

# Rat Genome and Model Resources

Mary Shimoyama, Jennifer R. Smith, Elizabeth Bryda, Takashi Kuramoto, Laura Saba, and Melinda Dwinell

Mary Shimoyama, PhD, is an associate professor in the Department of Biomedical Engineering at Marquette University and the Medical College of Wisconsin in Milwaukee, Wisconsin. Jennifer R. Smith, MS, is a biocurator at the Rat Genome Database in the Department of Biomedical Engineering at Marquette University and the Medical College of Wisconsin in Milwaukee, Wisconsin. Elizabeth Bryda, PhD, is a professor in the Department of Veterinary Pathobiology at the College of Veterinary Medicine, University of Missouri in Columbia, Missouri. Takashi Kuramoto, PhD, is an associate professor in the Institute of Laboratory Animals at the Graduate School of Medicine, Kyoto University in Kyoto, Japan. Laura Saba, PhD, is an assistant professor in the Department of Pharmaceutical Sciences at in the Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus in Aurora, Colorado. Melinda Dwinell, PhD, is an associate professor in the Department of Physiology at the Medical College of Wisconsin in Milwaukee, Wisconsin.

Address correspondence to Dr. Mary Shimoyama, Department of Biomedical Engineering, Marquette University and Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, or email [shimoyama@mcw.edu](mailto:shimoyama@mcw.edu).

## Abstract

Rats remain a major model for studying disease mechanisms and discovery, validation, and testing of new compounds to improve human health. The rat's value continues to grow as indicated by the more than 1.4 million publications (second to human) at PubMed documenting important discoveries using this model. Advanced sequencing technologies, genome modification techniques, and the development of embryonic stem cell protocols ensure the rat remains an important mammalian model for disease studies. The 2004 release of the reference genome has been followed by the production of complete genomes for more than two dozen individual strains utilizing NextGen sequencing technologies; their analyses have identified over 80 million variants. This explosion in genomic data has been accompanied by the ability to selectively edit the rat genome, leading to hundreds of new strains through multiple technologies. A number of resources have been developed to provide investigators with access to precision rat models, comprehensive datasets, and sophisticated software tools necessary for their research. Those profiled here include the Rat Genome Database, PhenoGen, Gene Editing Rat Resource Center, Rat Resource and Research Center, and the National BioResource Project for the Rat in Japan.

**Key words:** bioinformatics; database; disease; genomics; phenotype; rat; *Rattus norvegicus*; resource

## Introduction

Rats remain a major model for studying disease mechanisms and discovery, validation, and testing of new compounds to

improve human health. The rat's value continues to grow as indicated by the more than 1.4 million publications (second to human) at PubMed documenting important discoveries using

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

this model. Advanced sequencing technologies, genome modification techniques, and the development of embryonic stem cell protocols ensure the rat remains an important mammalian model for disease studies. The 2004 release of the reference genome has been followed by the production of complete genomes for more than two dozen individual strains utilizing NextGen sequencing technologies; their analyses have identified over 80 million variants (Atanur et al. 2013; Baud et al. 2013). This explosion in genomic data has been accompanied by the ability to selectively edit the rat genome, leading to hundreds of new strains through technologies using the CRISPR/Cas9 system (Li et al. 2013), zinc finger nucleases (Geurts and Moreno), transcription activator-like effector nucleases (Tong et al. 2012), transposons (Carlson et al. 2011), and meganucleases (Menoret et al. 2013) as well as embryonic stem cells (Buehr et al. 2008) in the rat. A number of resources have been developed to provide investigators with access to precision rat models, comprehensive datasets, and sophisticated software tools necessary for their research. Those profiled here include the Rat Genome Database, PhenoGen, Gene Editing Rat Resource Center, Rat Resource and Research Center, and the National BioResource Project for the Rat in Japan.

## Rat Resources

### Rat Genome Database

The Rat Genome Database (RGD, <http://rgd.mcw.edu>) is the premier site for rat genomic and phenotype data. It also provides a sophisticated suite of tools for mining, analyzing, and visualizing this data. RGD is also responsible for providing official nomenclature for rat genes, QTL, and strains. RGD provides a critical research platform that also provides easy access to comprehensive human and mouse data, including genes, variants, QTL, and interactions and the integration of these data into multiple software tools. This is particularly important to those involved in translational research into disease mechanisms, pharmacology, and toxicology, in which cross-organism comparisons and analyses are vital. As Table 1 illustrates, RGD provides a comprehensive catalogue of genomic elements for three organisms, interacting with major resources such as NCBI (Tatusova 2016), Ensembl (Aken et al. 2016), mirBase (Kozomara and Griffiths-Jones 2014), and UniProt (Pundir et al. 2015). RGD also continuously updates datasets for genomic elements, including revised annotations and map positions on new genome assemblies, and provides confirmation and validation of genomic element identity and improved, validated ortholog assignments through its quality control pipelines and manual curation efforts. These validated and adjusted data are returned to originating sources to ensure synchronization of rat data worldwide.

Increasing the value of these genomic elements are the functional annotations attached to them. RGD uses a combination of manual curation and data import pipelines to provide

**Table 1** Counts of genomic elements at RGD

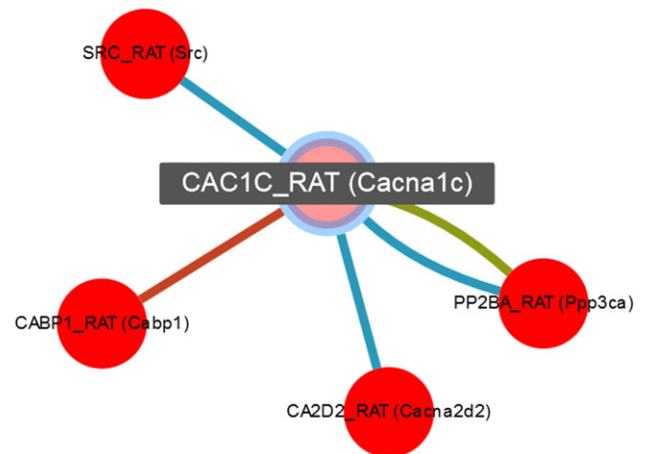
	Rat	Human	Mouse
Protein coding Genes	24,632	19,338	26,141
Pseudogenes	9738	13,120	9751
Promoters	119	60,382	53,674
Transcripts	171,437	187,031	175,919
microRNAs	20,218	31,497	30,749
tRNAs	470	537	412
QTL	2608	1911	5163

researchers with extensive functional information as illustrated by the types and number of annotations for genes shown in Table 2. While the Gene Ontology annotations for rat, along with many of the pathway, disease, and phenotype annotations, are manually curated from published literature, the drug/chemical interaction data is imported from the Comparative Toxicogenomics Database (Davis et al. 2016) and other annotations are also acquired from ClinVar (NCBI Resource Coordinators 2016), GOA (Huntley et al. 2015), MGD (Bult et al. 2016), and KEGG (Kanehisa et al. 2016). RGD also provides data and tools for the exploration of protein-protein interactions, microRNAs and their targets, and pathways. Users can explore protein-protein interactions through its InterViewer tool (Figure 1). Input data types include UniProtKB IDs, RGD Gene IDs, and Gene Symbols. Users can choose to see the interactions in a single organism (rat, human, mouse) or to view the interactions for all three organisms at the same time. Interaction types are indicated and illustrated with different colors and users can choose from multiple types of visualizations. A table is presented with the protein interactors, their symbols, associated gene symbols, association type, and identifiers. Users can also download all of the data presented.

RGD's pathway portal (Hayman et al. 2013; Petri et al. 2011, 2014a, 2014b, 2016) incorporates interactive diagrams and

**Table 2** Number of functional annotations for rat, human, and mouse genes

	Rat	Human	Mouse
GO: Molecular Function	138,019	141,376	114,763
GO: Biological process	207,810	141,890	150,741
GO: Cellular Component	122,480	133,113	102,690
Disease	84,727	165,592	84,482
Phenotype	1557	119,569	11,489
Pathway	30,313	29,400	28,932
Drug/chemical Interaction	757,148	773,753	761,797



**Figure 1** RGD's InterViewer tool. The InterViewer is a visualization tool for protein-protein interactions. In this example, a UniProt identifier for the rat Cacna1c gene, P22002, was entered. The results show four interaction partners for this protein. The red color of the nodes indicates that these are all rat proteins. The colors of the edges, that is, the lines between the nodes, indicate that there are three types of interactions: physical association, dephosphorylation reaction, and colocalization. Additional information about the interactors and their interactions is available within the tool and via links to other pages at RGD and other external sites.

