OXFORD GENETICS

GENETICS, 2023, 224(4), iyad078
https://doi.org/10.1093/genetics/iyad078
Advance Access Publication Date: 29 April 2023
Genetic Models of Rare Diseases

Rare disease research resources at the Rat Genome Database

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

[1]The Rat Genome Database, Department of Physiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA
[2]Clinical and Translational Science Institute, Medical College of Wisconsin, Milwaukee, WI 53226, USA
[3]Finance and Administration, Medical College of Wisconsin, Milwaukee, WI 53226, USA
[4]Joint Department of Biomedical Engineering, Marquette University & Medical College of Wisconsin, Milwaukee, WI 53226, USA

*Corresponding author: FAHA, Department of Physiology, Rat Genome Database, Medical College of Wisconsin, B5235 BSB, 8701 Watertown Plank Rd, Milwaukee, WI 53226. Email: akwitek@mcw.edu

Abstract

Rare diseases individually affect relatively few people, but as a group they impact considerable numbers of people. The Rat Genome Database (https://rgd.mcw.edu) is a knowledgebase that offers resources for rare disease research. This includes disease definitions, genes, quantitative trail loci (QTLs), genetic variants, annotations to published literature, links to external resources, and more. One important resource is identifying relevant cell lines and rat strains that serve as models for disease research. Diseases, genes, and strains have report pages with consolidated data, and links to analysis tools. Utilizing these globally accessible resources for rare disease research, potentiating discovery of mechanisms and new treatments, can point researchers toward solutions to alleviate the suffering of those afflicted with these diseases.

Keywords: rare diseases, Rat Genome Database, ontologies

Introduction

Certain diseases are considered rare, defined by a 1984 amendment to the 1983 Orphan Drug Act (P.L. 97-414; https://www.fda.gov/media/99546/download) as a disease or condition that affects fewer than 200,000 people in the United States (21 USC 360bb; (https://www.govinfo.gov/app/details/USCODE-2010-title21/USCODE-2010-title21-chap9-subchapV-partB-sec360bb). However, collectively, rare diseases affect millions of individuals with conditions that can be serious and life threatening. Estimates are that 80% or more of rare diseases have a genetic cause [Institute of Medicine (US) Committee on Accelerating Rare Diseases Research and Orphan Product Development 2010]. Because of their rarity in human populations, "rare diseases" are often more efficiently studied in preclinical research, in addition to clinical settings. A PubMed search for "rare disease" finds >326,000 results, and a search for "rare disease, genetic" finds >81,000 results. Sifting through all these results to find information about a disease, or a preclinical model of disease to study, would be problematic and onerous. Here in this manuscript, we summarize rare disease information that can be found and analyzed at The Rat Genome Database (RGD) (https://rgd.mcw.edu/) (Smith *et al.* 2020) as a resource to benefit researchers, and walk through some examples of how the resources can be utilized.

For some rare conditions, scientific progress has brought dramatic improvements in the length and quality of life for patients. For example, in the mid-20th century, children with cystic fibrosis faced an average life expectancy of less than 10 years. Today, there is not a cure, but targeted treatments have helped increase average life expectancy to nearly 40 years (https://www.cff.org/). Progress can take the form of research that provides a more thorough understanding of a disease, or therapies that treat symptoms/phenotypes, or may be mutation-targeted or stem cell therapies (Rafeeq and Murad 2017). Gene-editing approaches have been tested *in vitro*, and therapies based on clustered regularly interspaced short palindromic repeats [CRISPR (Jinek *et al.* 2012)] may be on the horizon (Lee *et al.* 2021). Technological advances including DNA sequencing and analysis, computer-aided tools, and online resources are allowing a more thorough understanding of rare disorders (Pogue *et al.* 2018).

Advances in rare disease research have and will continue to also illuminate disease mechanisms and treatment avenues for more common conditions. For example, research on Wilms tumor, a rare pediatric kidney cancer, has increased the understanding of the genetics, epigenetics, and molecular biology of many cancers, and has even elucidated a developmental paradigm for nephrogenesis in general (Feinberg and Williams 2003).

Studies of Fanconi anemia (FA), a genome instability syndrome, have illuminated disease mechanisms of bone marrow failure, breast cancer, and resistance to chemotherapy (D'Andrea 2010), and have discovered a novel DNA repair mechanism required for maintaining genomic stability and preventing cancer (Kee and D'Andrea 2010). A recent study shows that the primary

Received: January 13, 2023. **Accepted:** April 19, 2023

genomic signature of FA repair deficiency is the presence of high numbers of structural variants, and implicates these variants in the frequency and severity of squamous cell carcinomas in these subjects (Webster *et al.* 2022).

RGD rare disease resources

RGD incorporates extensive rat disease and phenotype data that is integrated with rat strain, genetic, genomic, and other genome-scale information. This professionally curated knowledgebase is supplemented with complementary human (and eight additional species) genomic and phenotypic data organized within an infrastructure of standardized ontologies and bioinformatic tools that allow users to explore disease/gene/strain connections. The primary mission of RGD is to provide novel ways to integrate the vast amount of genomic and biological data for the rat with other mammalian disease models, and with similar data facets for human, to provide a unique comparative discovery platform for researchers to identify and evaluate precision rat models, to test novel hypotheses, and to make vital discoveries related to human health and disease. Imported and professionally curated annotations at RGD encompass multiple ontologies, as shown in Table 1.

RGD strain reports

While clinical research and computational modeling show promise (Zhao and Wei 2018; Ehrhart *et al.* 2021), basic research tools available to biomedical investigators are critical for rare disease research, particularly animal models of disease. Rats are well-utilized models of human diseases, with genetic variants and phenotypes that can be exploited to study mechanisms, treatments, and more (Mashimo *et al.* 2005). RGD houses an extensive strain registry and associated annotations, including those that link models to rare diseases and phenotypes. These rat strain models encompass all types, including inbred strains that develop disease spontaneously or with induction, and may be polygenic and complex in genomic nature. Also, strains with a modified chromosome (consomic) or modified chromosomal region (congenic) are listed. There are strains with gene-targeted modifications [N-ethyl-N-nitrosourea (ENU) (van Boxtel *et al.* 2010), transcription activator-like nucleases (TALEN) (Ménoret *et al.* 2014), CRISPR, (Sato *et al.* 2022), etc.] represented. Each strain has a strain report page that collects all relevant information in one location (Fig. 1). Strain nomenclature, IDs, alleles, type (e.g. mutant, inbred), sources, origins, curated references and annotations, and more are available on the report page for each strain. Strain annotations may have additional information about different aspects of disease being modeled by the strain, for example disease penetrance vs induced disease.

For example, the rat model of Cockayne Syndrome developed by Xu *et al.* (2019) recapitulates the phenotypes that characterize the human disease more completely than any prior mouse model, including cerebellar cortex atrophy and dysmyelination (Xu *et al.* 2019; Pacak and Brooks 2020). This rat model SD-*Ercc6*^{em1Cgen} (RGDID: 126925978) has a CRISPR mutation in the *Ercc6* gene. A strain report page is available at RGD with links to disease and phenotype annotations (Fig. 1). An allele report page for *Ercc6*^{em1Cgen} is available directly or via the strain report page as well. There is a gene report page for *Ercc6* that follows the same general format as strain report pages with consolidated information for gene symbol, name, definitions, orthologs, alleles, and variants. Sections are provided for imported information [Clinvar (Landrum *et al.* 2020; https://www.ncbi.nlm.nih.gov/clinvar/), OMIM (Hamosh *et al.* 2000; https://www.omim.org/), ChEBI (Hastings *et al.* 2016; https://www.ebi.ac.uk/chebi/), Gene Ontology (Gene Ontology Consortium 2021; http://geneontology.org), etc.], and manually curated information (mammalian phenotype, molecular pathway, strains, and Gene Ontology for rat) derived from published literature, and mapping and sequence data.

Similarly, a strain developed to study autosomal recessive cerebellar ataxia, or ataxia telangiectasia, F344-*Atm*^{em1} (RGDID: 12879400), has multiple disease and phenotype annotations recorded in RGD (https://rgd.mcw.edu/rgdweb/report/strain/main.html?id=12879400). This model displays phenotypes that are similar to patients with milder forms of the disorder and is suitable for studying the neurodegeneration characteristic of the disease (Quek *et al.* 2017). *Atm* knockout in primary rat neurons has demonstrated response to anti-inflammatory drugs as treatment, and has contributed to an increased understanding of the mechanisms involved, lending support to this treatment in patients

Table 1. RGD imported and curated by ontologies.

Ontology	# terms with annotations
RDO: RGD Disease Ontology; (Hayman *et al.* 2016; Schriml *et al.* 2022) https://rgd.mcw.edu/rgdweb/portal/index.jsp	10733
MP: Mammalian Phenotype; (Smith and Eppig 2015) https://www.informatics.jax.org/vocab/mp_ontology	10797
HP: Human Phenotype (Köhler *et al.* 2021) https://hpo.jax.org/app/	9512
GO: Biological Process (Gene Ontology Consortium 2021) http://geneontology.org/docs/ontology-documentation/	13751
GO: Cellular Component (Gene Ontology Consortium 2021) http://geneontology.org/docs/ontology-documentation/	1962
GO: Molecular Function (Gene Ontology Consortium 2021) http://geneontology.org/docs/ontology-documentation/	4909
ChEBI: Chemical Entities of Biological Interest (Hastings *et al.* 2016) https://www.ebi.ac.uk/chebi/	6741
PW: Pathway Ontology (Petri *et al.* 2014) https://rgd.mcw.edu/wg/home/pathway2/	1018
VT: Vertebrate Trait Ontology (Park *et al.* 2013) https://rgd.mcw.edu/rgdweb/ontology/search.html	180
CMO: Clinical Measurement (Smith *et al.* 2013) https://rgd.mcw.edu/rgdweb/ontology/search.html	368
MMO: Measurement Methods (Smith *et al.* 2013) https://rgd.mcw.edu/rgdweb/ontology/search.html	143
XCO: Experimental Condition (Smith *et al.* 2013) https://rgd.mcw.edu/rgdweb/ontology/search.html	123
RS: Rat Strains (Nigam *et al.* 2013) https://rgd.mcw.edu/rgdweb/ontology/search.html	4153
Total number of (unique) terms used for annotations	64397

For more information regarding each ontology, respective references and websites are provided.

