ground Squirrel	Pig	Green Monkey	Naked Mole rat
Genes	66,544	82,791	75,641	33,266	41,707	52,756	34,418	32,117	41,609	37,447
Proteins	36,395	211,685	88,816	99	43,653	50,115	25,489	122,578	19,525	27,185
Maps/Assemblies	14	20	14	2	4	8	3	4	4	2
Cell Lines	2135	96,476	21,110		33	787	2	248	45	12
Promoters	12,720	66,331	60,623			7,545				
Transcripts	338,715	1,661,329	892,774	100,941	190,658	632,708	89,632	214,434	135,019	74,574
Variants (SNP, Indel)	75,519,796	1,093,201				28,468,158				
Markers (SSLP/EST)	50,134	320,148	55,165							
Strains	4028									
QTL	2378	1911	6828							

Counts reflect data included as of August 2021

paniscus (bonobo) spontaneously develops a number of human disease conditions including hypertension and cardiomyopathies (Lowenstine et al. 2016; Strong et al. 2016; Celestino-Soper et al. 2018). Although most commonly thought of as agricultural, *Sus scrofa* (pig) is also a model for studies of cardiovascular disease, skin conditions, brain physiology and function, surgical procedures, and the microbiome (Ploug et al. 2008; Samulin et al. 2009; Tsang et al. 2016; Warr et al. 2020). *Ictidomys tridecemlineatus* (13-lined ground squirrel) is an established model for retinal function and dysfunction, and for hypoxia/reperfusion injury due to long spells of hibernation/torpor interspersed with short periods of euthermic arousal (Storey 2010; Luu et al. 2016; Ballinger et al. 2017; Tessier et al. 2019). *Chinchilla lanigera* (long-tailed chinchilla) is often used to study the development, physiology, and pathophysiology of the auditory system (Shimoyama et al. 2016). Most recently, RGD incorporated data of *Chlorocebus sabaeus* (green monkey), which is used in research as a model of neurodegeneration, diabetes, hypertension, and HIV/AIDS (Martin et al. 1990; Palmour et al. 1997). In fact, human immunodeficiency virus (HIV) likely evolved from the simian immunodeficiency virus (SIV), for which the green monkey is the natural host (Compton et al. 2012). *Heterocephalus glaber* (naked mole-rat) is studied for aging/longevity and disease resistance and is also used in social science research on colony structure and communication (Jarvis et al. 1994; Buffenstein 2005).

Orthologous gene information across species is easily accessed in RGD through a general gene symbol search, using the Gene-Ortholog Location Finder (GOLF) tool, or through RGD gene report pages. Links to orthologs in all species integrated into RGD and additional species at the Alliance can be accessed from each gene report page. In addition, species-centric portals can be accessed directly from the RGD homepage (Fig. 1). These portals give consolidated access to species-specific data and tools, presented in a consistent format for ease of finding and using relevant information. An OntoMate link is provided with the search preset for the species of interest. Collated links to manual and imported disease and pathway annotations, GO annotations, and RGD's data download site are also included in each species portal, along with direct access to all RGD analysis tools, geared toward that species, as well as NCBI and Ensembl species-specific pages. Another important feature of the species portals is a collection of relevant external resources, a curated sampling of selected publications and review articles, and websites considered for informational value on that species.

Multiple species can model monogenic disease—Wilson disease

Investigators often select a model system that best allows them to address their research question, often that is to investigate a human disease or other biomedical or basic science study. If that disease is monogenic, the availability of cross-species genetic models offers ideal opportunities to study a disease mechanism in the model with the most similar pathophysiology to the human disease. An example of a disease in humans that is modeled by genetic defects in multiple species is Wilson disease (Buiakova et al. 1999; Meng et al. 2004; Merle et al. 2006; Fieten et al. 2012, 2016; Reed et al. 2018; Gerosa et al. 2019; Saba et al. 2019). Wilson disease is a rare autosomal recessive disease characterized by a severe buildup of copper in the brain, liver, and other organs, as a result of abnormal copper metabolism, which results in liver and neurological damage. Although there are a number of other genes that serve as biomarkers and modifiers of the disease, Wilson disease is primarily a monogenic disease caused by mutations in the *ATP7B* gene. Using cross-species data from RGD and the Alliance, and analysis tools available in RGD, a Wilson disease researcher can find extensive evidence supporting the relationship between *ATP7B*

and Wilson disease and has the ability to choose the best model organism for further studies of mechanism or potential treatments.

Data related to Wilson disease are easily accessed by searching for the term Wilson disease in RGD's general search or "Ontology & Annotation" search (Fig. 2a). As shown in Fig. 2b, the ontology report page for Wilson disease contains data for all of RGD's species. On the Wilson disease ontology report page, the result table for human shows data for cell lines, variants, and genes, including *ATP7B*, annotated to the term in the Disease Ontology. A number of these annotations have been manually curated from the literature by RGD or by the Comparative Toxicogenomics Database (CTD), in addition to gene annotations derived from over 1100 ClinVar variants that have been associated with Wilson disease. The human *ATP7B* gene page (Fig. 2c), accessible from the ontology report page by clicking the gene symbol, shows annotations to Wilson disease (Fig. 2d) as well as related diseases such as acute liver failure and liver cirrhosis and phenotypes such as atypical or prolonged hepatitis, high nonceruloplasmin-bound serum copper, polyneuropathy, hepatomegaly, and hepatic failure (Fig. 2e). Both the disease annotations imported from ClinVar and the Clinical Variants section of the human gene page

Fig. 2 Integrated phenotypic and genotypic data annotations for human Wilson disease: **a** RGD's general search bar or Ontology & Annotation search finds disease data; **b** ontology report page for Wilson disease contains data for all of RGD's species; **c** selecting Human in the species list opens Wilson disease annotated genes, and the

human *ATP7B* gene symbol links to the gene report page; **d** annotations to Wilson disease found on the gene report page; **e** phenotype annotations; **f** clinical variants; and **g** clickable icons for finding the gene report page for other species such as mouse

(Fig. 2f) give links to RGD records for individual *ATP7B* variants associated with Wilson disease.