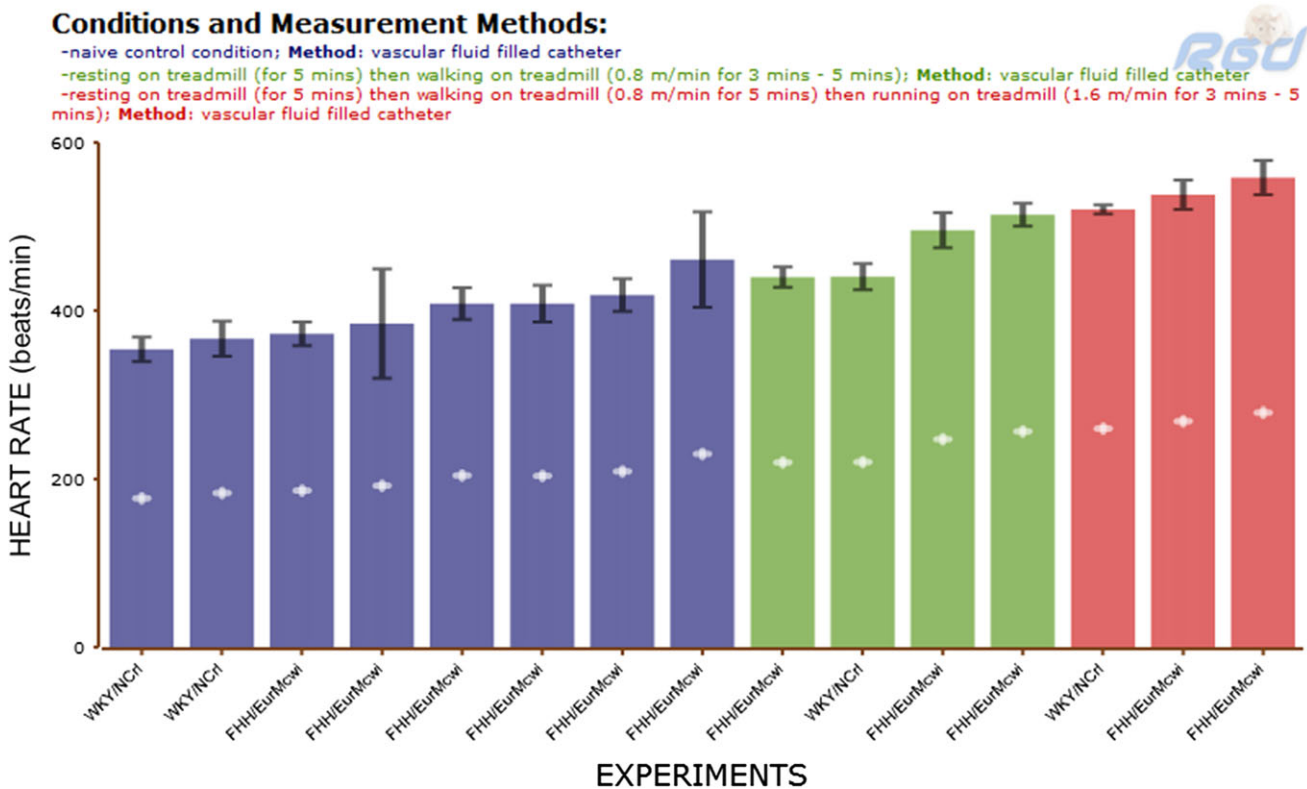
summary data for metabolic, signaling, regulatory, drug, and disease pathways. Each diagram provides dynamic links to reports for each participating gene and links to interacting pathways and their diagrams. A comprehensive description of the pathway including the players, targets, and function of the pathway is also provided along with a reference list with links to the papers. Researchers can also view the disease and additional pathway annotations for each of the genes involved and can also download this data. Pathway Suites and Networks are a unique feature that provides easy access to interacting and related pathways associated with particular physiological traits, diseases, and complex molecular processes.

Much of the value of the rat as a model organism lies with the wide variety of available strains, including inbred and manipulated strains (Table 3). RGD strain reports include

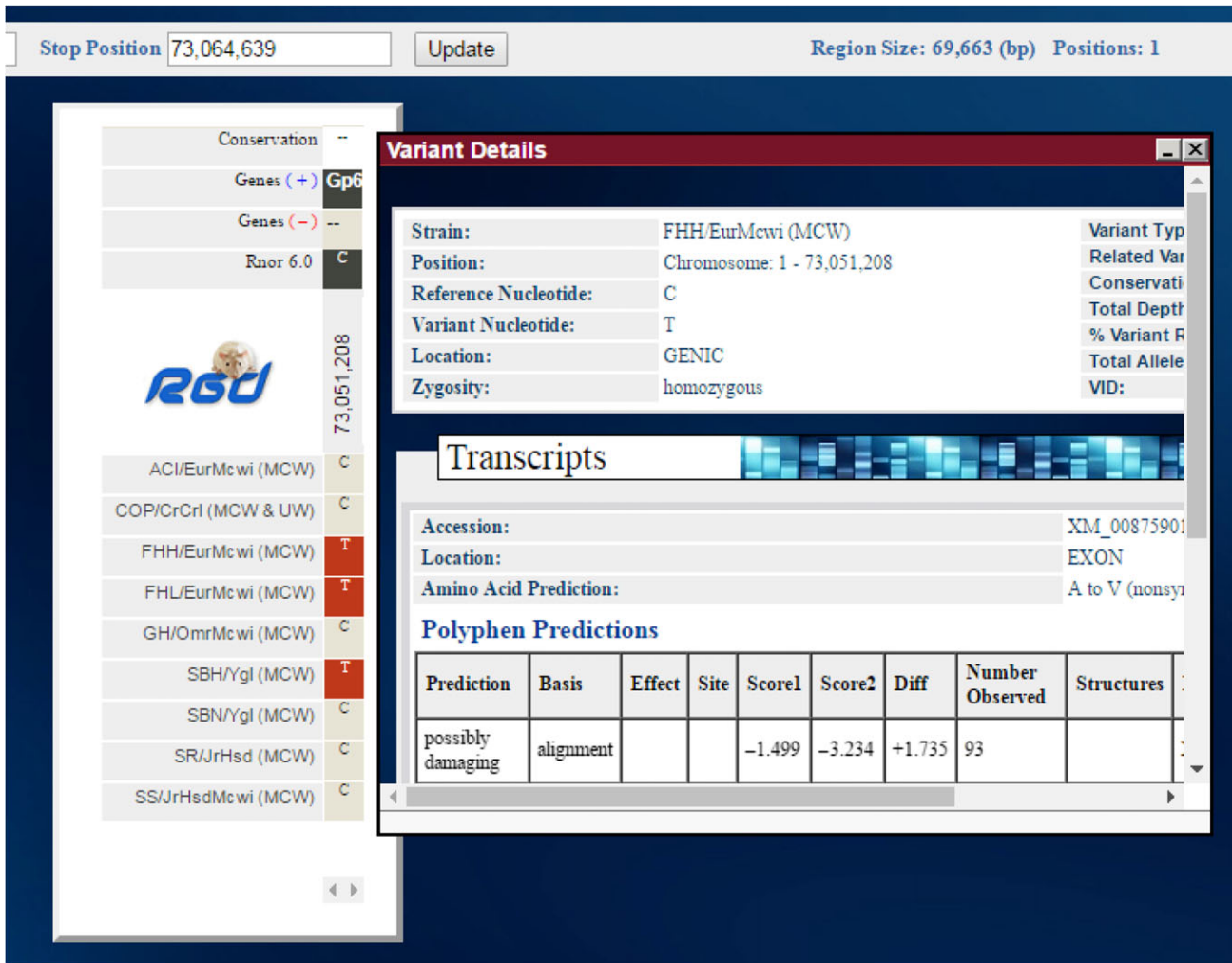
**Table 3** Number of rat strains at RGD by strain type

Strain type	Number
Inbred	738
Outbred	60
Consomic	91
Congenic	1200
Recombinant Inbred	131
Segregating inbred	13
Mutant	706
Transgenic	247

information on the source and availability of the strain, type of manipulation used to derive the strain, related substrains, and QTL identified in experiments using the strain. RGD provides two tools for researchers to identify the best precision model for their studies. The first of these, PhenoMiner (Laulederkind et al. 2013; Shimoyama et al. 2012; Smith et al. 2013; Wang et al. 2015) incorporates quantitative measurements from multiple experiments and currently houses over 60,000 records covering a wide variety of phenotypes, including those related to the cardiovascular, respiratory, endocrine, renal, immune, musculoskeletal, and nervous systems as well as reproductive, morphological, and behavioral traits. Researchers can easily query and retrieve data based on strain, measurement type, method used, and experimental conditions employed. Results are presented in charts and downloadable table formats (Figure 2). Accompanying this phenotypic profile of each strain is the variant profile provided by the Variant Visualizer tool (Figure 3). The Variant Visualizer accepts gene lists, as well as queries by function or genomic position. The user can filter results by type of amino acid change, predicted effect, and structural location. Results are visualized with detailed information on the predicted effects and location of each variant. Variants from a wide variety of strains are also presented in tracks on JBrowse, RGD's genome browser. In addition, JBrowse provides tracks for mutant and congenic strains, RNA-Seq data, and QTL. Users can view disease-specific tracks as well as those for chemical/drug-gene interactions. RGD also has human and mouse JBrowse tools to provide easy cross-species comparisons.



**Figure 2** Comparison of heart rates across strains and experiments. RGD's PhenoMiner tool allows users to select one or more measurements, rat strains, measurement methods, and experimental conditions of interest, then compare the results across the selected strains and conditions, not only within a single study but across multiple studies. Here, heart rates for the WKY/NCrl and FHH/EurMcwi strains are compared under naïve control conditions after walking at 0.8 m/min for 5 minutes and after walking for 5 minutes then running at 1.6 m/min for 5 minutes. A tabular view of the results (not shown) is also available to view and download.



**Figure 3** Variant Visualizer showing damaging variants in gene across strains. RGD’s Variant Visualizer tool leverages the whole genome sequencing data for a number of rat strains to provide variant profiles for genes or genomic regions across any or all of the sequenced strains. Here a variant predicted to be “possibly damaging” for protein functions is shown to be present in the FHH, FHL, and SBH strains but not in the other strains selected. Clicking on the square corresponding to the strain and variant of interest opens a detail window that shows information about the sequencing, the variant(s) called, and the predicted consequence of the sequence change.