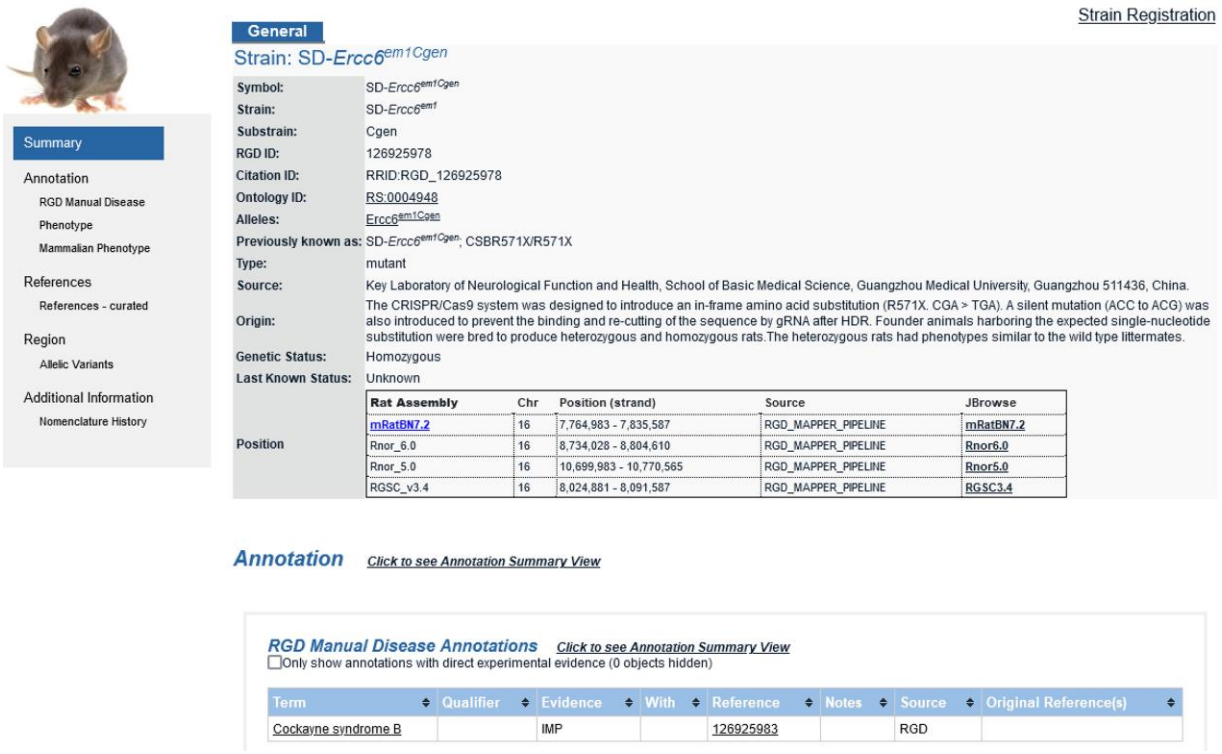

Fig. 1. RGD strain report pages. Strain nomenclature, IDs, alleles, type (mutant, inbred), sources, origins, curated references and annotations, and more are available on the report page for each strain https://rgd.mcw.edu/rgdweb/report/strain/main.html?id=126925978).

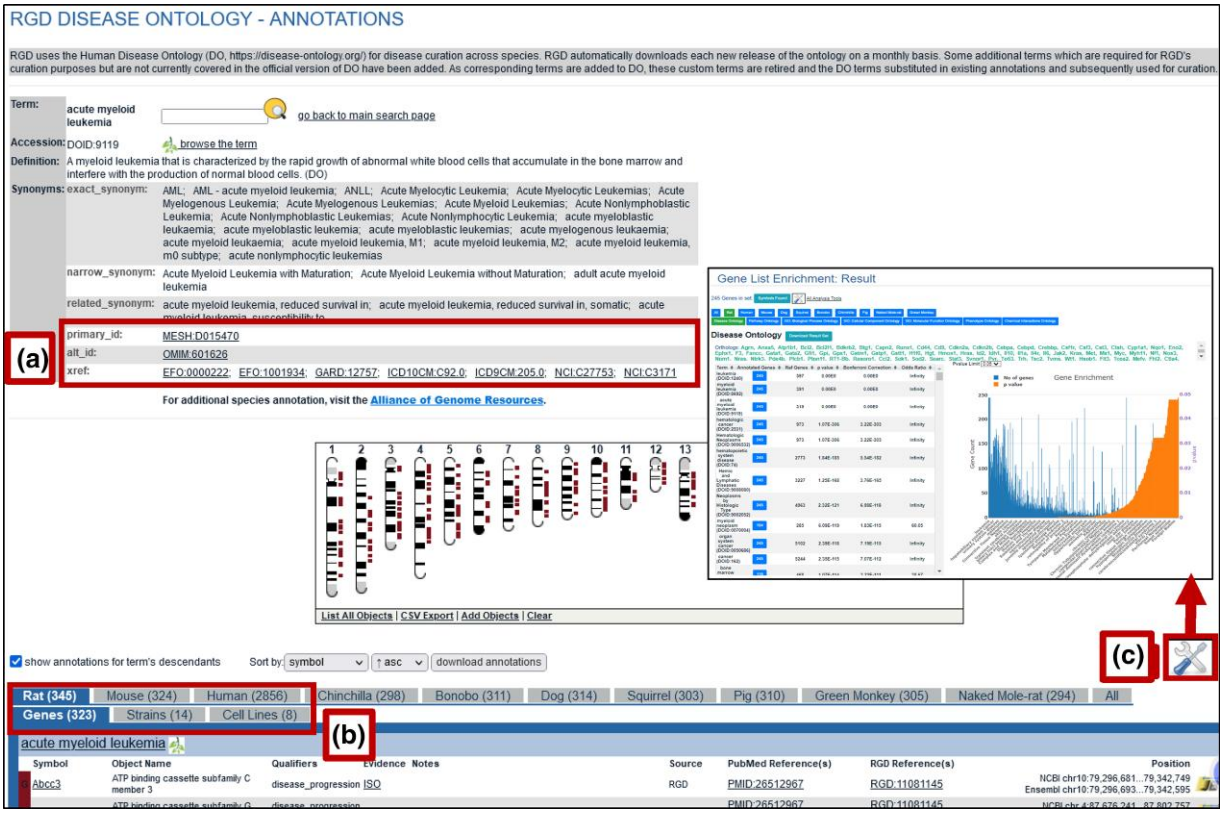

Fig. 2. RGD disease ontology term report page. Disease term report pages include consolidated information including definitions, synonyms, links, gene/strain lists, and tools. a) Indicates the cross references to multiple disease ontology resources and external links. b) Indicates categories of lists on the disease report page for genes in rat (default), other species by sequence orthology, rat strains, and cell lines with annotations to the disease. Where applicable, human clinical variants will be listed when the human tab is selected. Lists are downloadable. c) Indicates the RGD toolbox icon that allows the gene list to be analyzed directly in any of the RGD suite of analysis tools, for example a MOET gene list enrichment analysis. (https://rgd.mcw.edu/rgdweb/ontology/annot.html?acc_id=DOID:9119).

Table 2. RGD disease terms with annotations to genes and strains.