In addition to supporting evidence in human for the association between *ATP7B* and Wilson disease, RGD provides data for several animal models of Wilson disease. The mouse icon in the "RGD Orthologs" section of the human gene page (Fig. 2g) gives one click access to RGD's mouse gene report page (Fig. 3a). In addition to annotations to Wilson disease, experimental ChEBI annotations imported from CTD (Fig. 3b) give evidence for interactions between copper and the mouse Atp7b protein. Mammalian Phenotype ontology annotations (Fig. 3c) imported from Mouse Genome Informatics (MGI) list numerous phenotypes observed in genetically modified (e.g., knockout) mice related to abnormal copper levels in multiple organs, liver cirrhosis, neurological abnormalities, and postnatal lethality. Applicable models in mouse can be explored via the link to MGI in the "External Database Links" section of the mouse gene report page, or by the mouse icon in the "Alliance Genes" list at the top of the RGD gene report page to navigate to the Alliance of Genome Resources website (Fig. 3d) where additional information is available on mouse models of the disease.

From RGD's mouse gene page, the dog icon (Fig. 3e) in the RGD orthologs list gives easy access to the dog gene report page for *ATP7B* (Fig. 3f). Labrador Retrievers, Doberman Pinschers, and several terrier breeds have been shown to occasionally develop Wilson disease (Kruitwagen and Penning 2019). RGD's dog *ATP7B* gene page (https://rgd.mcw.edu/rgdweb/report/gene/main.html?id=12299630#importedAnnotationsOMIA) shows this association via a manually curated annotation imported from the Online Mendelian Inheritance in Animals (OMIA) database (Fig. 3g). According to OMIA, Fieten et al. found a G > A mutation at position 225,112 on chromosome 22 in a Labrador Retriever which displayed increased copper accumulation and copper toxicosis (Fieten et al. 2016). In RGD's Variant Visualizer tool for dog (https://rgd.mcw.edu/rgdweb/front/config.html?mapKey=631), selection of the Labrador Retriever breed group, searching for the *ATP7B* gene, and limiting results to non-synonymous variants returns a heterozygous G > A variant at this position in two of the five samples (Fig. 3h). Clicking on the variant provides access to more detailed information about it, including the information that this variant is predicted to be "possibly damaging" by the Polyphen algorithm. By editing the breed search to include Doberman Pinschers and all available terrier breeds, additional cases of the G > A mutation were seen in a Doberman Pinscher, a Black Russian Terrier, and a German Hunting Terrier.
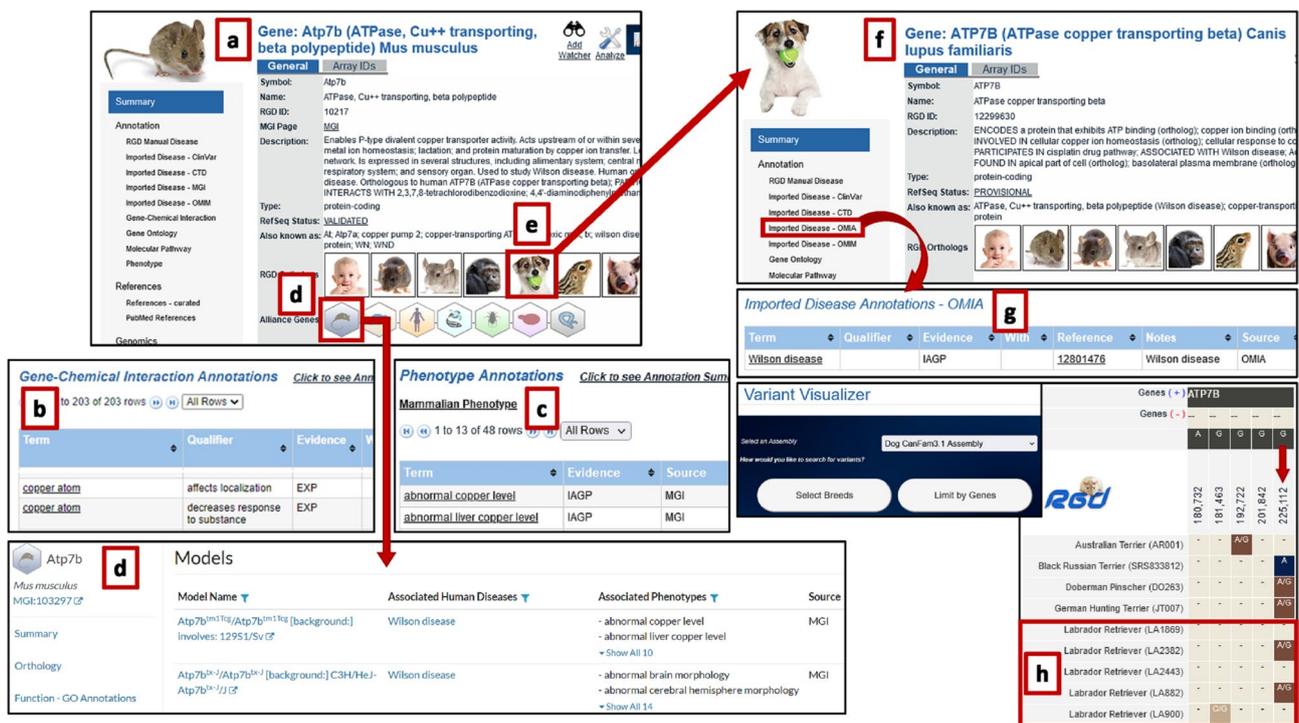

Fig. 3 Multispecies data identify multiple *Atp7b* models for Wilson disease: **a** the mouse gene report page provides access to annotations such as **b** ChEBI gene–chemical interactions, **c** phenotype annotations; **d** external links to Alliance of Genome Resources website; **e** multispecies links such as the dog icon that will open the **f** gene report page for dog, which shows **g** manually curated annotations imported from OMIA; and **h** RGD's Variant Visualizer tool with data for an extensive list of breeds, some of which have possibly damaging variants and have been shown to develop Wilson disease

Additional non-synonymous variants were identified in other terrier breeds, including variants having predicted effects ranging from benign to probably damaging.