Providing easy access to important datasets, navigation among data types and tools, and analysis of retrieved data forms the focus of RGD’s continuing expansions and innovations. The Object List Generator & Analyzer (OLGA) assists researchers in assembling datasets and subsets easily (Figure 4). Users can search for and retrieve rat, human, or mouse gene sets based on functional similarities (disease, phenotype, pathway, Gene Ontology, or drug/chemical-gene interactions) by entering a genomic region or QTL symbol to obtain a list of overlapping genes or by uploading their own list of genes. Once a gene list is generated, the researcher is given the option to retrieve a second gene list based on other search criteria. Users can then join the lists to make a composite list, subtract one list from another, or create a gene subset from an intersection of the two datasets. Users can continue to retrieve datasets and create additional groupings of data elements. At any point in the process, users can download the active result set or send it to a variety of tools for visualization and analysis (Figure 4, top). These include the InterViewer to identify protein-protein interactions for the final result list, the Variant Visualizer to identify damaging variants across strains for these genes, and

a Genome Viewer that allows users to visualize the genes relative to the karyotype of their species of interest. In addition, the user can send the dataset to three functions of the Gene Annotator tool: Functional Annotation, Annotation Distribution, and Annotation Comparison. Functional Annotation provides a comprehensive gene report for each of the genes within the list. The downloadable data report includes IDs for and links to external databases for the gene, ortholog identification, and functional annotations, including those for disease, pathway, phenotype, Gene Ontology, and drug/chemical-gene interactions for the gene and its orthologs. The Annotation Distribution option provides a functional overview of the gene set itself, indicating the percentage of genes within the list associated with each type of functional classification. The user can then retrieve the subsets of genes in each of these categories, or the genes associated with two or more categories, and analyze them further. The Annotation Comparison function (Figure 4, bottom) presents a heatmap visualization of the gene set. Users select the functional domains, such as disease, pathway, etc., to display on the X and Y axes. The heatmap shows how the genes in the originating set are distributed across the major categories of

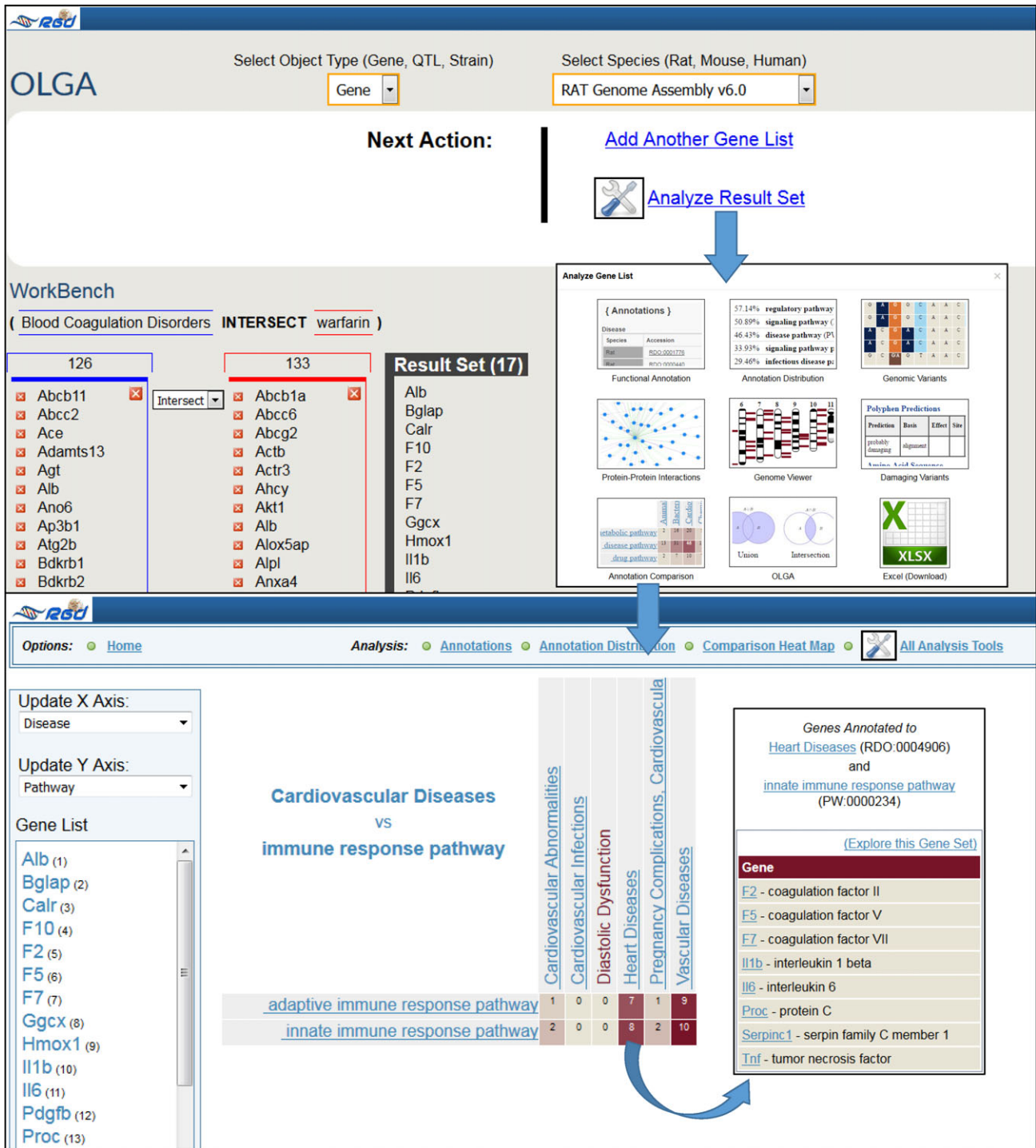


Figure 4 RGD’s OLGA. The OLGA tool is a powerful and flexible advanced query engine that can be used for bulk queries or searches by functional attributes or genomic position for rat, mouse and human genes and QTL, and rat strains. At each step, a list of objects, in this case genes, is produced and the user can choose how to combine the lists. In the example on the top, the user selected to view the intersection between lists of rat genes associated with blood coagulation disorders and genes that interact with the drug warfarin. The user then has the option to add another gene list to the current result set or to analyze the current list. Selecting “Annotation Comparison” in the toolbox takes the user to the Gene Annotator (GA) Tool with their list of 17 genes already displayed in a comparison heatmap which compares disease and pathway associations for the genes in the list. By selecting more specific disease and pathway terms, the user retrieves the list of eight genes that are associated with blood coagulation disorders and heart diseases, are involved in the innate immune response pathway, and interact with warfarin.

these domains. Users can “drill down” further by clicking on a term on either axis to see the next layer of more specific categories under that term. The heatmap now shows how the subset of genes from the higher category are distributed among its

subcategories. A click on any of the squares in the heatmap will retrieve the subset of genes associated with the corresponding terms on the X and Y axes. This new list of genes can then be further explored with any of the tools. As such, these tools allow

researchers to easily identify and analyze datasets without needing to download and reformat data.

One of the newest innovations on the RGD website is MyRGD. With this tool, a researcher can create a free online account and receive notifications of updates to data of most interest to them. After creating an account, the user can click on the binoculars icon found throughout the website to “watch” genes, QTL, strains, or functional categories such as diseases or pathways. The investigator will then receive weekly notifications of any changes or additions to the data associated with those elements or annotation types. For example, if a new gene has been identified as being associated with arthritis, a researcher watching either the term arthritis or that specific gene will receive a notification. Alternatively, if a gene has had a functional annotation, new map position, or sequence attached to it, researchers following that gene will receive notification in a weekly email. Users can add to or delete items from their account any time to customize as their research expands and changes.

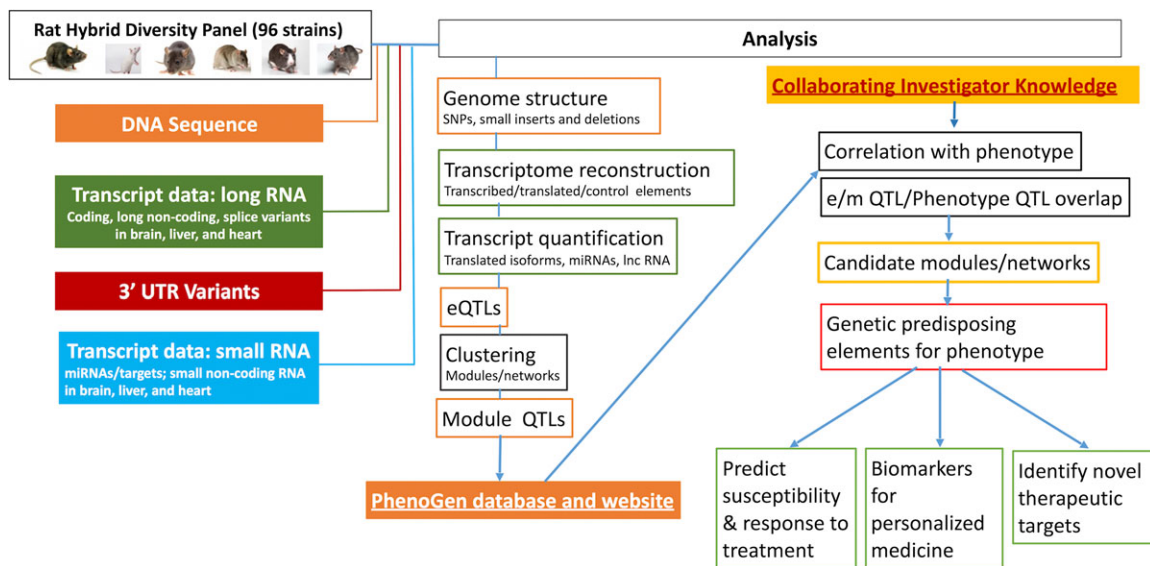
With over 190,000 uses per year from researchers in more than 160 countries, RGD remains the premier resource for rat genomic and phenotypic data providing comprehensive datasets and innovative tools. As the utility of rat as a model for human disease continues to grow with new techniques for genomic and phenotypic manipulation, RGD will continue to expand its resources and tools to aid investigators in choosing precision models for their studies.

## PhenoGen

The PhenoGen website (<https://phenogen.ucdenver.edu/PhenoGen/>) has been developed to give researchers an interactive platform to explore DNA variants, RNA expression, and complex traits in the rat and to apply a systems genetics approach to the study of complex traits. A systems genetics approach (Figure 5) requires the integration of multiple data

sets for a global analysis of molecular factors that contribute to disease phenotypes (Civelek and Lusis 2014). Thus, a critical requirement to allow for a systems genetics approach is a genetically stable population, which can be studied many times, over many generations, and will allow for accumulation of data on DNA, RNA, proteins, and metabolites to eventually facilitate a complete systems genetic analysis of a complex trait. PhenoGen’s current data accumulation efforts consist of quantitative measures of brain, liver, and heart total RNA expression and DNA sequence/variant information that is generated from such a genetically stable population, which is a subset of strains from the Hybrid Rat Diversity Panel (HRDP). The full HRDP will be a combination of recombinant inbred [HXB/BXH (Printz et al. 2003) and FXLE/LEXF (Voigt et al. 2008)] and classic inbred strains of rats that optimizes genetic mapping power and precision (Bennett et al. 2010).