Disease	Annotations #	Annotated rat genes	Annotated strains (RGDID)
Acute erythroid leukemia, DOID:0080780	168	Cad, Ddx41, Dhodh, Flt3, Gfi1b, Hoxb9, Kmt2a, Npm1, Nup98, Rb1, Tp53, Umps	LE/Stm (629485), LEXF2A/Stm (1302605), LEXF2D/Stm (7349321), LEXF3/Stm (1302707), LEXF4/Stm (1302604), LEXF5/Stm (1302723), LEXF6A/Stm (4140405), LEXF7B/Stm (1302649), LEXF8A/Stm (1302653), LEXF8B/Stm (7349322), LEXF8C/Stm (4140408), LEXF8D/Stm (1302699), LEXF9/Stm (1302618)
Cystic fibrosis, DOID:1485	4463	Adrb1, Adrb2, Adrb3, Ager, Akp3, Bglap, C5, Ccl11, Ccl17, Ccl2, Ccl4, Ccr3, Cd14, Cd40lg, Cftr, Cftrem1Ang, Cftrem1Sage, Cftrem2Ang, Clca1, Clcn2, Csf3r, Cxcl1, Cxcl2, Cxcl3, Cxcl9, Cxcr2, Cxcr3, Cyp1a1, Dctn4, Defb4, Defb5, Edn1, Eng, Ephx1, Fas, Faslg, Fcgr2a, Gclc, Gstm1, Gstm3, Gstm5, Gstp1, Gstt1, Havcr2, Hfe, Hmox1, Hspa1b, Hspd1, Igf1, Igfbp3, Il13, Il17a, Il18, Il18bp, Il1a, Il1b, Il1rn, Il6, Il9, Il9r, Irf1, Lep, Lta, Mbl2, Mif, Mir155, Mmp9, Mpo, Muc1, Muc2, Muc5ac, Muc5b, Muc6, Ndufs1, Nos1, Nos2, Nos3, Plg, Ppara, Prss1, Ptgdr2, Ptgs2, Ptx3, RT1-M3-1, Scnn1a, Scnn1b, Scnn1g, Serpina1, Serpina3n, Sftpa1, Sftpb, Sftpc, Sftpd, Slc26a9, Slc6a14, Slc9a3, Tgfb1, Timp1, Tlr4, Tlr5, Tlr9, Tnf, Tnfrsf1a	SD-*Cftr*^{em1Ang} (126925992), SD-*Cftr*$^{em1Sage-/-}$14392815), SD-*Cftr*^{em2Ang} (126925994)
Cockayne syndrome type 3, DOID:2962	753	Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Ercc6, Ercc6em1Cgen, Ercc8, Ghr, Igf1, Ndufaf2, Polr1g, Xpa	SD-*Ercc6*^{em1Cgen} (126925978)
Cockayne syndrome type II, DOID:0080908	315	Ercc6, Ercc6em1Cgen	SD-*Ercc6*^{em1Cgen} (126925978)
Cockayne syndrome type III, DOID:2962	753	Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Ercc6, Ercc6em1Cgen, Ercc8, Ghr, Igf1, Ndufaf2, Polr1g, Xpa	SD-*Ercc6*^{em1Cgen} (126925978)
Acute myeloid leukemia, DOID:9119	5327	Abca1, Abcc3, Abcg2, Abl1, Acsl6, Adcy7, Agrn, Akt1, Anapc2, Anxa2, Anxa4, Anxa5, Anxa6, Aqp9, Arhgap26, Arid4a, Asmtl, Asxl1, Asxl2, Atg2b, Atp1b1, Baalc, Bach2, Bcl2, Bcl2l1, Bcl2l10, Bcor, Bdkrb1, Bdkrb2, Birc5, Bmi1, Brd4, Brd7, Btg1, Cad, Calr, Capg, Capn2, Casp7, Cbfb, Cbl, Cbr1, Ccl2, Ccna1, Ccnd1, Ccnd2, Cd33, Cd44, Cd86, Cd9, Cdh1, Cdk6, Cdkn1b, Cdkn2a, Cdkn2a_v1, Cdkn2a_v2, Cdkn2b, Cebpa, Cebpd, Cebpe, Cfhr1, Cflar, Chi3l1, Chic2, Chmp5, Cnr2, Coro7, Crebbp, Csf1r, Csf2, Csf3, Csf3r, Cst3, Ctcf, Ctla4, Ctnna1, Ctsh, Ctsz, Cxcr4, Cyp1a1, Cyp2b3, Cyp2d4, Dapk1, Dapk2, Dcaf7, Ddx41, Dhcr7, Dhodh, Dhx15, Dlec1, Dnmt1, Dnmt3a, Dnmt3b, Ehd3, Ehmt2, Eif4ebp1, Enah, Eno2, Ephx1, Epor, Erbb3, Ercc1, Ercc2, Erg, Etv6, Ezh2, F3, Fadd, Fanca, Fancc, Fas, Fermt2, Fermt3, Fgfr1, Fhl2, Flt3, Fndc3b, Foxo1, Fxyd6, Gas2l1, Gata1, Gata2, Gfi1, Gfi1b, Gli1, Gli2, Gmps, Gpatch1, Gphn, Gpi, Gpx1, Gskip, Gsr, Gstm1, Gstp1, Gstt1, Gtf2i, H1f0, H1f2, Hgf, Hmox1, Hoxa9, Hoxb9, Hras, Hspb1, Id2, Idh1, Idh2, Ier2, Ifi30, Ifng, Il10, Il17a, Il1a, Il4r, Il6, Inpp4b, Insl6, Irf2bp2, Itgal, Itgam, Itgav, Itgax, Itgb2, Itgb3, Jak1, Jak2, Jak3, Kansl1, Kat6a, Kcne2, Kit, Klf1, Kmt2a, Kmt2b, Kmt2c, Kmt2e, Kras, LOC100909954, Lat2, Lep, Lpar1, Lpp, Lrp3, Lrrc56, Ltc4s, Lyl1, Mdga1, Mdm2, Me1, Mecom, Mefv, Met, Mfsd11, Mir155, Mir802, Mlf1, Mllt10, Mn1, Mrtfa, Mt-nd6, Mtarc2, Mthfr, Mtrr, Mx1, Mybl2, Myc, Myh11, Nat2, Ncam1, Ncoa2, Nectin2, Nf1, Nos3, Npm1, Nqo1, Nras, Nsd1, Ntrk3, Numa1, Nup214, Nup98, Pcf11, Pdcd1, Pde4b, Phf6, Picalm, Pim2, Plat, Plcb1, Pml, Pou4f1, Pparg, Prame, Prkar1a, Psip1, Ptch1, Pten, Ptpn11, Pvr, Pxdn, RT1-Bb, Rac2, Rac3, Rad21, Rap1gap, Rara, Rasal3, Rasgrp1, Rb1, Retn, Rgs2, Rhpn2, Rock1, Rtel1, Runx1, Runx1t1, Runx3, S100a10, S100a8, Samd9, Samd9l, Samhd1, Septin9, Setd2, Setd4, Sf3b1, Sf3b2, Sgk1, Sh3gl1, Slc7a10, Slc9a2, Slit2, Smo, Sncb, Socs1, Sod2, Sparc, Spi1, Spry4, Srsf2, Srsf4, Stag2, Stat3, Svil, Syngr1, Tcea2, Tcl1a, Tcn2, Tek, Tert, Tet2, Tfpi2, Tfr2, Tgm6, Thbd, Tmem127, Tnfsf10, Tnfsf8, Tnfsf9, Top2a, Tp53, Trh, Trib3, Trio, Tsc1, Tsc2, Tubb2a, Tyms, U2af1, Umps, Vegfa, Vopp1, Vsig4, Wdr88, Wt1, Zbtb16, Zbtb7a, Zfp91, Zrsr2	BN/Rij (155804258), LE/Stm (629485), LEXF2A/Stm (1302605), LEXF2D/Stm (7349321), LEXF3/Stm (1302707), LEXF4/Stm (1302604), LEXF5/Stm (1302723), LEXF6A/Stm (4140405), LEXF7B/Stm (1302649), LEXF8A/Stm (1302653), LEXF8B/Stm (7349322), LEXF8C/Stm (4140408), LEXF8D/Stm (1302699), LEXF9/Stm (1302618)

An expanded version can be found in Supplementary Table 1, including links to the disease report pages that provide the provenance for data and annotations.

Table 3. RGD links to websites on disease ontology report pages.

Resource: URL	Description (taken from websites)
OrphaNet: https://www.orpha.net/consor/cgi-bin/index.php	Rare diseases and orphan drugs portal: Orphanet is a resource on rare diseases to improve the diagnosis, care and treatment of patients with rare diseases. Established in France by the INSERM (French National Institute for Health and Medical Research) in 1997, became a European endeavor from 2000. Orphanet has gradually grown to a Consortium of 40 countries, within Europe and across the globe.
MeSH terms: https://meshb.nlm.nih.gov/record/ui?ui=D007951	National Library of Medicine Medical Subject Headings (MeSH) thesaurus is a controlled and hierarchically-organized vocabulary produced by the National Library of Medicine.
OMIA: https://www.omia.org/home/	Online Mendelian Inheritance in Animals (OMIA) is a catalogue/compendium of inherited disorders, other (single-locus) traits, and associated genes and variants in 358 animal species (other than human, mouse, rats, and zebrafish, which have their own resources). It is developed and housed at the University of Sydney, Sydney School of Veterinary Science, Australia.
OMIM: https://www.omim.org/	Online Mendelian Inheritance in Man (OMIM) is a catalog of human genes and genetic disorders. OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal, medical, or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions. OMIM® and Online Mendelian Inheritance in Man® are registered trademarks of the Johns Hopkins University.
GARD: https://rarediseases.info.nih.gov/ https://ncats.nih.gov/gard	Established by the Rare Diseases Act of 2002, the Genetic and Rare Diseases (GARD) Information Center, part of the NIH National Center for Advancing Translational Sciences, is a public health resource that aims to support people living with a rare disease and their families with free access to reliable, easy to understand information, in English and Spanish. There is no advertising on this website, and GARD does not endorse or promote any companies, products, or services.
ICD9CM, ICD10CM: https://icdlist.com/icd-10/	ICD List is a reference website: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM and ICD-10-PCS codes) is a classification system of diagnosis codes representing conditions and diseases, related health problems, abnormal findings, signs and symptoms, injuries, and external causes of injuries and diseases.
NCI thesaurus: https://ncit.nci.nih.gov/ncitbrowser/pages/	National Cancer Institute (NCI) thesaurus (NCIt) provides reference terminology for many NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities. NCIt is a widely recognized standard for biomedical coding and reference, used by a broad variety of public and private partners both nationally and internationally including the Clinical Data Interchange Standards Consortium Terminology, the US Food and Drug Administration, the Federal Medication Terminologies, and the National Council for Prescription Drug Programs.
EFO: https://www.ebi.ac.uk/efo/	The Experimental Factor Ontology (EFO) provides a systematic description of many experimental variables available in European Bioinformatics Institute (EBI) databases, and for projects such as the genome-wide association studies catalog. It combines parts of several biological ontologies, such as UBERON anatomy, ChEBI chemical compounds, and Cell Ontology. The scope of EFO is to support the annotation, analysis, and visualization of data handled by many groups at the EBI and as the core ontology for Open Targets. EFO is developed by the EMBL-EBI Samples, Phenotypes, and Ontologies Team.

(Fang *et al.* 2016; Quek *et al.* 2017). Each disease discussed above, and many more, has a disease ontology term report page (Fig. 2) that consolidates all the information available, similar to the gene and strain report pages.

RGD phenotype data and the PhenoMiner repository and tool

Beyond rats being used as models of common and rare diseases, they also display phenotypes of disease, or carry defining genetic abnormalities identified as causal or related to disease processes. Recent advances in acute myeloid leukemia (AML) treatments are based on knowledge of the cellular mechanisms of the disease, elucidated using *in vitro* and *in vivo* preclinical models (Dozzo *et al.* 2022), including a number of rat models, as reviewed in McCormack *et al.* (2005) and Skayneh *et al.* (2019). In particular, the transplant model systems, especially the BNML rat (Vaughn *et al.* 1978; Hagenbeek and Martens 1980; Martens *et al.* 1990) model (RGD strain BN/Rij RGDID: 155804258), have proven to be invaluable experimental tools. Additionally, the study of T-lymphomas and erythro- and myeloid leukemias has been advanced by the susceptibility strains F344 and LE/Stm. The production and phenotyping of recombinant inbred (RI) strains from an intercross of these strains have informed researchers of a multifactorial genetic process involving several loci linked with susceptibility and resistance (Lu *et al.* 1999). Although this RI set was originally developed to study susceptibility to chemically induced tumors, it has been shown to