The rat *Atp7b* gene report page (Fig. 4a) has Wilson disease annotations based on phenotypes resulting from a spontaneous mutation in *Atp7b* in a rat with Wilson disease features (Fig. 4b). In addition to annotations to Wilson disease and related Mammalian Phenotype annotations, the rat gene page provides links to the mutant allele and mutant strains with aberrant *Atp7b* genes (Fig. 4c).

These same mutant strains can also be accessed by selecting the Find Models tool (https://rgd.mcw.edu/rgd-web/models/findModels.html) from the "Phenotypes & Models" menu in the header of most RGD pages. Searching for Wilson disease in the tool shows that LEC/Hok is an established model for this condition (Fig. 4c). In the result list, click the strain symbol to access the RGD report page for this strain which details the clinical manifestations of Wilson disease displayed in this mutant rat. When compared to the related LEA/Hok strain, LEC/Hok rats show acute liver failure, hepatitis, decreased circulating ceruloplasmin, increased liver copper, hepatic steatosis, oxidative stress, and marks of liver damage such as jaundice and increased circulating alanine transaminase, aspartate transaminase, and bilirubin levels (Fig. 4d), all of which recapitulate phenotypes seen in Wilson disease patients. Information on the LEC/Hok page lists *Atp7b*

as the mutated gene in this strain. A listing of rat models for liver diseases is also accessible through RGD's Liver Disease Portal page. The portal provides a link to a Rat Strain Models page where rat strains which have been curated as models of liver disease are compiled for ease of access. Similar curated information for rat strain models is provided as a link at the bottom of each of RGD's Disease Portal pages, updated to reflect the applicable disease category.

Cross-species studies of complex disease—thrombosis and blood coagulation

In contrast to the relatively small number of monogenic diseases, most diseases are complex, involving the action and interaction of environmental stressors and multiple genetic variations with varying effect sizes. In such cases, the use of genomic integration across multiple species can lead to better understanding of disease mechanisms and pathophysiology. As an example, a current topic of intense research interest is clotting mechanisms. While the mechanisms involved in blood clotting have been well established (Farris 1954), recent observations of coagulopathy in the context of viral infections such as SARS-CoV-2 have stimulated a wave of new investigations, e.g., (Abu-Farha et al. 2020; Campbell et al. 2021), particularly in the area of thrombosis.

Fig. 4 A spontaneous mutation in the LEC/Hok rat makes the strain a model for Wilson disease: **a** the gene report page for *Atp7b* in rat; **b** manual disease annotations for the gene; c. phenotype annotations for the gene; **c** genetic rat model with a mutation in *Atp7b*; and its **d** disease and phenotype annotations

Consolidated data for thrombosis can be found in the RGD Disease Portals (Fig. 5a) (Smith et al. 2020), from the main menu dropdown list. Specifically, selecting the Cardiovascular Disease Portal link (Fig. 5b) provides an integrated ontology browser to facilitate narrowing the data category being displayed (Fig. 5c). In the default view, the term cardiovascular system disease is highlighted in the center panel. Navigate to the child term vascular disease by selecting it in the right panel of the ontology browser. Once selected, it moves to the middle panel and the child term thrombosis can be chosen. While the broad category of cardiovascular system disease has thousands of associated genes, the more specific term of thrombosis gives a more focused list of 139 genes in rat. Annotations for other species can be selected directly by clicking on the species icon. Scrolling down the disease portal page provides the gene list and a genome view of the locations of the associated genes, QTL, and mapped strains, where available. Selecting the "A" icon next to a disease term will link out to a Disease Ontology report page similar to the one described above with annotations for that disease and ontologically related diseases, and with tabs for each species and each associated data type.

From the list of 139 rat genes in the Cardiovascular Disease Portal annotated to thrombosis, a gene of interest can be selected. Here, we have selected *Serpinc1* (Fig. 5d). The gene symbol is a link that will open the report page for the gene in the species selected on the Disease Portal (Fig. 5e). The rat gene report page for *Serpinc1*, for instance, shows RGD manual disease annotations including an experimental annotation to thrombosis. Looking across species, the mouse gene shows an experimental annotation to the more specific child term venous thrombosis. The human gene report page (Online Resource 1) also shows manual, experimental annotations for thrombosis, venous thrombosis, and venous thromboembolism assigned by RGD curators and imported from CTD, as well as annotations based on ClinVar variants. These experimental annotations are then assigned to the orthologous genes in other species as an indicator that, although it has not been shown in those species, the *Serpinc1* genes in those species might also be associated with thrombosis. This inference is indicated by the evidence code "ISO" (Inferred from Sequence Orthology).

Returning to the rat gene page, the Genetic Models section indicates that *Serpinc1* has been mutated in the strain SS.BN-(*D13Rat151-D13Rat197*)-*Serpinc1*^{em2Mcwi} (RGD:12,790,721) (Fig. 5f). Although not annotated to thrombosis, this strain has been studied for a role in the severity of renal ischemia/reperfusion injury. Gene Ontology

Fig. 5 Cardiovascular Disease Portal page for data linking *Serpinc1* to thrombosis in multiple species: **a** Main menu Diseases and **b** the subsequent Cardiovascular Disease Portal; **c** disease ontology is used to narrow the disease term to thrombosis; **d** a more focused disease term results in a shorter list of 139 genes in rat, and *Serpinc1* is selected; **e** the gene symbol is a link that opens the report page for that gene in the selected species; **f** the rat gene report page also provides information for Genetic Models, including in this case SS.BN-(D13Rat151-D13Rat197)-*Serpinc1*^{em2Mcwi} (RGD:12,790,721). For more detailed steps, see Online Resource 1

annotations show that rat *Serpinc1* is involved in blood coagulation and has molecular functions that include heparin binding. Reference IDs are given for each annotation, allowing the researcher to pursue the specifics of the connection (Ikezoe et al. 2013; Nakahara et al. 2013).