With the current focus on precision medicine and the related popularity of genome-wide association studies (GWAS), one unmet need is to provide a base of knowledge to link genetic variation to variations in phenotype. Systems genetics can greatly enhance the understanding of the biologic repercussions of genetic variation. The majority of human disease variants are likely to be involved in regulation of transcription (Nicolae et al. 2010; Schaub et al. 2012). The process of transcription is the first step in which environmental variables can interact with genetically defined elements of DNA to produce a quantitative phenotype, that is, steady-state levels or turnover rates of RNA. This quantitative information allows for use of analytical approaches to discern the variation in the trait of “transcription,” to parcel out the contributions of genetics and “environment” (i.e., calculate heritability), and to use quantitative genetics to map expression QTL (i.e., genetic loci contributing to transcriptional control [eQTL]). The overlap of eQTL with the disease-associated markers (QTL) can indicate a relationship between gene expression levels and a disease phenotype (Nica and Dermitzakis 2013). One can extend this QTL analysis



**Figure 5** The PhenoGen pipeline for systems genetic analysis. The PhenoGen website (<http://phenogen.ucdenver.edu>) was designed as an interactive platform to facilitate the exploration of DNA variants, RNA expression, and complex traits in the rat using a systems genetics approach. The data sets generated from the HRDP are represented by the orange, green, red, and blue boxes on the left. These data sets are then integrated to generate the information in the column of boxes in the middle. The processed data can be explored on the PhenoGen website at the individual gene level or various aspects of the data can be combined in a phenotype-level approach when an investigator has measured a phenotype on the HRDP or a subset of the HRDP. The final outcome of such an approach would be a candidate module/network that can be used for a multitude of purposes including those indicated in the boxes outlined in green on the bottom right of the graphic.

to genetic control of functionally related transcripts by use of graph theory to construct coexpression modules (networks of biological relevance) (Allocco et al. 2004) and subject such networks to genetic correlation analysis with higher level phenotypes. The identification and use of such network information further contributes to a systems genetic approach that can combine studies in humans with studies performed on nonhuman species to understand the relationships between the DNA genetic information and biology (Gusev et al. 2016).

Using these integrative techniques, the PhenoGen website provides data and tools to apply both phenotype-driven and gene-driven analyses to understand the contribution that a genetic locus can make to a phenotype. In a phenotype-driven approach, the researcher measures a phenotype in the rats of the HRDP, or a subset of the HRDP, and uses the DNA and RNA information for the HRDP provided on PhenoGen to identify QTL and genetic networks that are associated with the phenotype. It should be stressed that the information provided on PhenoGen reflects on the basal state (i.e., no treatment) and should be used to identify factors that predispose to a phenotype. Multiple groups (Hoffman et al. 2011; Leduc et al. 2012; Saba et al. 2015; Scott-Boyer et al. 2014; Vanderlinden et al. 2013) have successfully used this approach to study complex traits. The analysis of phenotypes and the associated genetic networks can be used to (1) predict clinical outcomes or response to treatment, (2) identify biomarkers for precision medicine, and (3) identify novel therapeutic targets. For a gene-driven analysis, a researcher may already have a candidate gene for a given trait based, for example, on current therapeutics or results from human GWAS. With the PhenoGen website, the researcher can then investigate the genetic mechanisms that control the RNA expression levels of that transcript and the coexpression network to which the transcript belongs. Knowledge gained from the analysis of genetic mechanisms can be used for identifying alternative therapeutic targets or potential adverse consequences of current therapies. The information on the coexpression network can bring insight into the biological context in which the gene is functioning in a particular tissue.

The remainder of this section outlines the types of data and tools currently available on PhenoGen and concludes with a description of future plans for the website and data repository.

#### Data Sets

The PhenoGen website is home to several large RNA expression data sets from both mouse and rat. For purposes of this discussion, the focus will be on rats and in particular, on the HXB/BXH recombinant inbred (RI) panel that represents a subset of the HRDP.

**DNA Sequence and Mapping Variants.** An RI panel is originally derived from crossing two genetically diverse inbred strains. This implies that the resulting RI strains represent different combinations of the DNA of the two progenitor strains. The PhenoGen website has DNA sequence information at 25× coverage from the two progenitor strains of the HXB/BXH RI panel, the SHR/OlaIpcv and BN-Lx/Cub strains (Hermsen et al. 2015). Single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) have been identified in the RN5 and RN6 versions of the rat genome. In addition, the PhenoGen website contains a marker set for the HXB/BXH RI panel that was originally derived by the STAR Consortium (STAR Consortium et al. 2008). This marker set has been converted to the RN6 version of the rat genome, and the genotyped SNPs have been subjected

to a rigorous quality control procedure to identify a set of SNPs appropriate for mapping studies (Harrall et al. 2016; Saba et al. 2015). PhenoGen users are able to download the strain-specific reference genomes for the two progenitor strains (i.e., the BN reference genome with alleles edited for SNPs within the SHR and BN-Lx genomes) and a mapping marker set for the HXB/BXH RI panel.

**RNA Sequencing.** The PhenoGen website contains deep brain, liver, and heart RNA sequencing data from three to four biological replicates within each of the HXB/BXH RI progenitor strains. These data contain information not only on protein-coding transcripts, but also on long and short noncoding RNAs. RNA was separated by size (>200 bp and <200 bp) prior to sequencing to optimize sequencing efficiency. Because the rat transcriptome is currently underannotated in comparison to human and mouse, tissue-specific transcriptomes were reconstructed using currently available annotation and the RNA-Seq data (Roberts et al. 2011). These tissue-specific transcriptomes include information on novel splice variants (including alternative 3'-untranslated regions) of protein coding genes and novel noncoding elements. Knowledge of the 3'-untranslated region isoforms and novel, as well as annotated, microRNAs expressed in different tissues allows for the identification of microRNA target sites that may be tissue or strain specific. Both the aligned reads and the reconstructed transcriptomes are available for download and are visualized within PhenoGen's genome browser.

**RNA Microarrays.** PhenoGen also has data from brain, heart, liver, and brown adipose tissue of the HXB/BXH RI panel generated on Affymetrix Rat Exon 1.0 ST arrays. This capitalizes on the DNA-Seq and RNA-Seq data generated on the progenitor strains to create expression data sets from these arrays with improved accuracy (eliminated nonspecific probes and probes that aligned to regions of the genome that contain a SNP or small indel in either progenitor strain) and with improved interpretability (probes that uniquely target a gene from the appropriate tissue-specific reconstruction are combined into a single expression estimate; for examples, see Saba et al. 2015 and Harrall et al. 2016). These microarray data have been used for heritability, eQTL, and coexpression network analyses.

#### Tools

Current development of the PhenoGen website has focused on visualization of the DNA and RNA information gathered on the HXB/BXH RI panel and other rodent models. The main point of entry for these visualizations is the PhenoGen Genome/Transcriptome Data Browser. However, PhenoGen contains many more customizable visualizations that can be accessed through the browser.

**Genome/Transcriptome Data Browser.** This browser allows the user to visualize the genome and transcriptome, and the Selected Feature section within the browser is the starting point for the gene-driven approach discussed above. The browser contains several customizable tracks that display the DNA and RNA data mentioned above and data from several public repositories, for example, Ensembl transcripts (Yates et al. 2016), RefSeq transcripts (O'Leary et al. 2016), behavioral and physiological QTL from the RGD (Shimoyama et al. 2015), and data from the UCSC RepeatMasker (Speir et al. 2016).

One track contains DNA sequence information for the progenitor strains, allowing the user to visualize the location of SNPs and indels with respect to transcribed elements in that region and with respect to microRNA target sites. The RNA

sequencing information can be viewed as histograms based on the aligned reads (small and long RNA fractions) from each tissue, as read coverage of splice junctions, and as the reconstructed transcriptomes, which allow comparison of the alternative isoforms expressed in the different tissues. The exon array data is also available for visualization in the browser. The locations targeted by the individual exon array probe sets are displayed and can be color-coded in each tissue to represent the percent of samples from the HXB/BXH RI panel with expression values above background, heritability of the expression values in the HXB/BXH RI panel, and the Affymetrix designation of annotation confidence (core, extended, or full).

**Selected Feature Options.** From within the browser, a user can select an individual transcript to obtain additional information including Gene Details, Gene eQTL, Probe Set Level Data, miRNA Targeting Gene (multiMiR), and WGCNA (weighted gene co-expression network analysis) (Zhao et al. 2010). The Gene Details section contains additional information about the selected gene, including any available annotation, a general summary of the microarray data for this gene, and eQTL information for the strongest eQTL in each tissue. The eQTL information for the gene can be explored further in the Gene eQTL section that contains a customizable Circos plot (Krzywinski et al. 2009) that compares the eQTL profile for the transcript across the four different tissues and includes the physical location of the gene. The Probe Set Level Data section has several graphics that compare RNA expression from the exon arrays across tissues and samples. The miRNA Targeting Gene (multiMiR) section includes the predicted and validated microRNA target sites on the transcript identified by the multiMiR package in R and its associated databases (Ru et al. 2014).

**Weighted Gene Coexpression Network Analysis.** Displayed within the WGCNA section are the coexpression modules from each tissue that contain the selected transcript. Coexpression modules are identified using WGCNA and the expression estimates from the exon array data sets. In this section, coexpression modules are initially visualized as nodes (circles) and edges (connecting lines) including details about the intra-modular connectivity of individual transcripts and the strength and direction of associations between transcripts. Users can also view a Circos plot with information on the module eigen-gene (first principal component) QTL. Finally, the user can access a Sunburst plot (Stasko et al. 2000) that contains details about the Gene Ontology (Gene Ontology Consortium 2015) categories that characterize the transcripts within the module and a diagram that identifies microRNA with target sites in multiple transcripts within the module.