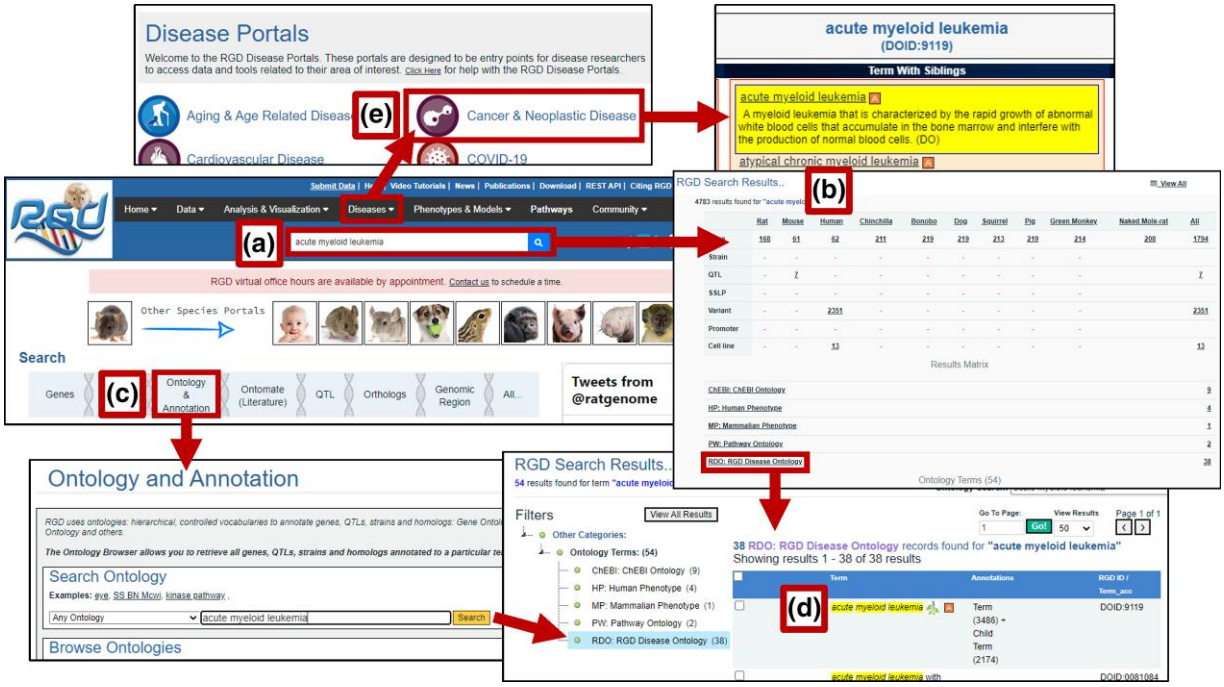

Fig. 3. Accessing rare disease report pages and data in RGD. a) Indicates the general search box on the RGD homepage, in which to enter a search term. b) Indicates the general search results page with categories for objects such as genes and ontologies. c) Indicates the Ontology and Annotation search tool, as an alternative search path to access ontology terms. d) Indicates the Ontology Search results, accessible from the Ontology and Annotation search or by selecting an ontology in the general search results. In the ontologies, a green branch icon will open the disease ontology tree. The red A box will open a disease report page (as shown in Fig. 2). e) Indicates the RGD Disease Portals landing page (https://rgd.mcw.edu/rgdweb/portal/index.jsp), which provides another path to explore the disease ontology, in this case for cancer and neoplastic diseases, and reach disease report pages (https://rgd.mcw.edu/rgdweb/ontology/view.html?acc_id=DOID:9119). Supplementary Figure 1 provides an expanded view of Fig. 3.

be powerful for mapping a wide spectrum of traits, including heart rate, organ development, and blood chemistry parameters (Voigt *et al.* 2008). Several of these strains have quantitative phenotype information in the PhenoMiner (Laulederkind *et al.* 2013) tool at RGD (https://rgd.mcw.edu/rgdweb/phenominer/ontChoices.html). The PhenoMiner repository and mining tool were developed to warehouse rat quantitative phenotype measurements, from both manual curation of scientific literature as well as uploaded data provided by investigators. The PhenoMiner repository, as of February 2023, houses more than 79,000 experimental records, and curation is an ongoing effort. Most of these are summary values curated from PubMed literature, but >34,000 of those records represent individual rat values from data submitted by researchers. Data includes detailed information about what (CMO, clinical measurement ontology), how (MMO, measurement method ontology), and under what conditions (XCO, experimental conditions ontology) phenotypes were measured in what animals (rat strain) for each measurement value. A tutorial for using PhenoMiner can be found at (https://rgd.mcw.edu/wg/home/rgd_rat_community_videos/phenominer-video/).

RGD organizes data using ontologies, which are controlled or standardized vocabularies, as introduced in Table 1. In particular, RGD leverages the Disease Ontology (https://disease-ontology.org/) to organize disease terms and cross-reference vocabularies from multiple resources, and has expanded the ontology with additional custom terms. The National Organization for Rare Disorders (https://rarediseases.org/) lists almost 1,300 rare disease terms, and the OrphaNet rare disease website (https://www.orpha.net/consor/cgi-bin/Disease.php?lng=EN) lists ~22,100 disease and phenotype terms. Of these terms, RGD directly matches more than 8,000 to a disease, phenotype, or a synonym term. A subset of this data is shown in Table 2, and a more encompassing dataset is provided in Supplementary Table 1. Within that list, there are >129,000 disease, phenotype, chemical, and other annotations in RGD, and >136,000 rat gene annotations. Among the disease term annotations are ~260 terms directly annotated to rat strains, in some cases multiple diseases for a strain and frequently multiple strains for a disease, for ~360 unique strain identifiers. There is PhenoMiner quantitative phenotype data available for at least 140 of these rat strains.

RGD disease report pages and disease portals

To find out what is known about a particular disease, the disease ontology term report page (Fig. 2) shows the RGD ontology term and provides synonyms and cross references to multiple disease ontology resources (Table 3: websites). The disease term report page includes the DOID term, definition, synonyms, cross-reference terms, chromosome location idiograms for the genes, strains and quantitative trait loci (QTLs) annotated to the disease, and a list of those annotations with links to external resources. Finding information, tracking curated references for a disease, or a list of genes or clinical variants, or finding a model to study the disease of interest, would start with the disease term report page. The disease ontology itself can be utilized for data mining. For example, a biomedical researcher can discover the existence of sub-types of a disease, related annotations, child terms, and relationships, by viewing the report page for the term of interest. Also, because of the structure of the ontology, a researcher can search for a more general term and retrieve results just for that term or for that term and all of the more granular terms under it in the ontology with a single search. The gene list can be analyzed directly in any of the RGD analysis tools suite, by clicking

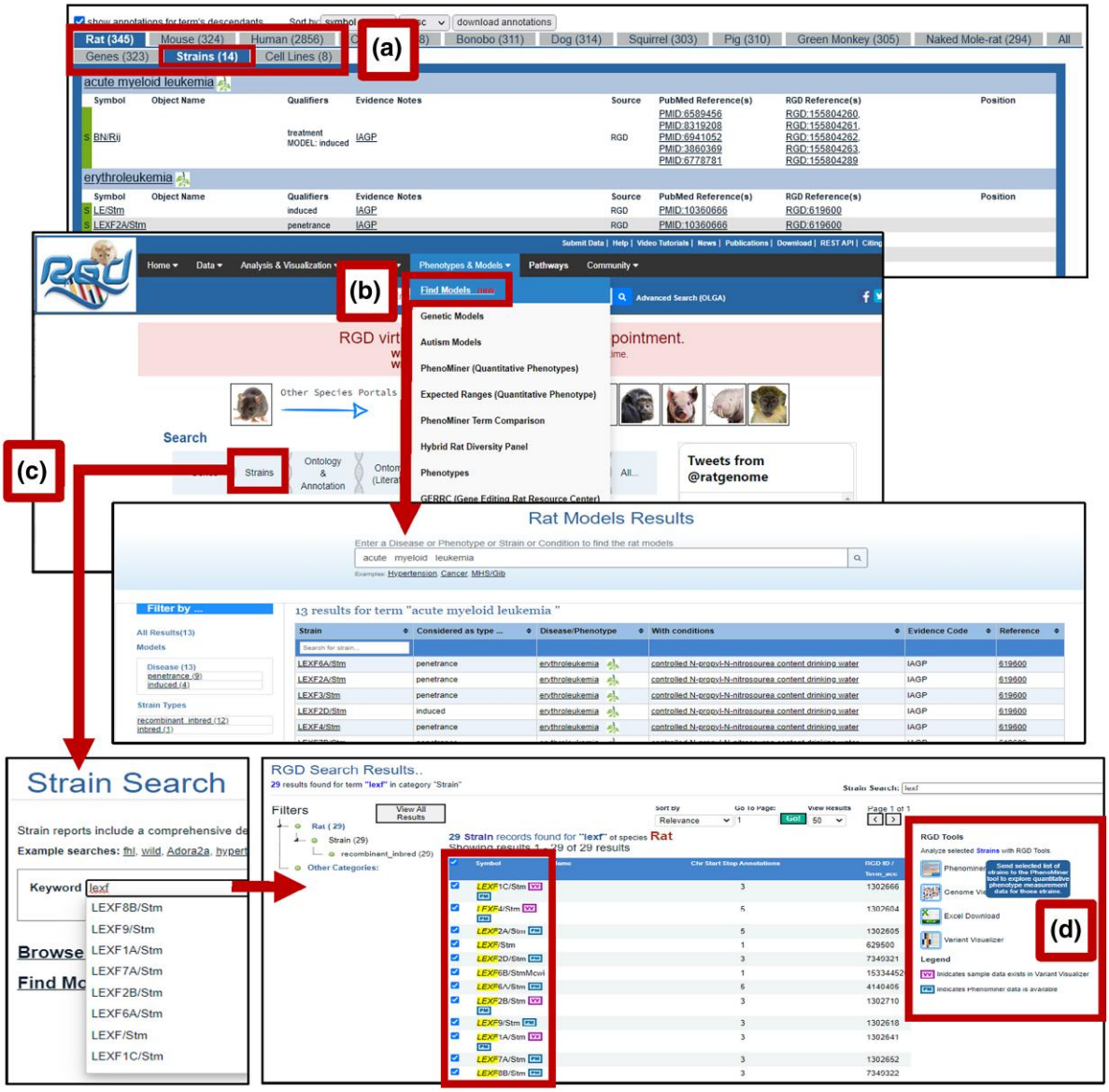

Fig. 4. RGD tools to search and analyze data. a) Indicates the categories of lists on the disease report page for genes, strains, and cell lines annotated to the disease. Selecting the Strains tab will change the display from the gene list in rat (default) to rat strains. b) Indicates the location of the Find Models tool on the RGD homepage in the Phenotypes & Models menu (https://rgd.mcw.edu/rgdweb/models/findModels.html). c) Indicates the direct Strain Search available on the RGD homepage (https://rgd.mcw.edu/rgdweb/search/strains.html). In the search results, a blue "PM" icon in the list of strains indicates that there is PhenoMiner data for that strain. Likewise, a purple "VV" icon indicates that RGD has variant data for that strain in the Variant Visualizer tool. d) Indicates the tools available on the right side of the strain search results page, including the ability to download the list, search for strain-specific variants using the Variant Visualizer, or explore quantitative measurements in the PhenoMiner tool for the strains listed and selected. Supplementary Figure 2 provides an expanded view of the details in Fig. 4.

on the Toolbox icon (Fig. 2c). For example, submitting a gene list to the Multi-Ontology Enrichment Tool (MOET), a statistical analysis tool developed by RGD (https://rgd.mcw.edu/rgdweb/enrichment/start.html; Vedi *et al.* 2022), will perform a gene set ontology enrichment analysis and report the results with terms and statistics and a graph with a modifiable *P*-value cutoff.

