Published in final edited form as:

Nat Genet. 2008 May ; 40(5): 523–527.

What everybody should know about the Rat Genome and its online resources

Simon N. Twigger, Kim D. Pruitt, Xose M. Fernandez-Suarez, Donna Karolchik, Kim Worley, Donna Maglott, Garth Brown, George Weinstock, Richard Gibbs, Jim Kent, Ewan Birney, and Howard J. Jacob

Abstract

It has been four years since the rat genome's original publication. Five groups are working together to assemble, annotate and release the next version of the genome for this key model system. As the prevailing model for physiology, complex disease and pharmacological studies, there is an acute need for the rat's genomic resources to keep pace with the rat's prominence in the laboratory. In this commentary we describe the current status of the rat genome sequence and the plans for its impending 'upgrade'. We then cover the key online resources providing access to the rat genome, including the new SNP views at Ensembl, NCBI's RefSeq and Genes databases, UCSC's Genome Browser and RGD's disease portals for cardiovascular disease and obesity.

The Rat Genome sequencing and assembly

The draft quality Norway rat (*Rattus norvegicus*, strain BN/NHsdMcwi) genome was sequenced in a public project supported by the NHGRI and the NHLBI and led by the Baylor College of Medicine Human Genome Sequencing Group. This assembly (RGSC v3.4) used a variety of sequence and map resources in a bacterial artificial chromosome (BAC) plus whole genome shotgun (WGS) strategy. Following publication in 2004¹, one subsequent update was released where portions of the draft were replaced with finished sequence from BACs². The sequence quality of the current assembly is good though some difficult-to-assemble regions remain to be addressed.

Additional sequence data has been produced since the initial assembly, further enhancing the sequence-based resources for the rat community as well as expanding the input data available for use in calculating the genome annotation. For example, ongoing sequencing projects by both individual labs and larger efforts have resulted in over 4,500 new mRNA submissions to GenBank in the past three years (See Box 1 for some general Rat genome and sequence statistics compared to Mouse and Human). Other data includes over 6000 cDNA sequences produced by the Mammalian Gene Collection and 3M WGS reads for SNP discovery released by BCM-HGSC. The scope of this extra sequence data extends all the way to 1x coverage of the Sprague Dawley (SD) genome completed by Applied Biosystems and a complete WGS assembly released by Celera. More details of these projects can be found in Worley et al.².

Genome Upgrade 2008

The additional sequence data now available provides an opportunity to upgrade the rat assembly, and work is underway to incorporate this new data in a genome upgrade planned for early 2008. The upgrade will combine the unique sequence data from the WGS-only assembly with that from the combined WGS-plus-BAC assembly resulting in a more complete representation of the genome. Also in development are plans to improve the coverage of the Y chromosome. The Y chromosome was originally sequenced only to about two-fold coverage as a consequence of the unusually large size of the BN Y chromosome. This will be addressed

by including additional Y sequence from a rat strain with a more moderately sized Y chromosome, combining WGS sequence and low coverage sequence from male BAC clones.

SNP identification and haplotype map resources

Another crucial genomic resource for any model system are single nucleotide polymorphisms. A great deal of recent progress has occurred in this area that will soon translate to a rapid increase in the number of SNPs available for the rat. Until recently there were less than 40,000 rat RefSNPs in dbSNP; however, a number of groups are working to greatly increase this number. The European STAR consortium together with Japanese colleagues have identified 2.9M SNPs from genomic DNA from six strains (S SD, SS/Jr, GK/Ox, WKY/Mdc, F344 and SHRSP/Mdc). 20,000 of these SNPs have been genotyped in over 300 inbred and hybrid strains³. Eight strains (PVG, F344, SS, LEW, BB, FHH, DA, and SHR) have been sequenced by the U.S. SNP discovery effort. The EURATools project intends to take these SNPs and genotype them across at least 150 inbred strains, enabling the development of a haplotype map. The previous rat SNP map was based on cDNA from four strains (SHRSP, BN, WKY, and SD) and had 12,395 interstrain polymorphic sites⁴. The databases and tools available for working with these SNP resources are described below.

Despite still being not being to ‘finished’ quality across the whole genome, the more complete and accurate genome sequence produced by the planned upgrade combined with additional SNP markers will clearly be of great use to many researchers. Access to these resources is predominantly via the various genome databases that analyze, annotate and curate the assembled genome sequence. The following sections outline some of resources available at these sites.