#### Future Directions

The PhenoGen website was established over a decade ago (Bhave et al. 2007) and continues to expand its data and functionality. In the near future, the main source of RNA expression levels will be RNA-Seq data rather than microarrays. The brain and liver transcriptomes of 30 strains of the HXB/BXH panel and 10 other inbred strains have already been sequenced. Processing of the data is currently underway with an expected release of December 2016. With the switch to RNA-Seq, coexpression networks will be expanded to include quantitative expression information on individual splice variants of protein-coding genes, annotated and unannotated long noncoding transcripts, and small RNA transcripts such as microRNAs and snoRNAs. With these data, detailed information on alternative

polyadenylation site usage, its drivers (e.g., genetics and/or sex) and its consequences (e.g., differences in microRNA target sites and binding sites for RNA processing proteins) can also be obtained. Thus far, the PhenoGen database has been generated using male rats, but a project to collect data on female rats has begun and will include examination of the influence of sex on RNA expression, genetic mechanisms controlling expression, and coexpression networks. In addition to continuously expanding the available data and enlarging the databases, PhenoGen strives to constantly improve both algorithms and visualization tools to maximize the relevancy of this database and website to the research public.

#### Gene Editing Rat Resource Center

The Gene Editing Rat Resource Center (GERRC, <http://rgd.mcg.edu/wg/gerrc>) is funded by the National Heart Lung and Blood Institute (R24HL114474) to produce rat models with gene modifications, distribute these models to the scientific community, and preserve the models. Rodent models have been used to study the function and underlying mechanisms of complex disease caused by genes implicated in human cohort studies. The development of gene-modified rat models provides investigators a platform on which to test GWAS nominated genes, validate genes contributing to QTL, and perform mechanistic studies to elucidate gene function. The GERRC allows investigators to request genetically modified rat models on appropriate background strains. Strain development is prioritized through a scientific merit review process. Applications are submitted from the scientific community and reviewed by a five-person external advisory board. Selection criteria focus on the value of the requested rat model to the nominating investigator and the potential utility to other investigators, the relevance to the NHLBI, and a rationale for needing the gene modification in rat, especially if the model exists in mouse.

The development of genetically modified rat models at the Medical College of Wisconsin began in 2009 when Drs. Geurts and Jacob and team used zinc-finger nucleases (ZFN) to target two rat genes, IgM and Rab38, on two different background strains (Geurts et al. 2009). ZFNs were used successfully to create 123 strains targeting 92 genes as part of the PhysGen Knockout program (<http://rgd.mcg.edu/wg/physgenknockouts>). The Gene Editing Rat Resource Center was built from the PhysGen Knockout infrastructure and has used TALEN, CRISPR/Cas9, and Sleeping Beauty technologies to target 91 genes on 15 different background strains to date. As new gene-modified strains are approved, the target gene and background strain are posted on the GERRC website. As the strain development is completed, the nominating investigator receives the first group of breeders. Additional rats or tissues are then made available for distribution to all investigators. Investigators receiving these models are obligated to share phenotype data generated from the rats upon publication. Phenotype data are integrated into PhenoMiner (<http://rgd.mcg.edu/phenotypes/>) at the Rat Genome Database as it becomes available to enable investigators to access the data by strain, clinical measurement, and experimental condition. After the strains are distributed and established at the nominating investigator's research facility, sperm from proven breeders are cryopreserved.

Investigators are invited to nominate novel strains twice each year. The goal of this resource is to benefit investigators and ensure that the models are easily available to the scientific community. The application review criteria include the significance and relevance of the work to human health, the



hypothesis and the need for the requested rat model, the experimental plan for the animal model, and the potential utility to more than one investigator. The External Advisory Board consists of investigators with expertise in heart, lung, and blood diseases and experience with animal models of these diseases. Applications are carefully reviewed, scored by at least three EAB members, and prioritized following a full EAB committee discussion. Strain production begins immediately following approval of the projects by the EAB. Most projects are completed within 9 to 12 months of approval. However, more complex models that involve conditional knockouts or allele-specific modifications often require more than 12 months for production.

### Rat Resource and Research Center

The major service functions of the Rat Resource and Research Center (RRRC, <http://www.rrrc.us/>) are to obtain valuable rat lines, cryopreserve them to prevent loss, rederive them to ensure that they are free of infectious pathogens, confirm their genotype, and distribute rats and cell lines to investigators needing to use these rats in their research. Cryopreservation of rat models safeguards unique genotypes from potential problems including genetic drift, genetic instability, genetic contamination, and loss due to disease or catastrophic disasters to housing facilities.

### Model Selection

The following criteria are considered in decisions regarding the recruitment and selection of rat lines for importation: scientific value including publications describing the model, nature of the mutation, potential estimates for the demand of the line, uniqueness and difficulty of replacement, existence of the line at other sources, special requirements, transgene construct information (when available), detailed observations of characteristics/phenotype, breeding scheme, difficulty of maintenance, and genotyping requirements. The RRRC does not accept rat strains/stocks that are commercially available or that are otherwise readily available from some other source. Investigators fill out an online form found on the RRRC website that provides basic information about the strain they wish to donate. These applications are evaluated regularly by an internal committee with broad scientific expertise so that a decision can be communicated to donating investigators within 3 weeks or less of submission of an application.

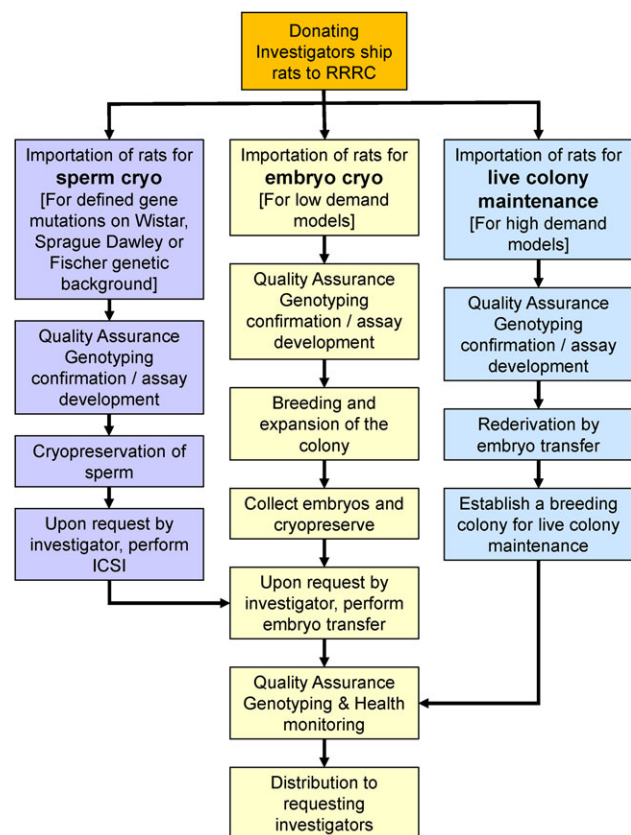
### Importation of Rat Models

Once models are accepted by the RRRC, the donating investigator completes a basic Material Transfer Agreement (MTA); transfers genotyping and, when appropriate, phenotyping protocols to the RRRC; and fills out an online breeding and husbandry form that details the reproductive characteristics of the strain and any recommended breeding schemes. This information becomes part of the strain data pages on the RRRC website. The strain is listed on the RRRC website under "Future Strains" and also appears on the homepage under "New RRRC Strains." There is an option for interested investigators who go to the strain page to indicate their interest, and donating investigators are encouraged to send other investigators to the page to register interest. Once all information is received from the donating investigator, the animals are imported into the RRRC. For lines with complex genetics (e.g., multiple mutations, mixed or unique genetic background), donating investigators transfer 5 to

10 breeding pairs along with any available additional females for the establishment of a colony to produce embryo donor females for cryopreservation or embryo transfer rederivation. For lines with simple genetics (e.g., a single mutation or transgene on a standard genetic background), donating investigators transfer 2 to 5 males for sperm cryopreservation or embryo transfer rederivation utilizing wild-type commercially available females. Upon arrival, animals are genotyped and assigned to one of the three importation groups shown in Figure 6.

### Cryopreservation of Rat Models

Rat models with a defined mutation on a standard genetic background, spermatozoa will be cryopreserved. Collection and freezing of sperm can be accomplished immediately upon arrival of the rats at the RRRC. Rat models cryopreserved in this manner will be reanimated by intracytoplasmic sperm injection (ICSI) upon request. For rat models in which the mutation is not known or the genetic background is unique or mixed, embryos are cryopreserved, since a haploid cell type (sperm) cannot be used to rederive these complex traits. For these lines, a breeding colony will be established for embryo cryopreservation. Reanimation will be performed by embryo transfer upon request of the line. A new strategy implemented for



**Figure 6** Flow chart of RRRC operations. Once a researcher has donated a rat strain to the RRRC, it will follow one of three paths depending on the characteristics of the model and the predicted demand for it. For rats with defined gene mutations on a standard rat background, the model undergoes sperm cryopreservation. For lines that are predicted to be low demand models with either unknown or complex mutations and/or backgrounds, the rats are bred to expand the colony before embryo collection and cryopreservation. For high demand models, live colonies are established for breeding, distribution, and preservation. In all cases, quality assurance, genotyping, and health monitoring are performed upon receipt of the animals and before distribution of products.

well-characterized congenic strains with standard genetic backgrounds is to freeze sperm rather than embryos. This necessitates cryorecovery by ICSI followed by a round of breeding to restore the homozygosity of the congenic region.

#### **Maintenance of Live Colonies**

For those lines that have sufficient demand to warrant a live colony for animal distribution, rats are imported into the RRRC and rederived by embryo transfer. Recipients, foster dams, and rederived pups are rigorously tested with a comprehensive health-monitoring panel to ensure successful elimination and exclusion of pathogens. Rederived rat models are housed in ventilated cage racks. Breeding schemes are developed that maintain the mutant allele(s), transgene, or other genetic alteration. Genotyping is performed on progeny to ensure the highest genetic control standards for all breeders and rats that are distributed to requesting investigators. The RRRC currently maintains 15 live strains.

#### **Genetic Monitoring of Rat Models**

The purpose of the genetic monitoring program is to ensure that all rats distributed by the RRRC are of the appropriate genotype or genetic background. Genotyping is performed upon arrival of donated animals, on animals within live colonies, and on all animals distributed to requesting investigators. Strain-specific genotyping protocols are developed whenever possible, and the use of generic genotyping assays (e.g., an assay that detects the GFP gene only) is avoided since these types of assays are ineffective in detecting genetic contamination due to the misidentification of strains carrying the same reporter genes. Developing strain-specific genotyping assays guarantees the highest genetic standards for the rats at the RRRC and provides users with the best tools for ensuring the genetic integrity of their animals at their own institutions. A detailed user-friendly genotyping protocol is written for every assay, including both validated assays based on the donating investigator's protocol and newly developed assays. These protocols are distributed to requesting investigators when rats are shipped, and they are freely accessible from the "Genetic Description" page that is available for every RRRC strain on the RRRC website.