There are several ways to reach the disease term report page from the RGD homepage (Fig. 3 and Supplementary Figure 1). The general search box, the Ontology and Annotation search tool, and the Disease Portal landing page accessible from the upper menu are all pathways to the disease ontology and disease report pages.

The search results page (Fig. 3b) lists all categories of data related to the search term. In addition to the disease ontology, the user can explore ChEBI, Phenotype, Pathway, gene, QTL, gene variant, and a list of references associated with the searched term. Navigating to the Disease Portals landing page via the menu item "Diseases" (Fig. 3e) finds categories and subsets of human disease, which have had special curation efforts (Hayman *et al.* 2016; Wang *et al.* 2016). This series of Disease Portals has had targeted efforts to curate research papers and data important to understanding the mechanisms of diseases in selected areas. Gene-disease relationships for rat, human, and mouse are specifically annotated to capture biomarkers, genetic associations, molecular mechanisms, and therapeutic targets. These organized annotations are associated with genes, strains, and QTLs, thus linking functional annotations to genome objects. RGD has developed a robust infrastructure of standardized ontologies, data

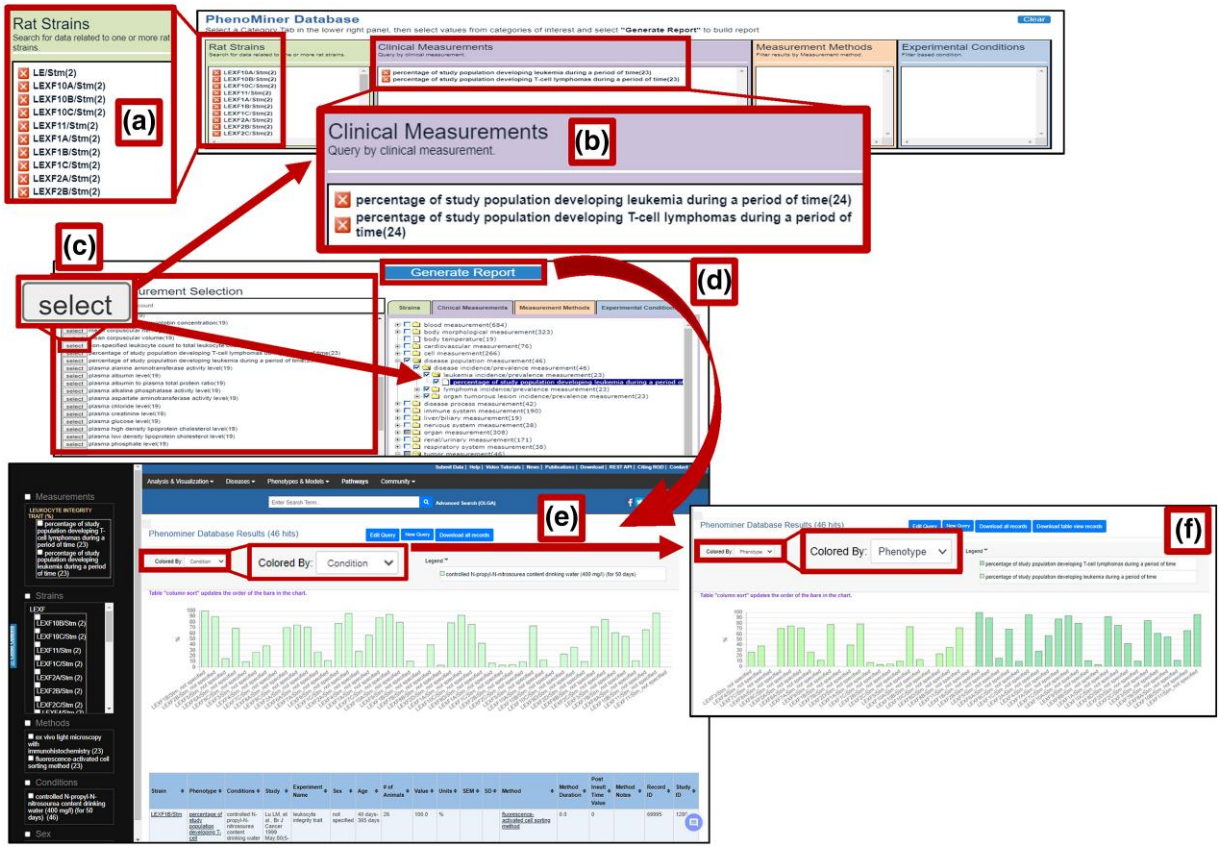

Fig. 5. RGD PhenoMiner repository of quantitative phenotype data (https://rgd.mcw.edu/rgdweb/phenominer/ontChoices.html). a) Indicates strains populating the Rat Strains bucket in the PhenoMiner data selection panels, as they would if entering via the Rat Strains search tool (from Fig. 4d). b) Indicates the Clinical Measurements tab, which will show phenotypes for which the selected strains have data. c) Indicates the Select buttons to load chosen phenotypes into the Clinical Measurements bucket. d) Indicates the Generate Report button, after selections are completed. e) Indicates the PhenoMiner report page, which consists of a graph (if phenotypes share a set of units of measure) and a table of the data, which can be manipulated or downloaded. The left-hand menu facilitates further filtering of data selections. 5f indicates the changes to the report page if "colored by" is changed to Phenotype, which also updates the graph legend; and the table is sorted, which will also sort the bars in the graph. Supplementary Figure 3 provides an expanded view of the details of Fig. 5.

formats, and disease- and species-centric portals, complemented with a suite of innovative tools for discovery and analysis.

How a researcher might utilize RGD rare disease resources

Example 1: acute myeloid leukemia

An example of a rare disease, to showcase RGD resources, is AML. In 2019, there were an estimated 69,700 people living with AML in the United States (https://seer.cancer.gov/statfacts/html/amyl.html). For AML, appropriate Disease Portals would be the Hematologic Disease Portal or the Cancer & Neoplastic Disease Portal. Using the Cancer Portal as an example, the user can optimize search and analysis with ever-increasing specificity within the ontology child terms for any of nine ontologies. Sequence orthology, when available, allows utilizing any of 10 catalogued mammalian species. Navigating the cancer ontology toward greater specificity to find AML will show 320 genes annotated to that disease term for rat and 13 strains from the RI set discussed above. Gene Set Enrichment can be performed using the MOET tool embedded within the Disease Portal pages. Selecting any gene, strain, or disease term on the portal page will transfer the user to the respective report page. The disease ontology term report page for AML, as shown in Fig. 2, has the disease definition, synonyms, and external website cross-referencing links. The Strains tab will change the display from the gene list in rat (default) to rat strains (Fig. 4 and Supplementary Figure 2). Alternative search methods to find a strain annotated to the disease of interest include utilizing the new Find Models tool from the Phenotypes & Models menu (Fig. 4b), or using the Strain Search from the search tools provided on the homepage (Fig. 4c).

To find out more and analyze rat strain models found, selecting strains in the RGD Search Results list and proceeding to any of the tools on the right side of the page (Fig. 4d) provide the ability to download the list, search for strain-specific variants using the Variant Visualizer, or explore quantitative measurements in the PhenoMiner tool for the strain listed.

With the strains selected in Fig. 4, choosing the PhenoMiner link on the right side will open the PhenoMiner quantitative data repository analysis tool, with the strains pre-loaded in the Rat Strains bucket (Fig. 5 and Supplementary Figure 3). Proceeding to the Clinical Measurements tab (Fig. 5b), the user can choose from the list of phenotypes that have data for the selected strains.

The PhenoMiner report page (Fig. 5e) consists of a graph, initially colored by Condition, with a dropdown legend and bars showing the data measurements for each strain for which data is available. Data selections can be manipulated using the left side menu. The table below the graph is sortable, and downloadable—for either all the phenotypes in the original query (when clicking Generate Report), or only results for the current view if terms have been