Choosing a different gene from the disease portal page, thrombomodulin (*Thbd*), will take the user to that specific gene report page (Fig. 6 and Online Resource 2). On the human thrombomodulin gene page, (Fig. 6a) manually curated and imported annotations can be found for references indicating that human recombinant thromobomodulin was used as a treatment for coagulopathy in mice and indicating its use in disseminated intravascular coagulation in humans (Fig. 6b) (Ikezoe et al. 2013; Nakahara et al. 2013). There is also evidence that *THBD* could be involved in related disease conditions in humans, notably stroke, myocardial infarction, and other thromboembolytic conditions (e.g., Kunz et al. 2000; Pilarska et al. 2010). Drilling into the information available on the gene page for human, RGD manual disease annotations are available that serve as a distillation of the salient information from the research while providing the user easy access to supporting research publications. Likewise, annotations

from ClinVar, CTD, OMIM, and the Human Phenotype Ontology group provide both focused data and links to the original publications (Fig. 6c) and/or the originating databases for additional information. Clinical variants imported from ClinVar are listed on the gene page, complete with associated conditions, chromosomal locations, and potential significance. In rat (Fig. 6d), this gene has been manually annotated for acute injury/failure for kidney, liver, and lung, particularly related to ischemia, but to coagulopathy only through gene orthology (Fig. 6e). There are genomic variants in a number of rat strains (Fig. 6f), also listed on the gene report page and available in Variant Visualizer, and a *Thbd*-mutant rat strain (SS-Thbd^{m1Mcwi}, RGD:1,642,273) has been generated.

Selecting Molecular Pathway (Fig. 6g) in the left sidebar of the *Thbd* gene report page and following the annotation to protein C anticoagulant pathway links to the term's ontology report page showing the list of six genes involved in the pathway, including *Thbd*. At the top of this page is a link to RGD's interactive pathway diagram for the protein C anticoagulation pathway (Fig. 7). This diagram in turn bidirectionally links to the coagulation cascade pathway diagram. The coagulation cascade is the series of events proceeding via the tissue factor (extrinsic) or the contact activation (intrinsic) pathway and converging in

Fig. 6 Cardiovascular disease portal page used to find data linking *THBD* to thrombosis in multiple species: **a** The human thrombomodulin (*THBD*) gene page shows **b** manually curated and imported annotations; **c** links to the original publications/originating databases; **d** selecting the rat species icon will open the gene page for rat with; **e** manual disease annotations; **f** gene variations in rat strains; and **g** molecular pathway annotations and links to the Pathway Ontology. For more detailed steps, see Online Resource 2
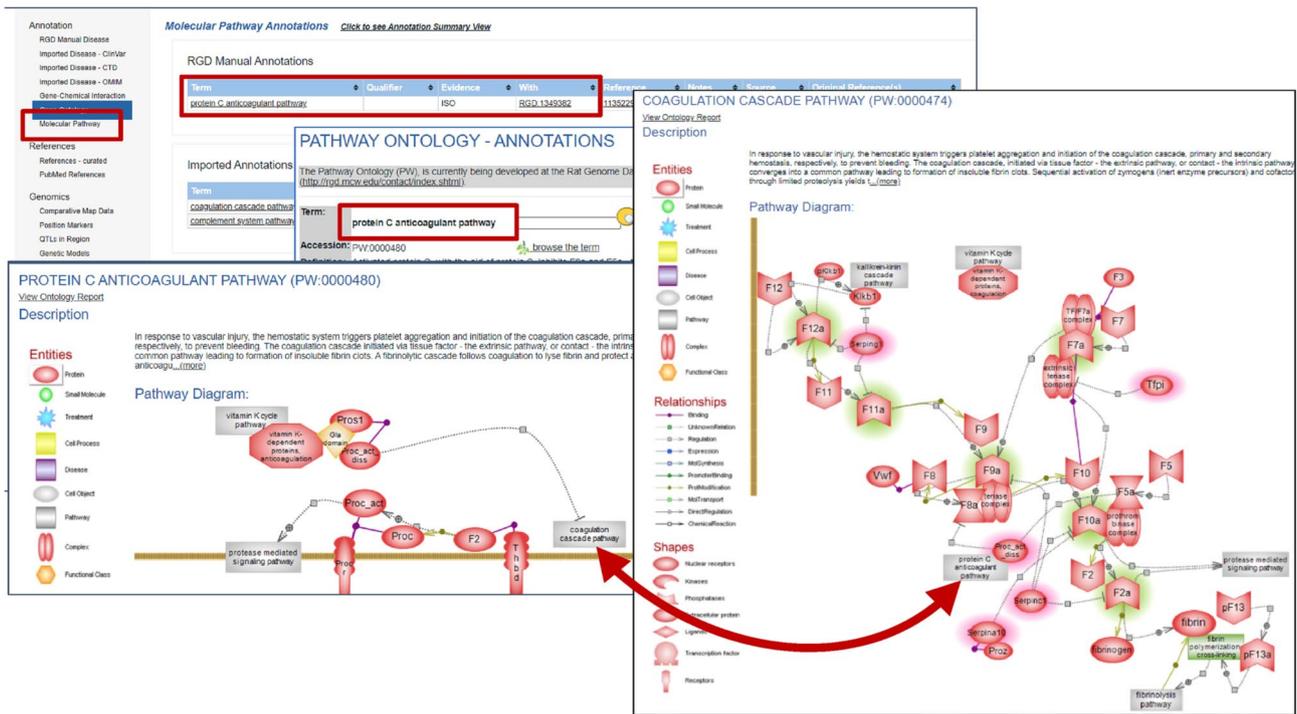
Fig. 7 Molecular pathway annotations and interactive pathway diagrams for *Serpinc1* and *Thbd*: Molecular Pathway Annotations are clickable links that open the Pathway Ontology—Annotations page, from which one can open the protein C anticoagulant pathway and the coagulation cascade pathway diagram. The two pathways are bidirectionally linked

the formation of fibrin clots. Both *Thbd* and *Serpinc1* participate in the anticoagulation systems that are in place to modulate the pathway which, when functioning correctly, prevent excessive clotting. The coagulation cascade and the complement pathway of immune responses are connected in a cross-talk. These diagrams are based on terms in the Pathway Ontology at RGD and offer visualization of molecular pathways and the interacting partners.