Rat annotation pipelines at Ensembl

Ensembl provides a comprehensive system integrating genome annotation visualization with a user-friendly web browser combined with open access to the underlying databases⁵. As with the other genomes found in Ensembl, the rat genome is annotated using a well established pipeline⁶ that begins with the latest assembly (currently RGSCv3.4) and ultimately delivers the Ensembl gene models for the rat. This pipeline has been streamlined and optimized in recent years, taking advantage of new data types such as more accurate cDNA^{6,7} sequences. In the current release (v47) this results in the annotation of over 35,000 transcripts and almost 23,000 protein-coding genes. Of growing interest in recent years, 2704 RNA genes (including miRNA, snoRNA snRNA and rRNA) are also annotated through this system⁵. A further unique contribution of this pipeline are EST genes, predicted splice variants based solely on EST evidence⁸. Ensembl also presents re-sequencing data from the Sprague-Dawley (SD) strain generated by Celera, enabling this to be compared to the BN reference genome.

SNPs and genome variation

Building on dbSNP Ensembl includes approximately 2.9M SNPs generated by the STAR project. New visualization tools such as **TranscriptSNPView** make it possible to compare this variation data between the different rat strains for which there is genotype data available⁹. Sequences from these strains can be visualized with **SequenceAlignView** highlighting SNPs that exist between a particular strain and the reference assembly (see Supplementary Fig. 1). Currently data from the SD, SS/Jr, GK/Ox, WKY/Mdc, F344 and SHRSP/Mdc strains are available, along with the reference strain BN/NHsdMcwi.

Comparative Genomics

Ensembl aligns whole genomes identifying constrained elements (putative functional elements) which are shown in **ContigView**. These multiple sequence alignments are produced

using PECAN, a progressive alignment algorithm. Based on these data **AlignSpliceView** can display rat alongside the syntenic mouse or human regions showing homologous genes in their genomic context, as well as conserved regions between the species. Homology relationships at the gene level are calculated using a phylogenetic approach and can be visualized in the **GeneTreeView** (see Supplementary Fig. 2). The one-to-one orthologies and other orthology categories (such as one-to-many and many-to-many) are available in the **GeneView** pages (see Supplementary Fig. 3).

Ensembl as a data integration platform

While Ensembl contains a great deal of information, there is much valuable data available outside of the Ensembl system. To address this issue, extensive use is made of the Distributed Annotation System (DAS) to integrate external resources into the Ensembl browser. This enables any DAS-compatible sources to be displayed alongside the primary Ensembl annotation. Microarray data can be integrated in this fashion, and GeneView allows users to visualize and access ArrayExpress data associated with a particular gene. Phenotype and expression data are of paramount importance in EURATools. Ensembl's contribution to this consortium involves integrating additional cDNA data obtained from the re-sequencing effort and presenting expression QTL (eQTL) data, using such tools as eQTL Explorer¹⁰ which relies on the DAS protocol.

Data mining with BioMart

BioMart¹¹ provides a data-mining tool to interact with the Ensembl database, enabling users to go beyond the genome browsers to retrieve information from Ensembl. For rat this contains the current Ensembl genome datasets, a SNP dataset and EURAMart which is a compendium of gene expression data populated with the Gene Expression Atlas from GNF¹³. BioMart has been deployed by several other databases (e.g. RGD¹²) and can be used to integrate data from these different installations. EURAMart can act as a bridge between Ensembl and RGD data allowing the integration of expression data (from EURAMart) with phenotype and disease information (from RGD).

Rat resources at the National Center for Biotechnology Information (NCBI)

NCBI maintains several resources that support the rat research community including the integrated suite of literature, sequence and BLAST databases, tools to query, retrieve, and display biological information contained in those databases, reference sequence¹⁴ and Gene¹⁵ records, variation data, two whole genome shotgun assemblies annotated by NCBI, and radiation hybrid and genetic maps. The Rat Genome Resources web page provides an up-to-date portal to access these and other rat-specific data (see Supplementary Table 1). The following highlights a subset of resources of interest to rat researchers. More information on NCBI resources is included in the Supplementary data and the online NCBI Handbook, and NCBI Help book.

Gene

Gene is the central resource for rat gene-specific information at NCBI and includes protein- and non-protein-coding genes, pseudogenes, and mapped phenotypes. The database reports assigned gene ontology (GO) terms, cytogenetic locations, names and symbols, pathways, protein interactions, publications including the GeneRIF annotated bibliography, sequences (GenBank and RefSeq), and links to numerous NCBI and other resources. Data is maintained through a combination of computation, collaboration, and ongoing curation by the Gene and RefSeq staff. Collaboration and curation enhance the content of Gene by: a) integrating information from, and establishing links to, databases such as the Rat Genome Database (RGD), RATMAP, Ensembl, and UniProt; b) resolving identified data conflicts and

ambiguities; and c) adding content such as sequences, names, phenotypes, and publications. For example regular updates synchronize rat gene nomenclature in the Gene database with that provided by RGD, add information based on new sequence submissions to GenBank, update GO terms obtained via FTP from the Gene Ontology Consortium, and, in collaboration with UniProtKB, update cross-links between Reference Sequence (RefSeq) proteins and the corresponding Swiss-Prot or TrEMBL proteins.