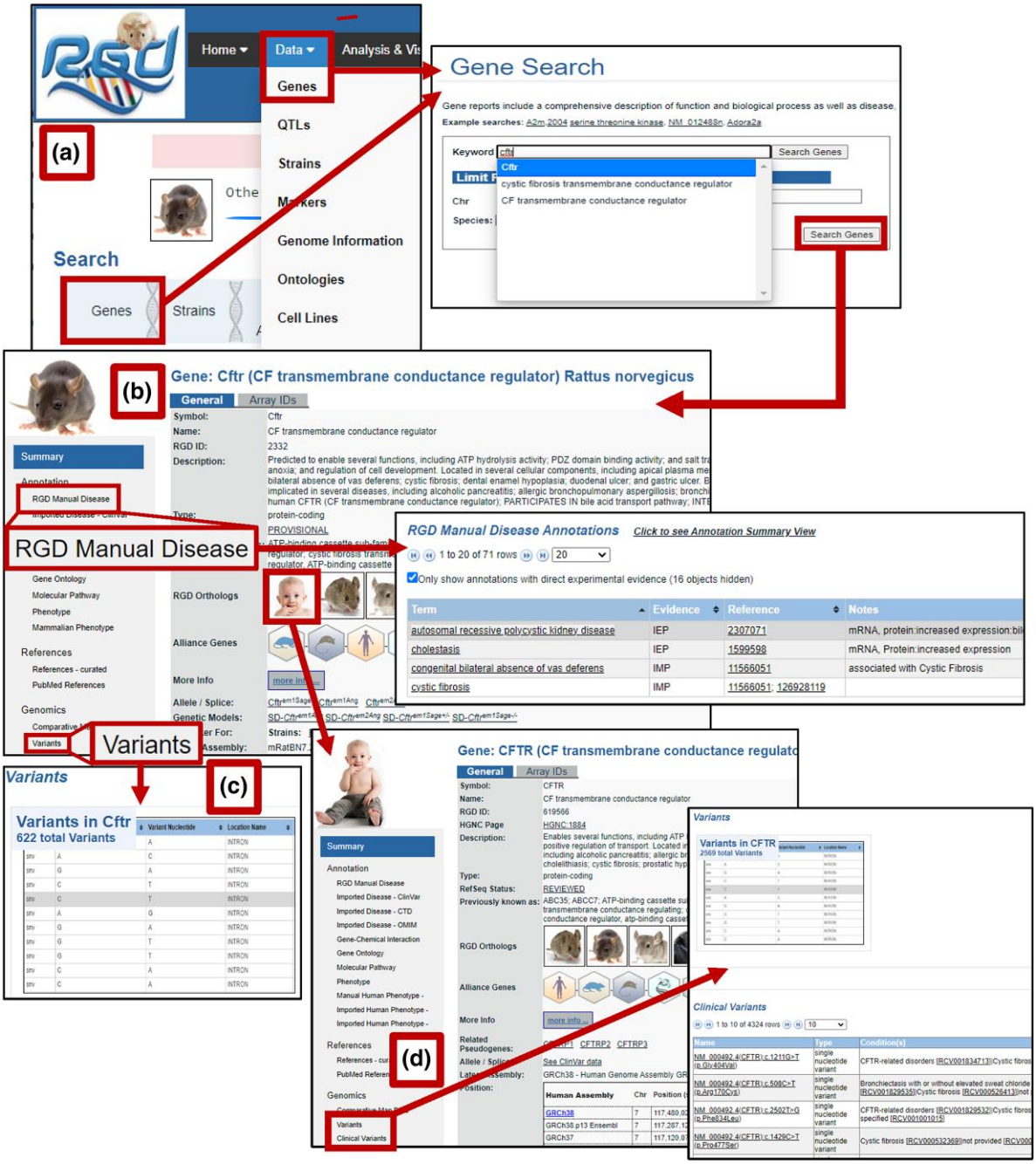

Fig. 6. RGD gene and strain search for cystic fibrosis disease. a) Indicates the RGD homepage Gene Search tool and the Data/Gene Search menu selections that are starting points for investigation of the *Cftr* gene. b) Indicates the gene report page (https://rgd.mcw.edu/rgdweb/report/gene/main.html?id=2332) and the menu option for the RGD manually curated disease annotations, including links to the referenced papers. c) Indicates the menu selection option of variants in the *Cftr* rat gene. d) Indicates the same menu selection after selecting the human gene report instead of rat. Supplementary Figure 4 provides an expanded view of the details of Fig. 6.

manipulated (dropped or added) using the report page filters. The bar colors can be changed using the dropdown boxes for "colored by" to reflect characteristics other than Condition, for example Phenotype. This change will update the legend. Sorting the table will also reorder the bars in the graph (Fig. 5f). It is now possible to see on the graph that, among the strains annotated for AML, phenotypes related to the susceptibility to induction of the disease, i.e. the percent of the study population developing either lymphoma or leukemia during the study period, vary greatly.

Example 2: cystic fibrosis

Another rare disease with considerable resources and data at RGD is cystic fibrosis (CF), which differs by having more specific identified causal and related genetic determinants. Approximately 40,000 people in the United States live with CF (https://www.cff.org/intro-cf/about-cystic-fibrosis). At RGD, this disease has 105 annotated genes in rat and three curated mutant rat strains. The rat strains are zinc-finger nuclease (ZFN) knockout rats with the *Cftr* gene as the target. Starting at RGD's homepage

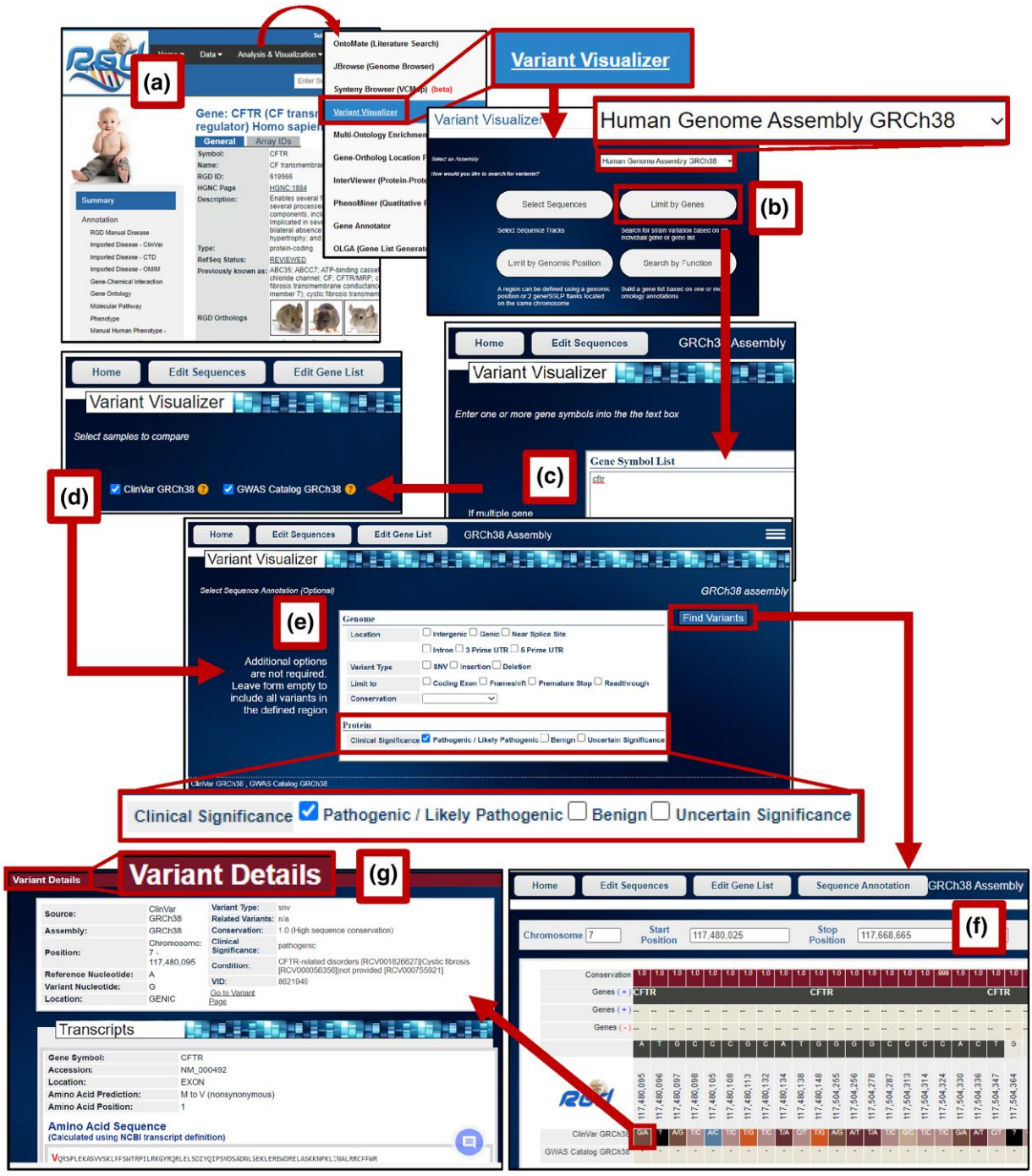

Fig. 7. RGD variant visualizer tool, *CFTR* example (https://rgd.mcw.edu/rgdweb/front/config.html). a) Indicates the RGD Analysis & Visualization menu dropdown for selecting the RGD tool Variant Visualizer. b), c), and d) Indicate the subsequent pages on which to select human, rat, or dog assembly and Limit by Genes to enter the gene symbol or a list, and select the sample in which to search for variants. e) Indicates selection of options, in this case to specify the clinical significance for variants listed in ClinVar as Pathogenic or Likely Pathogenic (https://rgd.mcw.edu/rgdweb/front/config.html?mapKey=38&geneList=cftr&chr=&start=&stop=&geneStart=&geneStop=&geneList=cftr&strain%5B%5D=2&sample1=2). f) Indicates the query result page, which displays the details for 624 pathogenic/likely pathogenic human variants in the *CFTR* gene with a scrollbar. g) Clicking on any one of the variants will display a pop-up with variant details. Supplementary Figure 5 provides an expanded view of the details of Fig. 7, and an additional explanation of rat variants in AML strains.

the same way as for AML, using the general search box, the Ontology and Annotation search tool (Fig. 3), or the Strain search tool, one can find genes and/or strains for cystic fibrosis. Alternatively, when proceeding through the Disease Portals landing page, for cystic fibrosis, one would select Respiratory Diseases. As noted in Fig. 1, strain report pages include phenotype annotations, which let the user see which aspects of the human disease are captured in the rat strain model.