From genes to mechanisms or treatments

One of the most powerful RGD functions is the ability to utilize analysis tools to interrogate a gene list leveraging our integrated data infrastructure. One such tool is the Multi-Ontology Enrichment Tool (MOET). MOET can analyze enrichment of a species-specific gene list annotated to a selected term, not just individual genes in the list. Within the Cardiovascular Disease Portal, with the specific disease selection for thrombosis, one finds 139 genes annotated in rat and 140 in mouse, as detailed in Fig. 5. Selecting "mouse" and choosing "MP: Phenotype Ontology" enrichment launches an analysis that utilizes MOET (Fig. 8a and Online Resource 3). The most highly enriched terms for this list of genes are closely related to thrombosis, hemostasis, blood coagulation, and vascular physiology. As detailed in

Online Resource 3, further analysis is possible by navigating the initial results page. Selecting any one phenotype category within the enrichment list will display the list of genes annotated to that term. One can choose "Explore This Gene Set," which will open the full MOET tool and provide additional options. Alternatively, a gene from the dropdown list can be selected to open the gene report page. Within the MOET tool, one may select a different species, a different ontology, or a subset list. For example, selecting "rat" and "CHEBI: Chemical/Drug Enrichment" (Fig. 8b), three of the top four overrepresented terms for the list of thrombosis-related genes are contraceptive drug terms; interestingly, there have been studies showing that use of contraceptive drugs can be associated with DVT and pulmonary embolism (Sitruk-Ware 2016). Selecting the term for oral contraceptives will open the gene list (see Online Resource 3).

Using the Excel download function, one can capture the thrombosis gene list from the Cardiovascular Disease Portal. This list can serve as input for a number of tools within the RGD toolbox (Online Resource 3). It is then possible to employ the power of multiple species and multiple ontologies to interrogate genetic associations, protein alterations, and interactions, even evaluate therapeutic targets. The Gene Annotator (GA) Tool, accessible from either the homepage icon or from the Analysis and Visualization dropdown menu, can accept a gene list for any species and find disease,
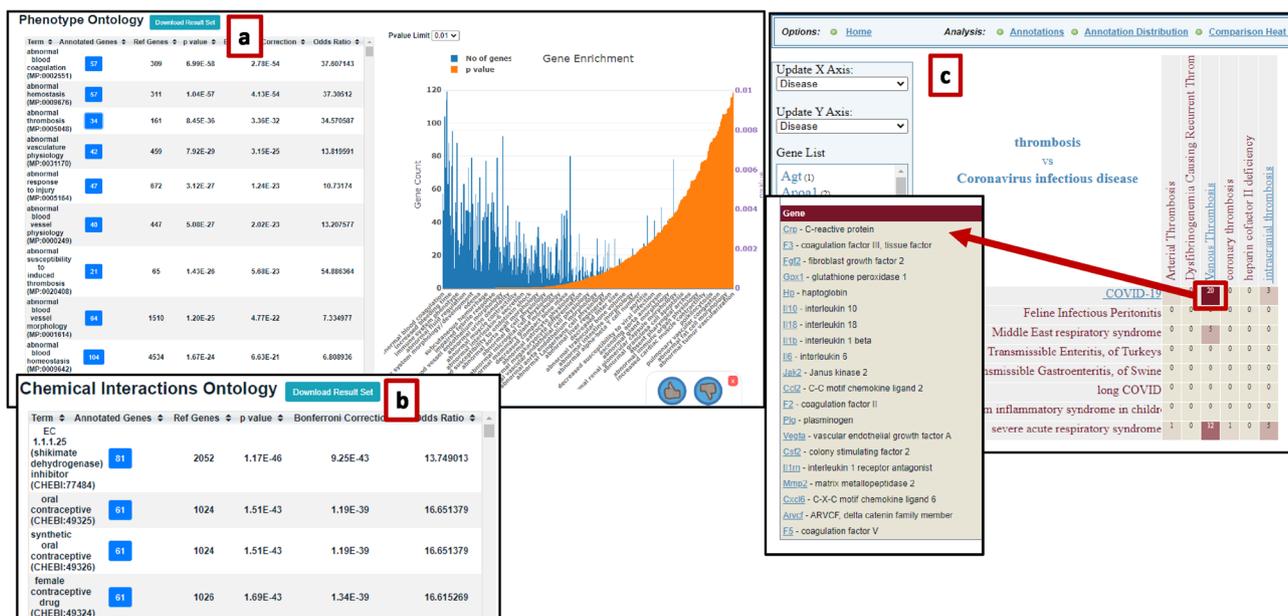
Fig. 8 Using functional analysis tools from the RGD toolkit to interrogate the thrombosis gene list: Within the Cardiovascular Disease Portal, with the specific disease selection for thrombosis, one finds 139 genes annotated in rat, as detailed in Fig. 5 and Online Resource 1. **a** Selecting "mouse" and choosing "MP: Phenotype Ontology" enrichment launches an analysis that utilizes MOET; **b** selecting "rat" and "CHEBI: Chemical/Drug Enrichment" alters the results accordingly; **c** downloading the gene list into Excel and entering it into the Gene Annotator (GA) Tool will return all annotations in any ontology selected and provide a feature of the GA tool that can generate a comparative heatmap of the genes in the intersections between and within ontologies. Here, we show the resultant heatmap of genes for the intersection of disease ontology child terms for thrombosis and Coronavirus infectious disease. For more detailed steps, see Online Resource 3

phenotype, pathway, GO, and ChEBI annotations for all of the genes on the list across all of RGD's species. One can select individual ontologies, gene information sources, and species, or select all (Online Resource 3). The analysis results return all annotations in any ontology selected. A feature of the GA tool is the ability to generate a comparative heatmap of the genes in the intersections between and within ontologies. Here, we show the resultant heatmap of genes for the intersection of disease ontology child terms for thrombosis and Coronavirus infectious disease (Fig. 8c). The cell at the intersection of the Venous Thrombosis column and the COVID-19 row is clickable and opens the list of the 20 genes from the original list of 139 which are annotated to both terms (Fig. 8c). Further analysis, detailed in Online Resource 3, is possible within the results page.