RefSeq

RefSeq data for rat includes the genomic reference (RGSCv3.4) and Celera assemblies as well as gene-specific products. Accessions are assigned to chromosomes (see Supplementary Table 2), scaffolds, and contigs. Gene-specific RefSeqs for RNAs and proteins include curated records based on submissions to GenBank, and predicted records that are generated as a product of computing annotation for the genome assemblies.

Curation of RefSeq transcripts and proteins for rat is a continuous process and serves to: a) ensure accurate, full-length sequence for the complete set of transcripts and proteins including loci that use selenocysteine or non-AUG codons; and b) provide additional RefSeq feature annotation such as mature peptides. In addition, the transcript-based curated RefSeq collection represents a high quality complement to genome annotation, because it can be used to identify genes which are not well-represented in one or both genomic assemblies. For example, genes missing from the RGSCv3.4 reference assembly include the smooth muscle alpha-actin (Acta2, GeneID:81633), thymidine kinase 1 (Tk1; GeneID:24834), and gamma-glutamyltransferase 1 (Ggt1, GeneID:116568).

Map Viewer (see Supplementary Fig. 4)

NCBI provides a unique service by annotating both available genome assemblies and displaying order of objects in multiple coordinate systems (sequence, centiMorgans). Genome annotation is computed based on alignments of the curated RefSeq collection described above, rat transcript data, and human, mouse, and rat protein data. The results are distributed in the genomic RefSeq collection, in the RefSeq and Map Viewer FTP sites, and are available for browsing and querying by accession, text, or sequence similarity (via BLAST) in the Map Viewer. Sequence-based data presented in the Map Viewer includes the annotated genome at the gene and transcript level plus an array of additional sequence details including repeats, STS markers, CpG islands, alignments of rat genomic records and of human, mouse, and rat transcript sequences, mapped phenotypes (QTLs) based on placement of flanking and peak markers, and variation data from dbSNP. Alternate displays provide tabular reports and download support (Data as Table View), present alignments supporting the annotation (Evidence Viewer), or support using transcript alignments to generate alternative transcript models for further evaluation (Model Maker). Map Viewer also supports comparative displays of human, mouse, and rat annotations as well as review of order and orientation of assemblies based on placement of markers common to the sequence, genetic, and radiation hybrid maps.

BLAST

A rat-specific BLAST page facilitates access to several custom BLAST databases. Options include the genome assemblies, RefSeq and GenBank RNAs and proteins, and trace reads from the Trace Archive. Query results for transcript and protein databases return links to the Gene, UniGene and/or GEO databases when an accession is known to that database. Query results for the genome assembly databases include links to view the results Map Viewer in the context of the genome annotation.

dbSNP

dbSNP processes submissions of variation of multiple classes (*e.g.* insertion/deletions, small tandem repeats, substitutions, *etc.*) and assigns them unique stable identifiers (ss). Submissions are clustered periodically by alignment to the genome, and submissions of the same variant are assigned an rs number. The placement of these variants on the genome, and calculation of the effect of a variant on an encoded protein, is reported in Map Viewer and the dbSNP GeneView display.

Rat Resources on the UCSC Genome Browser Website

The University of California, Santa Cruz (UCSC) Genome Bioinformatics Group provides a large collection of annotation data for the rat genome, along with a variety of web-based resources for displaying, querying, analyzing and downloading the assembly sequence and annotations. The complete toolset, downloadable data, documentation, and related information are available through links on the Genome Browser website at <http://genome.ucsc.edu/> (see Supplementary Table 3).

Rat sequence and annotation data

The three most recent rat assemblies from the Baylor Human Genome Sequencing Center are featured on the main Genome Browser website, and an older release is archived at <http://genome-rn1.cse.ucsc.edu/cgi-bin/hgGateway>. New assemblies are added as they become available. Supplementary Table 4 provides a list of the assemblies currently supported on the UCSC website.

A broad set of annotations generated by UCSC and its collaborators are available for each rat assembly (see Supplementary Table 5). The annotation data, which is stored in a MySQL database shared by all the tools on the website, is roughly grouped into seven different categories based on shared characteristics: mapping and sequencing data, phenotype and disease association data, genes and gene predictions, mRNA and EST data, expression and regulation data, comparative genomics data, and variations and repeats.