#### **Infectious Disease Screening for Rat Models**

To ensure that rats distributed by the RRRC are free of infectious pathogens, a rigorous health-monitoring program is employed. Because of this high level of biosecurity, many institutions will grant RRRC "approved vendor" status, meaning that animals shipped from this facility can often be brought into the receiving institution's animal facility directly without the need for the animals to be quarantined. This is of substantial benefit to the investigator, since they have immediate access to the animals.

#### **Reanimation of Cryopreserved Rat Lines**

Currently ~60% of lines accepted into the RRRC are single-gene mutations or transgenics on a standard genetic background. In these cases, cryopreserved sperm are utilized with ICSI into wild-type oocytes. RRRC is one of only a few laboratories in the world able to perform rat ICSI. A 4% to 5% live pup rate (the number of pups born/number of zygotes transferred) has been achieved for both outbred (e.g., Sprague Dawley) and inbred

(e.g., Fischer 344) genetic backgrounds. To date, the success rate for recovery of requested lines using ICSI has been 100%.

The remaining ~40% of lines are those with multiple or complex gene traits, often without developed genotyping methods and/or lines with unique or mixed genetic backgrounds. In these cases, embryo transfer of cryopreserved embryos continues to be utilized as the primary approach to reanimation. As mentioned previously, for genetically well-defined congenic and consomic lines, reanimation will involve ICSI followed by a round of breeding including detailed genetic analysis to confirm homozygosity and integrity of the congenic/consomic region.

#### **Embryonic Stem (ES) Cell Lines**

RRRC is one of the only sources for rat ES cells in the world, particularly with respect to lines with proven germline competence. Two novel germline competent ES cell lines have been isolated and characterized, both of which contain an EGFP reporter (Men et al. 2012; Men and Bryda 2013), and have become the most requested ES cell lines. All ES cell lines, whether imported from other investigators or generated in-house, are subjected to rigorous quality control standards. Cell lines are cryopreserved at the earliest passage available and then some cells are grown and expanded for 3 to 4 passages to generate inventory (typically a minimum of 40 vials of  $1 \times 10^6$  cells/vial). Quality control (QC) includes: (1) karyotyping; (2) confirmation of expression of the pluripotency markers Oct4, Nanog, Sox2, and Sox3; and (3) pathogen-testing for *Mycoplasma* spp., common murine pathogens, and parvovirus (KRV, RMV, H-1, RPV). QC assessment is carried out on cryopreserved cells and various media components used for expansion and cryopreservation. This QC is performed initially when the cell lines are first received or established and again, once the cells are expanded and frozen down for inventory and prior to distribution. Detailed data sheets about the characteristics of the cell lines have been generated and are posted on the specific cell line page on the RRRC website. In addition, detailed protocols containing all the information needed to work with rat ES cells, including pictures and recipes for media/reagents, are freely available on the website. Both the data sheets and protocols are sent to requesting investigators when vials of ES cells are shipped.

#### **Distribution of Rats, Cryopreserved Germplasm, and Other Materials**

Investigators may request live animals, cryopreserved spermatozoa or embryos, ES cells, genomic DNA, or tissues by using an online form available through the RRRC website that also includes a standard MTA

#### **Fee-For-Service**

The following additional services are available.

**Cryopreservation and Cryostorage.** Cryopreservation and cryostorage of sperm and embryos is performed routinely for the strains/stocks donated to the RRRC, and these services are made available to investigators for non-RRRC strains as insurance against loss of valuable models or to generate banks of frozen material for use to refresh foundation colonies to minimize genetic drift.

**Rederivation and Cryo-Resuscitation.** Embryo transfer or ICSI can be used to resuscitate and rederive strains/stocks. Investigators will receive recovered litters with confirmed genetics and a specific pathogen-free health status.

**Colony Management and Breeding Services.** The RRRC can maintain small colonies of rats for investigators who may not have the expertise or facility space to do so. Other services include moving genetic mutations to new genetic backgrounds using a speed congenic approach, timed matings, and embryo collection.

**Genetic Testing.** Services available for both RRRC and non-RRRC rats include, but are not limited to: genotyping (standard PCR, PCR followed by restriction endonuclease digest or nucleotide sequence analysis, RT-PCR, qPCR, and probe-based allelic discrimination), sex determination assays, genotyping assay development, validation/optimization of genotyping assays, SNP analysis, speed congenic assay development, fluorescent in situ hybridization, and karyotyping.

**Embryonic Stem Cell Lines.** The RRRC can assist with the isolation and characterization, including testing for germline competency, of new ES cells from rat strains of interest.

**Pathology.** Available services include, but are not limited to: pathogen detection through a collaborating diagnostic laboratory (IDEXX BioResearch), gross necropsy examination, tissue collection, and histopathologic evaluation of tissues.

**Microinjection.** The RRRC can perform pronuclear or cytoplasmic microinjection into zygotes or blastocysts to generate transgenic or genetically modified knockout or knockin rats, including the use of CRISPR/Cas9 technology and production of knockins via homologous recombination in ES cells.

**Microbiota Analysis.** In collaboration with the MU Metagenomics Center, a full line of services is available that includes, but is not limited to: targeted 16S rRNA amplicon sequencing and analysis, consultation on the impact of differing microbiota on model phenotype and reproducibility, manipulation of microbiota through rederivation or fecal transplants, and collaborative studies assessing the impact of microbiota on model phenotypes (Ericsson et al. 2015). Such studies suggest that the microbiota plays a previously underappreciated role in model phenotypes and that changes to the microbiome may account for some of the issues with research reproducibility that have been noted.

#### RRRC Website

The RRRC website is a critical link for RRRC repository information to the scientific community. It serves as the conduit of information for investigators seeking rat models, for investigators wishing to submit rat models to the RRRC, and for orders from requesting investigators. The website contains the following: (1) general information about the RRRC including contact information; (2) strain donation information including an online application and MTA form; (3) a list of available RRRC rat strains/ES cell lines that is searchable by strain name, gene name, donating investigator, and type of model; (4) strain request information including an online order form and MTA form; (5) information about strains that will be available in the

NBRP Rat No	Strain name	Common Name	Principal Investigator
0001	ACI/NKyo	August Copenhagen Irish	Tadao Serikawa
0002	ACI-Lyst <sup>bg-Kyo</sup> /Kyo	ACI/NKyo-Lyst <sup>bg-Kyo</sup> /Kyo, ACI/NKyo-Lyst <sup>bg-Kyo</sup> , ACI/N-bg	Tadao Serikawa
0003	DMY/Kyo	DMY, demyelination rat	Tadao Serikawa
0004	DON/Kyo	DON, Donryu rat, DONRYU	Tadao Serikawa
0005	F344.ACI-Lmx1 <sup>afc</sup> /Kyo	F344.ACI-Lmx1 <sup>afc</sup> , Queue courte rats	Tadao Serikawa
0006	HTX/Kyo	HTX, H-Tx	Tadao Serikawa
0007	KHR/Kyo	KHR, kaken hairless rat	Tadao Serikawa

**Figure 7** Search interface of the NBRP-Rat Strain Database. Several categories like “general strain information,” “preservation status,” “genetic category,” or “research category” can be combined. The search fields and checkboxes in the database queries are “AND” -linked, which supports the fine selection of specific rat strains.

**Table 4** Phenotypic measurements of the NBRP-Rat phenome project

Category	Characterization	Measurements
Functional observational battery (FOB)	Home cage observations (6) <sup>a</sup>	Body position, respiration, clonic and tonic involuntary movement, vocalization, palpebral closure
	Hand-held observations (8)	Reactivity, handling, palpebral closure, lacrimation, salivation, piloerection, skin color, others
	Open field activity (10)	Rearing, clonic and tonic involuntary movement, gait, movements, arousal, occurrence of stereotype, abnormal behavior, defecations, urinations
	Stimulus response (8)	Approach response, touch, eyelid reflex, pinna, sound, tail pinch, pupillary, righting
Behavioral studies	Nervous and muscle observations (5)	Abdominal and limb tone, forelimb and hindlimb grip strength (kgf) <sup>b</sup> , landing foot splay (mm)
	Locomotor activity (4)	0–10, 10–20, 20–30 min, and total (0–30 min)
Blood pressure	Passive avoidance (2)	Training (s), retention (s)
	Blood pressure (2)	Systolic blood pressure (mmHg), heart rate (1/min),
Biochemical blood tests	Body temperature (1)	Body temperature (°C)
	Biochemistry (16)	GOT, GPT, ALP, TP, ALB, A/G, GLU, T-CHO, HDL, LDL, TG, T-BIL, BUN, CRE, IP, Ca
Hematology	Plasma electrolytes (3)	Na <sup>+</sup> , K <sup>+</sup> , Cl <sup>-</sup> (mEq/l)
	Blood counts (8)	RBC, Hb, Ht, MCV, MCH, MCHC, WBC, platelet,
	White blood cells (7)	Bas. Eos. St. Seg. Lym. Mon. Other (%)
Urine parameters	Bleeding value (2)	PT (s), APTT (s)
	Urine (2)	Volume (ml)
Anatomy	Urinary electrolytes (6)	Na <sup>+</sup> , K <sup>+</sup> , Cl <sup>-</sup> (mEq/l)
	Body weight (3)	5, 6, 10 weeks (g)
Genotype	Organ weights (16)	Brain, heart, lung, liver, kidneys, adrenals, spleen, testes (g)
	Genotyping	369 SSLP markers

<sup>a</sup>The number of measurements of each type taken is shown in parentheses.