Conclusion

The laboratory rat has been used as a model to study human disease for over 160 years. In many cases, it is the model of choice for such studies, but it is not the best model in every case. For instance, although rat can be used in studies of middle ear infections, chinchilla is the preferred model for such studies because of the striking similarities between the anatomy of the chinchilla ear and that of human children. Likewise, the retina of a 13-lined ground squirrel is far more similar to the primate retina than nocturnal rodents such as rats, making squirrel the preferred model for studies of retinal function. Because of this and because researchers will often use multiple models to obtain as complete an answer as possible to their research questions, resources such as RGD and the Alliance, which integrate data for multiple species, have become even more valuable.

RGD has always had a commitment to providing data across species. This prompted the development of automated pipelines to import data. Because of this, RGD has been able to expand its repertoire of species that are well-studied models for one or more human diseases. The species have been selected and continue to be added on the basis of their status as established models for diseases of interest to RGD users. In addition, for species such as chinchilla and squirrel that are just moving into the realm of genomics and for which there are less species-specific data available, integration with more highly studied species can help to inform research going forward. For less well-studied species, annotations are assigned based on orthology between genes in those species and genes in other species. While there are caveats to this approach (i.e., that gene function can differ

in different species), having this information is a significant advantage for less-studied species and/or species that do not have curated databases of their own. Assignment of disease, pathway, and GO annotations based on orthology benefits researchers by leveraging that information to reveal genes for study in their model organism. A squirrel researcher, for example, could go to an RGD disease portal, find gene lists for squirrel that have been associated with disease in other species, and design experiments to determine if there are similar associations in squirrel.

The integration of both imported data from other sources and RGD's ongoing literature curation, the addition of new rat, bonobo, and other species' assemblies, and the ability to explore human clinical variants as well as variants in rat strains and dog breeds, all, help to elucidate the genomics of disease across species. The addition of new analysis tools and recent improvements to existing tools as well as links that allow users to directly submit the results obtained with one tool as input for another tool facilitate navigation between tools, simplifying researchers' analysis workflows and helping users assess the most efficacious next step in their analysis. This manuscript has demonstrated the utility of the integration of multiple species' genomics into the Rat Genome Database.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00335-021-09932-x.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abu-Farha M, Al-Sabah S, Hammad MM, Hebbar P, Channanath AM, John SE, Taher I, Almaeen A, Ghazy A, Mohammad A, Abubaker J, Arefanian H, Al-Mulla F, Thanaraj TA (2020) Prognostic genetic markers for thrombosis in COVID-19 patients: a focused analysis on D-dimer, homocysteine and thromboembolism. Front Pharmacol 11:587451

Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, Flint J, Gauguier D, Geurts AM, Gould M, Harris PC, Holmdahl R, Hubner N, Izsvak Z, Jacob HJ, Kuramoto T, Kwitek AE, Marrone A, Mashimo T, Moreno C, Mullins J, Mullins L, Olsson T, Pravenec M, Riley L, Saar K, Serikawa T, Shull JD, Szpirer C, Twigger SN, Voigt B, Worley K (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40:516–522

Alliance of Genome Resources Consortium (2020) Alliance of genome resources portal: unified model organism research platform. Nucleic Acids Res 48:D650-d658

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res 43:D789-798

Andres-Leon E, Gonzalez Pena D, Gomez-Lopez G, Pisano DG (2015) miRGate: a curated database of human, mouse and rat miRNA-mRNA targets. Database 2015:bav035

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 25:25–29

Baldridge D, Wangler MF, Bowman AN, Yamamoto S, Schedl T, Pak SC, Postlethwait JH, Shin J, Solnica-Krezel L, Bellen HJ, Westerfield M (2021) Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: current state and a future vision. Orphanet J Rare Dis 16:206

Ballinger MA, Schwartz C, Andrews MT (2017) Enhanced oxidative capacity of ground squirrel brain mitochondria during hibernation. American journal of physiology. Regul Integr Comp Physiol 312:R301-r310

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 41:D991-995

Bhave SV, Hornbaker C, Phang TL, Saba L, Lapadat R, Kechris K, Gaydos J, McGoldrick D, Dolbey A, Leach S, Soriano B, Ellington A, Ellington E, Jones K, Mangion J, Belknap JK, Williams RW, Hunter LE, Hoffman PL, Tabakoff B (2007) The PhenoGen informatics website: tools for analyses of complex traits. BMC Genet 8:59

Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66

Buffenstein R (2005) The naked mole-rat: a new long-living model for human aging research. J Gerontol A Biol Sci Med Sci 60:1369–1377

Buiakova OI, Xu J, Lutsenko S, Zeitlin S, Das K, Das S, Ross BM, Mekios C, Scheinberg IH, Gilliam TC (1999) Null mutation of the murine ATP7B (Wilson disease) gene results in intracellular copper accumulation and late-onset hepatic nodular transformation. Hum Mol Genet 8:1665–1671

Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE (2019) Mouse genome database (MGD) 2019. Nucleic Acids Res 47:D801-d806

Campbell RA, Boilard E, Rondina MT (2021) Is there a role for the ACE2 receptor in SARS-CoV-2 interactions with platelets? J Thromb Haemost 19:46–50

Celestino-Soper PBS, Lynnes TC, Zhang L, Ouyang K, Wann S, Clyde VL, Vatta M (2018) Genetic analyses in a bonobo (Pan paniscus) with arrhythmogenic right ventricular cardiomyopathy. Sci Rep 8:4350

Chenouard V, Remy S, Tesson L, Ménoret S, Ouisse LH, Cherifi Y, Anegon I (2021) Advances in genome editing and application to the generation of genetically modified rat models. Front Genet 12:615491

Compton AA, Hirsch VM, Emerman M (2012) The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. Cell Host Microbe 11:91–98

Davis BW, Ostrander EA (2014) Domestic dogs and cancer research: a breed-based genomics approach. Ilar j 55:59–68

Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE (2019) Apollo: Democratizing genome annotation. PLoS comput Biol 15:e1006790

Farris JM (1954) Thrombosis and embolism. Surg Clin North Am 1271–1287

Fieten H, Leegwater PA, Watson AL, Rothuizen J (2012) Canine models of copper toxicosis for understanding mammalian copper metabolism. Mamm Genome 23:62–75