Of the three mutant strains listed for cystic fibrosis, the two strains with "PM" icons in the result set have quantitative data available in the PhenoMiner repository. Selecting phenotypes of interest from the Clinical Measurements and generating the

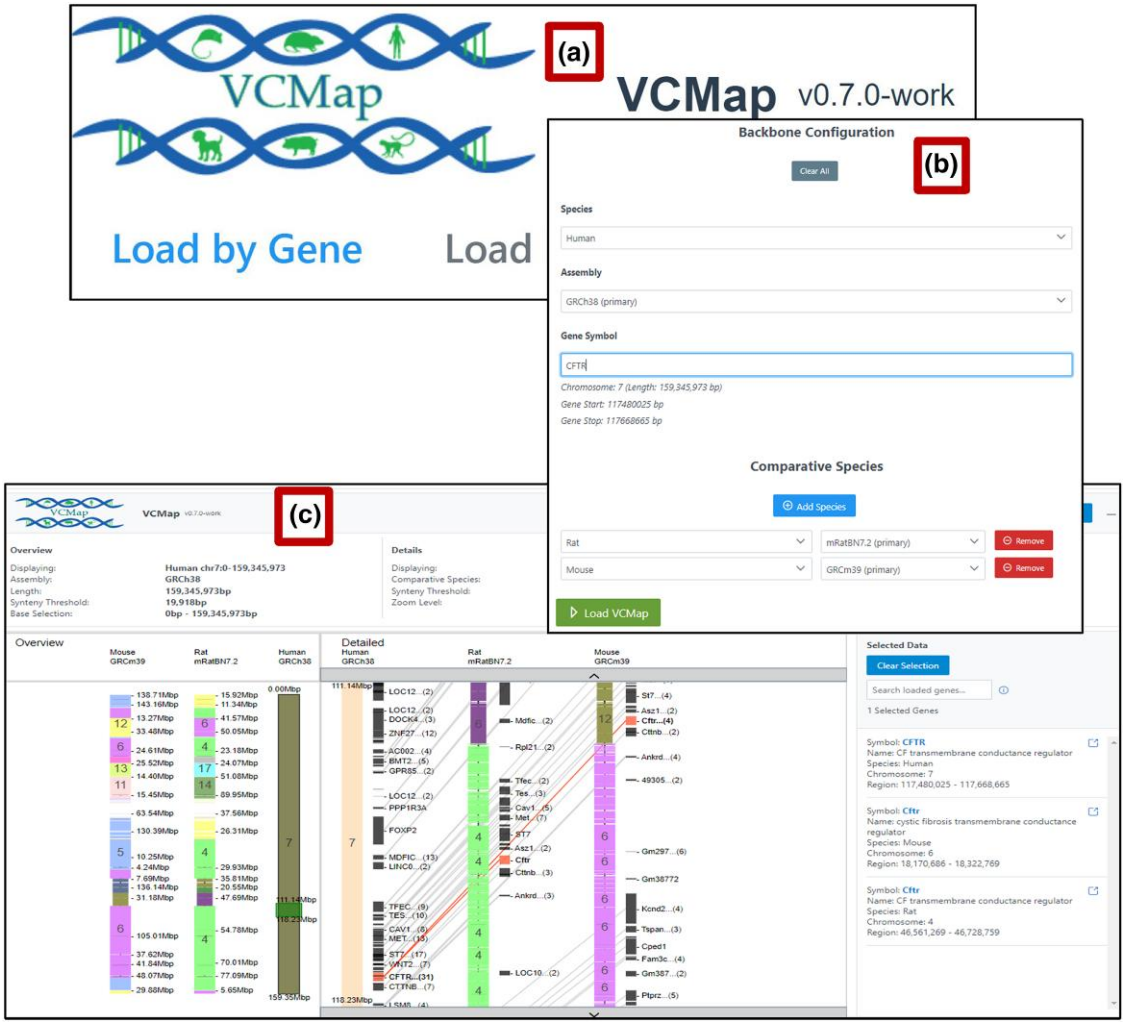

Fig. 8. RGD VCMap virtual comparative mapping (https://rgd.mcw.edu/vcmap/). a) Indicates the front page of the tool. b) Indicates the selection menu page. c) Indicates the results page showing multiple species' chromosome positions, based on sequence.

report allow visualization of the data, as well as a downloadable table for further study, just the same as the AML example in Fig. 5. Because the *Cftr* gene has been identified as critically involved in the incidence and development of cystic fibrosis, any investigation of this rare disease is likely to include research of the gene itself in addition to the utilization of strain models. Further investigation of the *Cftr* gene can begin on the homepage with the Gene Search box (Fig. 6a and Supplementary Figure 4), or from the dropdown menu under Data, selecting Genes. The gene report page (Fig. 6b) layout is very similar to the disease or strain report pages, simplifying the ability to find the information needed. All annotations for this gene available in RGD can be accessed from the gene report page, including variants in the *Cftr* rat gene (Fig. 6c). Without leaving the report page, the user can change to the human gene and see a set of human-centric annotations, curated references, and ClinVar variants (Fig. 6d).

The best way to view genetic variants is with the RGD Variant Visualizer tool. This tool is available from the Analysis & Visualization dropdown menu (Fig. 7 and Supplementary Figure 5). Working from the human gene page in this example, entering the gene in the Gene Symbol List and making selections as to assembly, samples, and genome parameters set up the search (Fig. 7b, c, d). It is also possible to specify the clinical significance to be specific for variants listed in ClinVar as Pathogenic or Likely Pathogenic (Fig. 7e). For human, the variants are from either the ClinVar or the genome-wide association studies (GWAS) Catalog (https://www.ebi.ac.uk/gwas/; Sollis *et al.* 2023). For rat and dog, the datasets represent rat strains and dog breeds, respectively. The catalog for associations to numerous diseases and RGD Variant Visualizer are regularly updated. The details for human variants in the *CFTR* gene are displayed with a scrollbar (Fig. 7f). Clicking on any one of the variants will display a pop-up with variant details, including pathogenicity (Fig. 7g).

Rat model strains can be useful for elucidating disease mechanisms, and switching the Variant Visualizer focus from human to the rat assembly allows a researcher to find inbred sequenced strains that have variants in the *Cftr* gene. Studying cystic fibrosis phenotypes in these strains might be key to understanding the variant effects, and perhaps even potential therapeutic strategies for humans.

While *CFTR* is the primary gene of interest for cystic fibrosis, the other example rare disease, AML, does not have specific causal genes identified. The research shown above implemented a cassette of RI strains to interrogate levels of susceptibility. Entering the genes annotated to AML into Variant Visualizer, selecting the LEXF recombinant inbred strains that are available and the two parent strains, and specifying Possibly/Probably Damaging as a search parameter show ~25 genes for which one or more of

the strains has a variant (computationally) predicted to be damaging (Supplementary Fig. 5). Comparing the strains with variants to the list of strains with increased or decreased susceptibility to AML induction protocol (Supplementary Fig. 5), it may be noted that those strains having possibly damaging variants have some of the highest rates of T-cell lymphoma, yet lowest rates of leukemia, upon induction.

Biomedical research frequently uses information from one species to understand the same processes in another. Insights into molecular pathways can be gleaned by comparing human, rat, and mouse utilizing comparative genomics. Genetic maps have been used successfully to find genes responsible for inherited disorders such as cystic fibrosis. While DNA markers often do not identify the gene responsible for the disease or trait, they do provide a rough indication of where the locus is on the chromosome. If a particular gene is close to a DNA marker, the gene and marker will likely stay together during the recombination process and be passed on together from one generation to the next. Cross-species gene mapping can assist with mapping resolution and identification of conserved gene regions and networks. A recent update to the RGD tool arsenal is Virtual Comparative Mapping (VCMap, Fig. 8), currently under active development and released as a beta-testing version on the RGD website. VCMap is a powerful way to view cross-species positions for genes and loci. Searching for human *CFTR* and adding comparative species of rat and mouse load a map of synteny for the region surrounding the gene of interest. The comparative map can also be explored using chromosome positions for the backbone species, using different assemblies, and/or swapping which species is the backbone/reference for the comparison. Comparative mapping helps to illuminate genes in common related to biological systems, and may translate into improvement in human health via innovative treatments (Kaldunski *et al.* 2022); (https://www.genome.gov/about-genomics/fact-sheets/Comparative-Genomics-Fact-Sheet; https://www.sciencedirect.com/topics/neuroscience/comparative-genomics).

Knowing a human gene or locus can allow one to utilize comparative mapping to locate the gene or region in rat or mouse and thus help determine an appropriate animal model for study. An example of this is illustrated in Heaney *et al.* (1998) in which it was found that given a similarity between human and mouse phenotypes for osteopetrosis/osteosclerosis, they found that the responsible genetic positions mapped to a region of conserved synteny.

Conclusion

RGD provides the disease research community with a wealth of rare disease information, relevant strains as research models, disease, gene, and strain report pages with consolidated data, and an arsenal of analysis tools with which to further biomedical research. Leveraging these resources for rare disease mechanism discovery, and potentially the discovery of new treatment modalities, will help further clinical progress for those afflicted with these diseases. RGD's data and tools are globally accessible. Information on how to contact us for questions or research assistance, including RGD virtual office hours available by appointment, can be found at our website https://rgd.mcw.edu.

Data availability

RGD abides by and implements FAIR data practices, with all information freely available under a CC BY 4.0 license. RGD is a Global Core Biodata Resource (Global Biodata Coalition 2022). The datasets and computer tools discussed in this paper are either publicly available at RGD or available from the corresponding author on request. Full datasets can be obtained from the RGD download site at https://download.rgd.mcw.edu/data_release/. Datasets that are the result of queries of repository data from within RGD tools are downloadable at the time using the Excel icon that appears next to gene, phenotype, QTL, strain, etc. lists. Tools software is available on our GitHub website (https://github.com/rat-genome-database).

Supplemental material available at GENETICS online.

Funding

RGD is grateful for funding support from the National Heart, Lung, and Blood Institute (R01HL064541) on behalf of the National Institutes of Health (NIH) and from the National Human Genome Research Institute (NHGRI) as part of the Alliance of Genome Resources (U24HG010859).

Conflicts of interest statement

The author(s) declare no conflict of interest.

Literature cited

D'Andrea AD. Susceptibility pathways in fanconi's anemia and breast cancer. N Engl J Med. 2010;362(20):1909–1919. doi:10.1056/NEJMra0809889.

Dozzo A, Galvin A, Shin JW, Scalia S, O'Driscoll CM, Ryan KB. Modelling acute myeloid leukemia (AML): what's New? A transition from the classical to the modern. Drug Deliv Transl Res. 2022. doi:10.1007/s13346-022-01189-4.

Ehrhart F, Willighagen EL, Kutmon M, van Hoften M, Curfs LMG, Evelo CT. A resource to explore the discovery of rare diseases and their causative genes. Sci Data. 2021;8(1):124. doi:10.1038/s41597-021-00905-y.

Fang EF, Kassahun H, Croteau DL, Scheibye-Knudsen M, Marosi K, Lu H, Shamanna RA, Kalyanasundaram S, Bollineni RC, Wilson MA, *et al.* Nad(+) replenishment improves lifespan and healthspan in ataxia telangiectasia models via mitophagy and DNA repair. Cell Metab. 2016;24(4):566–581. doi:10.1016/j.cmet.2016.09.004.

Feinberg AP, Williams BR. Wilms' tumor as a model for cancer biology. Methods Mol Biol. 2003;222:239–248. doi:10.1385/1-59259-328-3:239.