<sup>b</sup>Units are shown in parentheses.

near future through the RRRC; (6) information about the NIH Sharing Policy and how the RRRC can be used to meet the sharing policy requirements; (7) information about current research efforts; (8) protocols including genotype assays for all strains, cryorecovery protocols, ES cell protocols, and health-monitoring reports; (9) information on consulting RRRC scientists and staff concerning all aspects of rat model development and characterization; (10) pricing information; (11) publications generated by RRRC investigators; (12) links to other resources and databases of use to the rat user community; and (13) a list of services including both standard services offered for RRRC materials as well as fee-for-service options. The “Latest News” and “News and Events” sections on the homepage highlight new strains and post information of interest to the rat user community.

An important component of the website is the strain information pages. Every rat strain has the following pages: (1) strain profile page that lists name of strain, general information about the genetic and phenotypic characteristics of the strain, types of products available for the strain (live animals, cryopreserved germplasm, etc.), and any distribution restrictions and relevant publications; (2) genetic description page that describes the type and nature of the genetic alteration including hyperlinks to relevant NCBI and RGD gene pages, the strain background, the types of genotypes available, and a link to the strain-specific genotyping protocol; and (3) Breeding and Husbandry page, which details the physical and reproductive characteristics of the strain, the recommended breeding scheme, and any notes about special husbandry needs. The RRRC works closely with the NIH-funded RGD to ensure that strain nomenclature conforms to the nomenclature rules and that all strains held at the RRRC are cross-listed on the RGD website. A separate button on the strain profile page links directly to the relevant page on RGD allowing investigators to go back and forth between both sites with ease. For ES cell

lines, there are two pages: a strain profile page and a genetic description page. Cell lines are also cross-referenced in RGD.

The website information is updated frequently to provide the most current information with respect to rat lines available, research progress, and updated protocols.

#### Collaboration

The RRRC encourages research collaborations with investigators and can provide consultation on colony management, husbandry, reproductive biology, gamete and embryo cryopreservation, microinjections, genetics, model generation and characterization, including phenotyping, histopathology, and behavioral testing.

#### Education

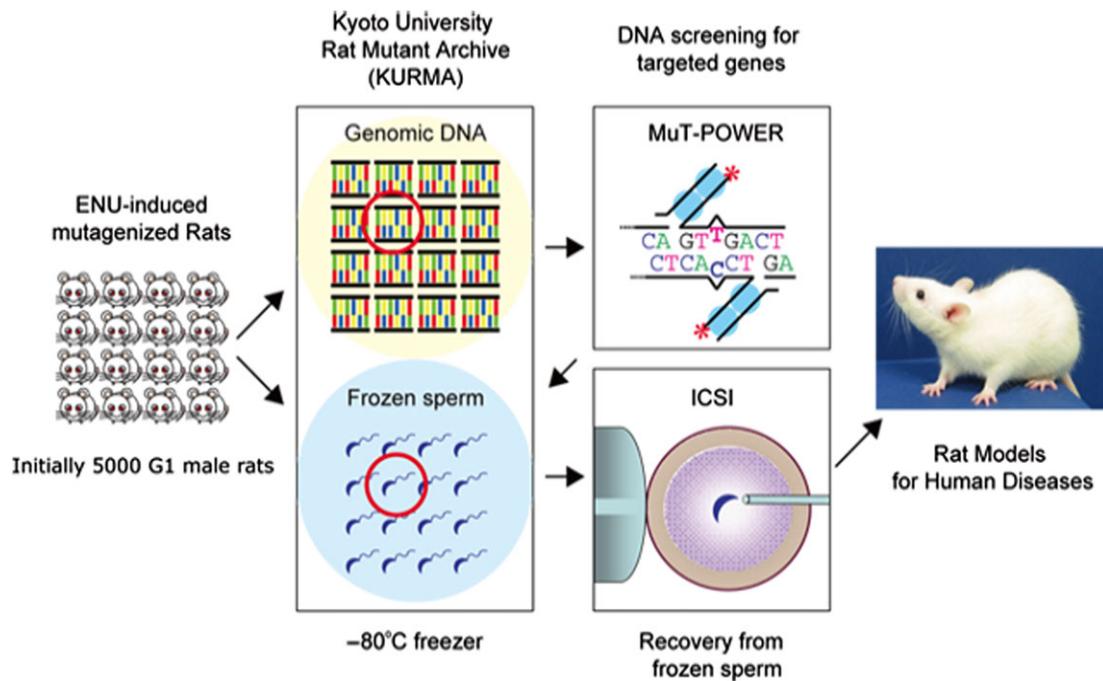
The RRRC is affiliated with the University of Missouri Comparative Medicine Program (<http://cmp.missouri.edu/>), a training program for veterinarians pursuing careers in biomedical research. The RRRC provides training for veterinary residents and veterinary students that includes hands-on experiences with genotyping, cryopreservation, embryo manipulation, surgical procedures, colony management as well as didactic sessions involving health monitoring, animal shipping, rodent reproduction, strain nomenclature, etc.

#### National BioResource Project for the Rat in Japan

The National BioResource Project for the Rat in Japan (NBRP-Rat, <http://www.anim.med.kyoto-u.ac.jp/NBR/>) was formed to collect, maintain, and preserve rat strains, characterizing them both genetically and phenotypically, and to develop and maintain a publicly available database with information about the deposited strains (Serikawa et al. 2009). NBRP-Rat maintains live stocks of strains under specific pathogen-free conditions







**Figure 10** Schematic diagram of the generation of gene-targeted rat models of human diseases. ENU is injected intraperitoneally into 9- and 10-week-old male F344/NSLc rats. They are mated 10 weeks after injection. DNA and sperm of their offspring (G1) are stored; the DNA can be screened through a newly developed method (MuT-POWER) and affected sperm are revitalized by ICSI to derive viable offspring from the G1 sperms.

also of expertise available from the RGD, PhenoGen, the Gene Editing Rat Resource Center, the RRRC, and the NBRP-Rat in Japan are invaluable resources available to assist and support them in their research efforts.

## Acknowledgments

The authors gratefully acknowledge our generous funders: the U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, grant #R01HL064541 (RGD); the U.S. Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism, grant #R24AA013162 (PhenoGen); the U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, grant #R24HL114474 (GERRC); the U.S. Department of Health and Human Services, National Institutes of Health, NIH Office of the Director, grant #P40OD011062 (RRRC); and the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (NBRP).

## References

- Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, Garcia Giron C, Hourlier T, Howe K, Kahari A, Kokocinski F, Martin FJ, Murphy DN, Nag R, Ruffier M, Schuster M, Tang YA, Vogel JH, White S, Zadissa A, Flicek P, Searle SM. 2016. The Ensembl gene annotation system. Database (Oxford) 2016.
- Allocco DJ, Kohane IS, Butte AJ. 2004. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* 5:18.
- Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, Tschannen MR, Kaisaki PJ, Otto GW, Ma MC, Keane TM, Hummel O, Saar K, Chen W, Guryev V, Gopalakrishnan K, Garrett MR, Joe B, Citterio L, Bianchi G, McBride M, Dominiczak A, Adams DJ, Serikawa T, Flicek P, Cuppen E, Hubner N, Petretto E, Gauguier D, Kwitek A, Jacob H, Aitman TJ. 2013. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* 154(3): 691–703.
- Baud A, Hermesen R, Guryev V, Stridh P, Graham D, McBride MW, Foroud T, Calderari S, Diez M, Ockinger J, Beyeen AD, Gillett A, Abdelmagid N, Guerreiro-Cacais AO, Jagodic M, Tuncel J, Norin U, Beattie E, Huynh N, Miller WH, Koller DL, Alam I, Falak S, Osborne-Pellegrin M, Martinez-Membrives E, Canete T, Blazquez G, Vicens-Costa E, Mont-Cardona C, Diaz-Moran S, Tobena A, Hummel O, Zelenika D, Saar K, Patone G, Bauerfeind A, Bihoreau MT, Heinig M, Lee YA, Rintisch C, Schulz H, Wheeler DA, Worley KC, Muzny DM, Gibbs RA, Lathrop M, Lansu N, Toonen P, Ruzius FP, de Bruijn E, Hauser H, Adams DJ, Keane T, Atanur SS, Aitman TJ, Flicek P, Malinauskas T, Jones EY, Ekman D, Lopez-Aumatell R, Dominiczak AF, Johannesson M, Holmdahl R, Olsson T, Gauguier D, Hubner N, Fernandez-Teruel A, Cuppen E, Mott R, Flint J. 2013. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat Genet* 45:767–775.
- Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang WP, He A, Kayne P, Gargalovic P, Kirchgessner T, Pan C, Castellani LW, Kostem E, Furlotte N, Drake TA, Eskin E, Lusis AJ. 2010. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res* 20(2):281–290.
- 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.

- Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A. 2008. Capture of authentic embryonic stem cells from rat blastocysts. *Cell* 135(7):1287–1298.
- Bult CJ, Eppig JT, Blake JA, Kadin JA, Richardson JE. 2016. Mouse genome database 2016. *Nucleic Acids Res* 44(D1):D840–847.
- Carlson DF, Geurts AM, Garbe JR, Park CW, Rangel-Filho A, O'Grady SM, Jacob HJ, Steer CJ, Largaespada DA, Fahrenkrug SC. 2011. Efficient mammalian germline transgenesis by cis-enhanced Sleeping Beauty transposition. *Transgenic research* 20(1):29–45.
- Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nat Rev Genet* 15(1):34–48.
- Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wieggers J, Wieggers TC, Mattingly CJ. 2016. The Comparative Toxicogenomics Database: Update 2017. *Nucleic Acids Res*.
- Ericsson AC, Akter S, Hanson MM, Busi SB, Parker TW, Schehr RJ, Hankins MA, Ahner CE, Davis JW, Franklin CL, Amos-Landgraf JM, Bryda EC. 2015. Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota. *Oncotarget* 6(32):33689–33704.
- Gene Ontology Consortium. 2015. Gene Ontology Consortium: Going forward. *Nucleic Acids Res* 43(Database issue):D1049–1056.
- Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R. 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 325(5939):433.
- Geurts AM, Moreno C. Zinc-finger nucleases: new strategies to target the rat genome. *Clin Sci (Lond)* 119(8):303–311.
- Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA, Sullivan PF, Nikkola E, Alvarez M, Civelek M, Lusis AJ, Lehtimaki T, Raitoharju E, Kahonen M, Seppala I, Raitakari OT, Kuusisto J, Laakso M, Price AL, Pajukanta P, Pasaniuc B. 2016. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48(3):245–252.
- Harrall KK, Kechris KJ, Tabakoff B, Hoffman PL, Hines LM, Tsukamoto H, Pravenec M, Printz M, Saba LM. 2016. Uncovering the liver's role in immunity through RNA co-expression networks. *Mamm Genome* 27(9–10):469–484.
- Hayman GT, Jayaraman P, Petri V, Tutaj M, Liu W, De Pons J, Dwinell MR, Shimoyama M. 2013. The updated RGD Pathway Portal utilizes increased curation efficiency and provides expanded pathway information. *Hum Genomics* 7:4.
- Hermesen R, de Ligt J, Spee W, Blokzijl F, Schafer S, Adami E, Boymans S, Flink S, van Boxtel R, van der Weide RH, Aitman T, Hubner N, Simonis M, Tabakoff B, Guryev V, Cuppen E. 2015. Genomic landscape of rat strain and substrain variation. *BMC Genomics* 16:357.
- Hoffman PL, Bennett B, Saba LM, Bhave SV, Carosone-Link PJ, Hornbaker CK, Kechris KJ, Williams RW, Tabakoff B. 2011. Using the Phenogen website for 'in silico' analysis of morphine-induced analgesia: Identifying candidate genes. *Addict Biol* 16(3):393–404.
- Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C. 2015. The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res* 43(Database issue):D1057–1063.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44(D1):D457–462.
- Kozomara A, Griffiths-Jones S. 2014. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42(Database issue):D68–73.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645.
- Laulederkind SJ, Liu W, Smith JR, Hayman GT, Wang SJ, Nigam R, Petri V, Lowry TF, de Pons J, Dwinell MR, Shimoyama M. 2013. PhenoMiner: Quantitative phenotype curation at the rat genome database. *Database (Oxford)* 2013:bat015.
- Leduc MS, Blair RH, Verdugo RA, Tsaih SW, Walsh K, Churchill GA, Paigen B. 2012. Using bioinformatics and systems genetics to dissect HDL-cholesterol genetics in an MRL/MpJ x SM/J intercross. *J Lipid Res* 53(6):1163–1175.
- Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, Zhao Y, Liu M. 2013. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol* 31(8):681–683.
- Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, Nakajima R, Kato M, Hirabayashi M, Kuramoto T, Serikawa T. 2008. An ENU-induced mutant archive for gene targeting in rats. *Nat Genet* 40(5):514–515.
- Men H, Bauer BA, Bryda EC. 2012. Germline transmission of a novel rat embryonic stem cell line derived from transgenic rats. *Stem Cells Dev* 21(14):2606–2612.
- Men H, Bryda EC. 2013. Derivation of a germline competent transgenic Fischer 344 embryonic stem cell line. *PLoS One* 8(2):e56518.
- Menoret S, Fontaniere S, Jantz D, Tesson L, Thinard R, Remy S, Usal C, Ouisse LH, Fraichard A, Anegon I. 2013. Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J* 27(2):703–711.
- NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 44(D1):D7–d19.
- Nica AC, Dermitzakis ET. 2013. Expression quantitative trait loci: Present and future. *Philos Trans R Soc Lond B Biol Sci* 368(1620):20120362.
- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet* 6(4):e1000888.
- O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvermin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44(D1):D733–745.
- Petri V, Hayman GT, Tutaj M, Smith JR, Laulederkind S, Wang SJ, Nigam R, De Pons J, Shimoyama M, Dwinell MR. 2016. Disease, models, variants and altered pathways-journeying RGD through the magnifying glass. *Comput Struct Biotechnol J* 14:35–48.
- Petri V, Hayman GT, Tutaj M, Smith JR, Laulederkind SJ, Wang SJ, Nigam R, De Pons J, Shimoyama M, Dwinell MR, Worthey EA, Jacob HJ. 2014a. Disease pathways at the Rat Genome Database Pathway Portal: Genes in context-a network



- approach to understanding the molecular mechanisms of disease. *Hum Genomics* 8:17.
- Petri V, Jayaraman P, Tutaj M, Hayman GT, Smith JR, De Pons J, Laulederkind SJ, Lowry TF, Nigam R, Wang SJ, Shimoyama M, Dwinell MR, Munzenmaier DH, Worthey EA, Jacob HJ. 2014b. The pathway ontology—updates and applications. *J Biomed Semantics* 5(1):7.
- Petri V, Shimoyama M, Hayman GT, Smith JR, Tutaj M, de Pons J, Dwinell MR, Munzenmaier DH, Twigger SN, Jacob HJ. 2011. The Rat Genome Database pathway portal. *Database (Oxford)* 2011:bar010.
- Printz MP, Jirout M, Jaworski R, Alemayehu A, Kren V. 2003. Genetic Models in Applied Physiology. HXB/BXH rat recombinant inbred strain platform: A newly enhanced tool for cardiovascular, behavioral, and developmental genetics and genomics. *J Appl Physiol* (1985) 94(6):2510–2522.
- Pundir S, Magrane M, Martin MJ, O'Donovan C. 2015. Searching and navigating UniProt databases. *Curr Protoc Bioinformatics* 50:1.27.21–10.
- Roberts A, Pimentel H, Trapnell C, Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27(17):2325–2329.
- Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, Mahaffey S, Rossi S, Calin GA, Bemis L, Theodorescu D. 2014. The multiMiR R package and database: Integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res* 42(17):e133.
- Saba LM, Flink SC, Vanderlinden LA, Israel Y, Tampier L, Colombo G, Kiianmaa K, Bell RL, Printz MP, Flodman P, Koob G, Richardson HN, Lombardo J, Hoffman PL, Tabakoff B. 2015. The sequenced rat brain transcriptome—its use in identifying networks predisposing alcohol consumption. *Febs J* 282(18):3556–3578.
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. 2012. Linking disease associations with regulatory information in the human genome. *Genome Res* 22(9):1748–1759.
- Scott-Boyer MP, Praktiknjo SD, Llamas B, Picard S, Deschepper CF. 2014. Dual linkage of a locus to left ventricular mass and a cardiac gene co-expression network driven by a chromosome domain. *Front Cardiovasc Med* 1:11.
- Serikawa T, Mashimo T, Takizawa A, Okajima R, Maedomari N, Kumafuji K, Tagami F, Neoda Y, Otsuki M, Nakanishi S, Yamasaki K, Voigt B, Kuramoto T. 2009. National BioResource Project-Rat and related activities. *Exp Anim* 58(4):333–341.
- Shimoyama M, De Pons J, Hayman GT, Laulederkind SJ, Liu W, Nigam R, Petri V, Smith JR, Tutaj M, Wang SJ, Worthey E, Dwinell M, Jacob H. 2015. The Rat Genome Database 2015: Genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res* 43(Database issue):D743–750.
- Shimoyama M, Nigam R, McIntosh LS, Nagarajan R, Rice T, Rao DC, Dwinell MR. 2012. Three ontologies to define phenotype measurement data. *Front Genet* 3:87.
- Smith JR, Park CA, Nigam R, Laulederkind SJ, Hayman GT, Wang SJ, Lowry TF, Petri V, Pons JD, Tutaj M, Liu W, Worthey EA, Shimoyama M, Dwinell MR. 2013. The clinical measurement, measurement method and experimental condition ontologies: Expansion, improvements and new applications. *J Biomed Semantics* 4(1):26.
- Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, Heitner S, Harte RA, Haussler M, Guruvadoo L, Fujita PA, Eisenhart C, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. 2016. The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res* 44(D1):D717–725.
- STAR Consortium, Saar K, Beck A, Bihoreau MT, Birney E, Brocklebank D, Chen Y, Cuppen E, Demonchy S, Dopazo J, Flicek P, Foglio M, Fujiyama A, Gut IG, Gauguier D, Guigo R, Guryev V, Heinig M, Hummel O, Jahn N, Klages S, Kren V, Kube M, Kuhl H, Kuramoto T, Kuroki Y, Lechner D, Lee YA, Lopez-Bigas N, Lathrop GM, Mashimo T, Medina I, Mott R, Patone G, Perrier-Cornet JA, Platzer M, Pravenec M, Reinhardt R, Sakaki Y, Schilhabel M, Schulz H, Serikawa T, Shikhagaie M, Tatsumoto S, Taudien S, Toyoda A, Voigt B, Zelenika D, Zimdahl H, Hubner N. 2008. SNP and haplotype mapping for genetic analysis in the rat. *Nat Genet* 40(5):560–566.
- Stasko J, Catrambone R, Guzdial M, Mcdonald K. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *Int J Hum Comp Stud* 53(5):663–694.
- Tatusova T. 2016. Update on genomic databases and resources at the National Center for Biotechnology Information. *Methods Mol Biol* 1415:3–30.
- Tong C, Huang G, Ashton C, Wu H, Yan H, Ying QL. 2012. Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. *J Genet Genomics* 39(6):275–280.
- Vanderlinden LA, Saba LM, Kechris K, Miles MF, Hoffman PL, Tabakoff B. 2013. Whole brain and brain regional coexpression network interactions associated with predisposition to alcohol consumption. *PLoS One* 8(7):e68878.
- Voigt B, Kuramoto T, Mashimo T, Tsurumi T, Sasaki Y, Hokao R, Serikawa T. 2008. Evaluation of LEXF/FXLE rat recombinant inbred strains for genetic dissection of complex traits. *Physiol Genomics* 32(3):335–342.
- Wang SJ, Laulederkind SJ, Hayman GT, Petri V, Liu W, Smith JR, Nigam R, Dwinell MR, Shimoyama M. 2015. PhenoMiner: A quantitative phenotype database for the laboratory rat, *Rattus norvegicus*. Application in hypertension and renal disease. *Database (Oxford)* 2015.
- Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P. 2016. Ensembl 2016. *Nucleic Acids Res* 44(D1):D710–716.
- Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. 2010. Weighted gene coexpression network analysis: State of the art. *J Biopharm Stat* 20(2):281–300.