Fieten H, Gill Y, Martin AJ, Concilli M, Dirksen K, van Steenbeek FG, Spee B, van den Ingh TS, Martens EC, Festa P, Chesi G, van de Sluis B, Houwen RH, Watson AL, Aulchenko YS, Hodgkinson VL, Zhu S, Petris MJ, Polishchuk RS, Leegwater PA, Rothuizen J (2016) The Menkes and Wilson disease genes counteract in copper toxicosis in Labrador retrievers: a new canine model for copper-metabolism disorders. Dis Model Mech 9:25–38

Foley S, Ku C, Arshinoff B, Lotay V, Karimi K, Vize PD, Hinman V (2021) Integration of 1:1 orthology maps and updated datasets into Echinobase. Database 2021

Gene Ontology Consortium (2021) The gene ontology resource: enriching a GOld mine. Nucleic Acids Res 49:D325-d334

Gerosa C, Fanni D, Congiu T, Piras M, Cau F, Moi M, Faa G (2019) Liver pathology in Wilson's disease: from copper overload to cirrhosis. J Inorg Biochem 193:106–111

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera HRA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J,

Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45:580–585

Harris TW, Chen N, Cunningham F, Tello-Ruiz M, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Bradnam K, Chan J, Chen CK, Chen WJ, Davis P, Kenny E, Kishore R, Lawson D, Lee R, Muller HM, Nakamura C, Ozersky P, Petcherski A, Rogers A, Sabo A, Schwarz EM, Van Auken K, Wang Q, Durbin R, Spieth J, Sternberg PW, Stein LD (2004) WormBase: a multi-species resource for nematode biology and genomics. Nucleic Acids Res 32:D411-417

Howe DG, Blake JA, Bradford YM, Bult CJ, Calvi BR, Engel SR, Kadin JA, Kaufman TC, Kishore R, Laulederkind SJF, Lewis SE, Moxon SAT, Richardson JE, Smith C (2018) Model organism data evolving in support of translational medicine. Lab Anim 47:277–289

Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin FL, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, Ilsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P (2021) Ensembl 2021. Nucleic Acids Res 49:D884-d891

Ikezoe T, Takeuchi A, Chi S, Takaoka M, Anabuki K, Kim T, Sakai M, Taniguchi A, Togitani K, Yokoyama A (2013) Effect of recombinant human soluble thrombomodulin on clinical outcomes of patients with coagulopathy after hematopoietic stem cell transplantation. Eur J Haematol 91:442–447

Jarvis JU, O'Riain MJ, Bennett NC, Sherman PW (1994) Mammalian eusociality: a family affair. Trends Ecol Evol 9:47–51

Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, Rustici G, Williams E, Parkinson H, Brazma A (2010) Gene expression atlas at the European bioinformatics institute. Nucleic Acids Res 38:D690-698

Karlsson EK, Lindblad-Toh K (2008) Leader of the pack: gene mapping in dogs and other model organisms. Nat Rev Genet 9:713–725

Kruitwagen HS, Penning LC (2019) Preclinical models of Wilson's disease, why dogs are catchy alternatives. Ann Transl Med 7:S71

Kunz G, Ireland HA, Stubbs PJ, Kahan M, Coulton GC, Lane DA (2000) Identification and characterization of a thrombomodulin

gene mutation coding for an elongated protein with reduced expression in a kindred with myocardial infarction. Blood 95:569–576

Kwitek AE, Tonellato PJ, Chen D, Gullings-Handley J, Cheng YS, Twigger S, Scheetz TE, Casavant TL, Stoll M, Nobrega MA, Shiozawa M, Soares MB, Sheffield VC, Jacob HJ (2001) Automated construction of high-density comparative maps between rat, human, and mouse. Genome Res 11:1935–1943

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, Lyoshin V, Maddipatla Z, Maiti R, Mitchell J, O'Leary N, Riley GR, Shi W, Zhou G, Schneider V, Maglott D, Holmes JB, Kattman BL (2020) ClinVar: improvements to accessing data. Nucleic Acids Res 48:D835-d844

Laulederkind SJF, Hayman GT, Wang SJ, Hoffman MJ, Smith JR, Bolton ER, De Pons J, Tutaj MA, Tutaj M, Thota J, Dwinell MR, Shimoyama M (2019) Rat genome databases, repositories, and tools. Methods Mol Biol 2018:71–96

Liu W, Laulederkind SJ, Hayman GT, Wang SJ, Nigam R, Smith JR, De Pons J, Dwinell MR, Shimoyama M (2015) OntoMate: a text-mining tool aiding curation at the Rat Genome Database. Database 2015

Lowenstine LJ, McManamon R, Terio KA (2016) Comparative pathology of aging great apes: bonobos, chimpanzees, gorillas, and orangutans. Vet Pathol 53:250–276

Luu BE, Biggar KK, Wu CW, Storey KB (2016) Torpor-responsive expression of novel microRNA regulating metabolism and other cellular pathways in the thirteen-lined ground squirrel, Ictidomys tridecemlineatus. FEBS Lett 590:3574–3582

Martin S, Palmour RM, Goldwater R, Gutkowsa J, Hughes C, Hamet P, Ervin FR (1990) Characterization of a primate model of hypertension. The response of hypertensive and normotensive male vervets (Cercopithecus aethiops) to cold pressor stress, captopril administration, and acute bolus of atrial natriuretic factor. Am J Hypertens 3:27–32

Meng Y, Miyoshi I, Hirabayashi M, Su M, Mototani Y, Okamura T, Terada K, Ueda M, Enomoto K, Sugiyama T, Kasai N (2004) Restoration of copper metabolism and rescue of hepatic abnormalities in LEC rats, an animal model of Wilson disease, by expression of human ATP7B gene. Biochem Biophys Acta 1690:208–219

Merle U, Encke J, Tuma S, Volkmann M, Naldini L, Stremmel W (2006) Lentiviral gene transfer ameliorates disease progression in Long-Evans cinnamon rats: an animal model for Wilson disease. Scand J Gastroenterol 41:974–982

Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, Carbon S, Conlin T, Dunn N, Engelstad M, Foster E, Gourdine JP, Jacobsen JO, Keith D, Laraway B, Lewis SE, NguyenXuan J, Shefchek K, Vasilevsky N, Yuan Z, Washington N, Hochheiser H, Groza T, Smedley D, Robinson PN, Haendel MA (2017) The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res 45:D712-d722

Nakahara M, Ito T, Kawahara K, Yamamoto M, Nagasato T, Shrestha B, Yamada S, Miyauchi T, Higuchi K, Takenaka T, Yasuda T, Matsunaga A, Kakihana Y, Hashiguchi T, Kanmura Y, Maruyama I (2013) Recombinant thrombomodulin protects mice against histone-induced lethal thromboembolism. PloS one 8:e75961

NCBI Resource Coordinators (2018) Database resources of the national center for biotechnology information. Nucleic Acids Res 46:D8-d13

Nicholas FW (2021) Online Mendelian Inheritance in Animals (OMIA): a record of advances in animal genetics, freely available on the Internet for 25 years. Anim Genet 52:3–9

Palmour RM, Mulligan J, Howbert JJ, Ervin F (1997) Of monkeys and men: vervets and the genetics of human-like behaviors. Am J Hum Genet 61:481–488

Pilarska E, Lemka M, Bakowska A (2010) Thrombomodulin and antibeta2-glycoprotein I in stroke in children. Med Sci Monit 16:Cr348-351

Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG, Ostrander EA (2019) Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun 10:1489

Ploug KB, Sørensen MA, Strøbech L, Klaerke DA, Hay-Schmidt A, Sheykhzade M, Olesen J, Jansen-Olesen I (2008) K ATP channels in pig and human intracranial arteries. Eur J Pharmacol 601:43–49

Prentice H, Webster KA (2004) Genomic and proteomic profiles of heart disease. Trends Cardiovasc Med 14:282–288

Reed E, Lutsenko S, Bandmann O (2018) Animal models of Wilson disease. J Neurochem 146:356–373

Ruiz-Arenas C, Cáceres A, López M, Pelegrí-Sisó D, González J, González JR (2020) Identifying chromosomal subpopulations based on their recombination histories advances the study of the genetic basis of phenotypic traits. Genome Res 30:1802–1814

Saba L, Tiwari A, Biswas M, Gupta SK, Godia-Cuadrado E, Chaturvedi A, Turk M, Suri HS, Orru S, Sanches JM, Carcassi C, Marinho RT, Asare CK, Khanna NN, Madhusudhan BK, Suri JS, (2019) Wilson's disease: a new perspective review on its genetics, diagnosis and treatment. Front Biosci 11:166–185

Samulin J, Berg PR, Sundvold H, Grindflek E, Lien S (2009) Expression of DLK1 splice variants during porcine adipocyte development in vitro and in vivo. Anim Genet 40:239–241

Shimoyama M, Smith JR, De Pons J, Tutaj M, Khampang P, Hong W, Erbe CB, Ehrlich GD, Bakaletz LO, Kerschner JE (2016) The Chinchilla Research Resource Database: resource for an otolaryngology disease model. Database 2016

Sitruk-Ware R (2016) Hormonal contraception and thrombosis. Fertil Steril 106:1289–1294

Smith JR, Bolton ER, Dwinell MR (2019) The rat: a model used in biomedical research. Methods Mol Biol 2018:1–41

Smith JR, Hayman GT, Wang S-J, Laulederkind SJF, Hoffman MJ, Kaldunski ML, Tutaj M, Thota J, Nalabolu HS, Ellanki SLR, Tutaj MA, De Pons JL, Kwitek AE, Dwinell MR, Shimoyama ME (2020) The year of the rat: the rat genome database at 20: a multi-species knowledgebase and analysis platform. Nucleic Acids Res 48:D731–D742

Storey KB (2010) Out cold: biochemical regulation of mammalian hibernation—a mini-review. Gerontology 56:220–230

Strong VJ, Grindlay D, Redrobe S, Cobb M, White K (2016) A systematic review of the literature relating to captive great ape morbidity and mortality. J Zoo Wildlife Med 47:697–710

Szpirer C (2020) Rat models of human diseases and related phenotypes: a systematic inventory of the causative genes. J Biomed Sci 27:84

Szpirer C (2021) Rat models of human diseases and related phenotypes: a novel inventory of causative genes. Mamm Genome

Tessier SN, Wu CW, Storey KB (2019) Molecular control of protein synthesis, glucose metabolism, and apoptosis in the brain of hibernating thirteen-lined ground squirrels. Biochem Cell Biol 97:536–544

Tholey A, Taylor NL, Heazlewood JL, Bendixen E (2017) We are not alone: the iMOP initiative and its roles in a biology- and disease-driven human proteome project. J Proteome Res 16:4273–4280

Tsang HG, Rashdan NA, Whitelaw CB, Corcoran BM, Summers KM, MacRae VE (2016) Large animal models of cardiovascular disease. Cell Biochem Funct 34:113–132

Twigger S, Lu J, Shimoyama M, Chen D, Pasko D, Long H, Ginster J, Chen CF, Nigam R, Kwitek A, Eppig J, Maltais L, Maglott D, Schuler G, Jacob H, Tonellato PJ (2002) Rat genome database (RGD): mapping disease onto the genome. Nucleic Acids Res 30:125–128

Twigger SN, Nie J, Ruotti V, Yu J, Chen D, Li D, Mathis J, Narayana-samy V, Gopinath GR, Pasko D, Shimoyama M, De La Cruz N, Bromberg S, Kwitek AE, Jacob HJ, Tonellato PJ (2004) Integrative genomics: in silico coupling of rat physiology and complex traits with mouse and human data. Genome Res 14:651–660

UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506-d515

Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, Girón CG, Griffin DK, Hall R, Hannum G, Hourlier T, Howe K, Hume DA, Izuogu O, Kim K, Koren S, Liu H, Manchanda N, Martin FJ, Nonneman DJ, O'Connor RE, Phillippy AM, Rohrer GA, Rosen BD, Rund LA, Sargent CA, Schook LB, Schroeder SG, Schwartz AS, Skinner BM, Talbot R, Tseng E, Tuggle CK, Watson M, Smith TPL, Archibald AL (2020) An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience 9