Gene Ontology Consortium. The gene ontology resource: enriching a gold mine. Nucleic Acids Res. 2021;49(D1):D325–d334. doi:10.1093/nar/gkaa1113.

Global Biodata Coalition. Global core biodata resources: Concept and selection process 2022. 105281/zenodo5845115.

Hagenbeek A, Martens AC. The pathogenesis of a rat model for human acute myelocytic leukemia. Haematologica. 1980;65(3):293–308.

Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian inheritance in man (omim). Hum Mutat. 2000;15(1):57–61. doi:10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G.

Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. Chebi in 2016: improved services and an expanding collection of metabolites. Nucleic Acids Res. 2016;44(D1):D1214–D1219. doi:10.1093/nar/gkv1031.

Hayman GT, Laulederkind SJ, Smith JR, Wang SJ, Petri V, Nigam R, Tutaj M, De Pons J, Dwinell MR, Shimoyama M. The disease portals, disease-gene annotation and the rgd disease ontology at the

rat genome database. Database (Oxford). 2016;2016:baw034. doi: 10.1093/database/baw034.

Heaney C, Shalev H, Elbedour K, Carmi R, Staack JB, Sheffield VC, Beier DR. Human autosomal recessive osteopetrosis maps to 11q13, a position predicted by comparative mapping of the murine osteosclerosis (oc) mutation. Hum Mol Genet. 1998;7(9): 1407–1410. doi:10.1093/hmg/7.9.1407.

Institute of Medicine (US) Committee on Accelerating Rare Diseases Research and Orphan Product Development. Rare diseases and orphan products: Accelerating research and development Washington (DC): National Academies Press (US); Copyright © 2010, National Academy of Sciences; 2010.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-rna-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816–821. doi:10.1126/science.1225829.

Kaldunski ML, Smith JR, Hayman GT, Brodie K, De Pons JL, Demos WM, Gibson AC, Hill ML, Hoffman MJ, Lamers L, *et al.* The rat genome database (RGD) facilitates genomic and phenotypic data integration across multiple species for biomedical research. Mamm Genome. 2022;33(1):66–80. doi:10.1007/s00335-021-09932-x.

Kee Y, D'Andrea AD. Expanded roles of the fanconi anemia pathway in preserving genomic stability. Genes Dev. 2010;24(16): 1680–1694. doi:10.1101/gad.1955310.

Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, *et al.* The human phenotype ontology in 2021. Nucleic Acids Res. 2021;49(D1):D1207–d1217. doi:10.1093/nar/gkaa1043.

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, *et al.* Clinvar: improvements to accessing data. Nucleic Acids Res. 2020;48(D1):D835–d844. doi:10.1093/nar/gkz972.

Laulederkind SJ, Liu W, Smith JR, Hayman GT, Wang SJ, Nigam R, Petri V, Lowry TF, de Pons J, Dwinell MR, *et al.* Phenominer: quantitative phenotype curation at the rat genome database. Database (Oxford). 2013;2013:bat015. doi:10.1093/database/bat015.

Lee JA, Cho A, Huang EN, Xu Y, Quach H, Hu J, Wong AP. Gene therapy for cystic fibrosis: new tools for precision medicine. J Transl Med. 2021;19(1):452. doi:10.1186/s12967-021-03099-4.

Lu LM, Shisa H, Tanuma J, Hiai H. Propylnitrosourea-induced T-lymphomas in LEXF RI strains of rats: genetic analysis. Br J Cancer. 1999;80(5–6):855–861. doi:10.1038/sj.bjc.6690432.

Martens AC, Van Bekkum DW, Hagenbeek A. The bn acute myelocytic leukemia (BNML) (a rat model for studying human acute myelocytic leukemia (AML)). Leukemia. 1990;4(4):241–257.

Mashimo T, Voigt B, Kuramoto T, Serikawa T. Rat phenome project: the untapped potential of existing rat strains. J Appl Physiol (1985). 2005;98(1):371–379. doi:10.1152/japplphysiol.01006.2004.

McCormack E, Bruserud O, Gjertsen BT. Animal models of acute myelogenous leukaemia—development, application and future perspectives. Leukemia. 2005;19(5):687–706. doi:10.1038/sj.leu.2403670.

Ménoret S, Tesson L, Rémy S, Usal C, Thépenier V, Thinard R, Ouisse LH, De Cian A, Giovannangeli C, Concordet JP, *et al.* Gene targeting in rats using transcription activator-like effector nucleases. Methods. 2014;69(1):102–107. doi:10.1016/j.ymeth.2014.02.027.

Nigam R, Munzenmaier DH, Worthey EA, Dwinell MR, Shimoyama M, Jacob HJ. Rat strain ontology: structured controlled vocabulary designed to facilitate access to strain data at RGD. J Biomed Semantics. 2013;4(1):36. doi:10.1186/2041-1480-4-36.

Pacak CA, Brooks PJ. The past, present, and future of modeling cockayne syndrome—a commentary on "rat model of cockayne syndrome neurological disease". DNA Repair (Amst). 2020;88: 102788. doi:10.1016/j.dnarep.2020.102788.

Park CA, Bello SM, Smith CL, Hu ZL, Munzenmaier DH, Nigam R, Smith JR, Shimoyama M, Eppig JT, Reecy JM. The vertebrate trait ontology: a controlled vocabulary for the annotation of trait data across species. J Biomed Semantics. 2013;4(1):13. doi:10.1186/2041-1480-4-13.

Petri V, Jayaraman P, Tutaj M, Hayman GT, Smith JR, De Pons J, Laulederkind SJ, Lowry TF, Nigam R, Wang SJ, *et al.* The pathway ontology—updates and applications. J Biomed Semantics. 2014; 5(1):7. doi:10.1186/2041-1480-5-7.

Pogue RE, Cavalcanti DP, Shanker S, Andrade RV, Aguiar LR, de Carvalho JL, Costa FF. Rare genetic diseases: update on diagnosis, treatment and online resources. Drug Discov Today. 2018;23(1): 187–195. doi:10.1016/j.drudis.2017.11.002.

Quek H, Luff J, Cheung K, Kozlov S, Gatei M, Lee CS, Bellingham MC, Noakes PG, Lim YC, Barnett NL, *et al.* A rat model of ataxia-telangiectasia: evidence for a neurodegenerative phenotype. Hum Mol Genet. 2017;26(1):109–123. doi:10.1093/hmg/ddw371.

Rafeeq MM, Murad HAS. Cystic fibrosis: current therapeutic targets and future approaches. J Transl Med. 2017;15(1):84. doi:10.1186/s12967-017-1193-9.

Sato M, Nakamura S, Inada E, Takabayashi S. Recent advances in the production of genome-edited rats. Int J Mol Sci. 2022;23(5):2548. doi:10.3390/ijms23052548.

Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, Baron JA, Jackson R, Bello SM, Bearer C, *et al.* The human disease ontology 2022 update. Nucleic Acids Res. 2022;50(D1):D1255–d1261. doi:10.1093/nar/gkab1063.

Skayneh H, Jishi B, Hleihel R, Hamieh M, Darwiche N, Bazarbachi A, El Sabban M, El Hajj H. A critical review of animal models used in acute myeloid leukemia pathophysiology. Genes (Basel). 2019; 10(8):614. doi:10.3390/genes10080614.

Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. J Biomed Semantics. 2015;6(1):11. doi:10.1186/s13326-015-0009-1.

Smith JR, Hayman GT, Wang S-J, Laulederkind SJF, Hoffman MJ, Kaldunski ML, Tutaj M, Thota J, Nalabolu HS, Ellanki SLR, *et al.* The year of the rat: the rat genome database at 20: a multi-species knowledgebase and analysis platform. Nucleic Acids Res. 2020; 48(D1):D731–D742. doi:10.1093/nar/gkz1041.

Smith JR, Park CA, Nigam R, Laulederkind SJ, Hayman GT, Wang SJ, Lowry TF, Petri V, Pons JD, Tutaj M, *et al.* The clinical measurement, measurement method and experimental condition ontologies: expansion, improvements and new applications. J Biomed Semantics. 2013;4(1):26. doi:10.1186/2041-1480-4-26.

Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J, *et al.* The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51-(D1):D977–d985. doi:10.1093/nar/gkac1010.

van Boxtel R, Gould MN, Cuppen E, Smits BM. Enu mutagenesis to generate genetically modified rat models. Methods Mol Biol. 2010;597:151–167. doi:10.1007/978-1-60327-389-3_11.

Vaughn WP, Burke PJ, Jung J. BN rat myeloid leukemia transferred to the (LEW x BN)F1 rat. J Natl Cancer Inst. 1978;61(3):927–929.

Vedi M, Nalabolu HS, Lin CW, Hoffman MJ, Smith JR, Brodie K, De Pons JL, Demos WM, Gibson AC, Hayman GT, *et al.* MOET: a web-based gene set enrichment tool at the rat genome database for multiontology and multispecies analyses. Genetics. 2022;220(4): iyac005. doi:10.1093/genetics/iyac005.

Voigt B, Kuramoto T, Mashimo T, Tsurumi T, Sasaki Y, Hokao R, Serikawa T. Evaluation of LEXF/FXLE rat recombinant inbred

strains for genetic dissection of complex traits. Physiol Genomics. 2008;32(3):335–342. doi:10.1152/physiolgenomics.00158.2007.

Wang SJ, Laulederkind SJ, Hayman GT, Petri V, Smith JR, Tutaj M, Nigam R, Dwinell MR, Shimoyama M. Comprehensive coverage of cardiovascular disease data in the disease portals at the rat genome database. Physiol Genomics. 2016;48(8):589–600. doi:10.1152/physiolgenomics.00046.2016.

Webster ALH, Sanders MA, Patel K, Dietrich R, Noonan RJ, Lach FP, White RR, Goldfarb A, Hadi K, Edwards MM, *et al.* Genomic signature of Fanconi anaemia DNA repair pathway deficiency in cancer. Nature. 2022;612(7940):495–502. doi:10.1038/s41586-022-05253-4.

Xu Y, Wu Z, Liu L, Liu J, Wang Y. Rat model of cockayne syndrome neurological disease. Cell Rep. 2019;29(4):800–809.e805. doi:10.1016/j.celrep.2019.09.028.

Zhao M, Wei DQ. Rare diseases: drug discovery and informatics resource. Interdiscip Sci. 2018;10(1):195–204. doi:10.1007/s12539-017-0270-3.

Editor: T. Harris