Of particular note for the rat annotation, UCSC provides its own set of predicted protein-coding genes, UCSC Known Genes, based on protein data from UniProt and mRNA data from RefSeq and GenBank. This gene set will be further refined in the next rat assembly (rn5) to pull in additional lines of evidence and include putative non-coding transcripts. Also of note is UCSC's extensive collection of rat comparative genomics data, which includes a measure of evolutionary conservation across several multiply-aligned vertebrates (Conservation), predictions of conserved elements within this group (Most Conserved), and pairwise alignments of several species to the rat genome showing both the alignment "chains" and the best chain for every part of the rat genome ("nets")¹⁶.

Browsing, analysis, and data-mining tools

The *Genome Browser*¹⁷ is a fast interactive web-based tool that displays a chromosome-oriented view of the annotation data aligned as horizontal "tracks" to the rat genome sequence (see Supplementary Fig. 5). The rat genome can be queried using a wide range of search parameters, including chromosomal coordinate ranges, mRNA or EST accessions, gene names, clone accession numbers, and keywords from mRNA GenBank descriptions. Navigation and configuration controls allow the user to adjust and customize the browsing window to focus on information of interest. Detailed description pages for the displayed data elements link to information from a wide range of external resources.

The **Table Browser** provides a graphical interface for downloading and manipulating the data underlying the rat Genome Browser annotations (see Supplementary Fig. 6). The user can view data tables, filter the output based on one or more criteria, intersect or correlate data from multiple tables, and restrict the output to specific coordinate ranges or lists of data elements. For more complex queries, Table Browser output may be exported to Penn State's Galaxy tool for additional processing, or users can directly access the rat database through UCSC's public MySQL server.

UCSC's **Custom Tracks** functionality offers a convenient way for users to display and compare their own data with the built-in rat annotation. User-generated custom tracks may be viewed in the Genome Browser and manipulated using the full functionality of the Table Browser. They also provide an ideal medium for presenting a view of data submitted for publication or for sharing data with collaborators.

The **Blat** tool offers a fast method for quickly mapping rat sequence to the genome (see Supplementary Fig. 7). With DNA searches, Blat quickly finds matches of 95% and greater similarity of length 25 bases or more, finding perfect sequence matches of 33 bases and sometimes as few as 20 bases. On proteins, BLAT finds sequences of 80% and greater similarity of length 20 amino acids or more.

The **Gene Sorter** provides a graphical interface for exploring expression, homology and other relationships among rat genes, as well as orthology with several model organisms such as human, mouse, zebrafish, fruitfly, worm, and yeast (see Supplementary Fig. 8).

The **Genome Graphs** tool may be used to upload and view genome-wide data sets such as QTL mapping and linkage studies.

The **Proteome Browser** allows the user to examine many characteristics (such as exon structure, domains, and structural features) of proteins associated with the gene (see Supplementary Fig. 9).

Getting more information

Most of the programs and utilities on the UCSC website are documented through online user's guides. The website also provides a FAQ, links to several online tutorials, and subscription information and archives for three publicly accessible technical support mailing lists. See Supplementary Table 6 for a list of online resources.

Rat annotation resources at RGD

The Rat Genome Database (RGD, <http://rgd.mcw.edu>) is the model organism database for the laboratory rat¹². Its goal is to provide data and tools that build upon the rat genome and related genomic resources to maximize the utility of the rat as a model organism.

The core of RGD is information curated from the published literature combined with data imported from other authoritative sources such as NCBI¹⁵, Ensembl⁵, UCSC¹⁸, MGI¹⁹ and Uniprot²⁰. RGD manually curates data for a variety of 'biological objects' and these are made available via the RGD web site and through downloadable files on the RGD FTP site. The website includes pages describing each rat gene and its function; an extensive catalog of rat strains used as experimental models with their phenotypes and known diseases; all rat quantitative trait loci that have been identified mapping phenotypes to specific regions of the rat genome. A complete list of all RGD data objects and links to example reports are available in Supplementary Table 7.

Bio-ontologies - attaching function to the genome

The rat is very much a 'functional model', it is widely used to study genes involved in specific systems-level phenotypes or as a model of a particular disease. As a result many researchers are looking for information to tie genes back to phenotypes or diseases of interest. In the current era, bio-ontologies²¹ provide the framework for describing this information in RGD and many other databases. As part of its ongoing curation RGD provides the Gene Ontology²² annotation for the rat, describing the molecular function of a gene, the cellular component(s) in which it has been found and biological processes it is involved in. These are augmented with pathway, mammalian phenotype²³ and disease annotations. QTL and strain records are also annotated with phenotype and disease ontology terms.

These annotations capture a wealth of functional knowledge within the database in a consistent fashion. They also tie the systems-level biology of disease, phenotype and pathway to the rat genome via the annotated QTL and genes. Comparative genomics and orthology can then enable this information to be applied to the genome of other species such as human and mouse.

RGD Disease portals - quick access to disease genes/QTL for Rat, Mouse and Human

A major area of interest for the rat is as an animal model of human disease. To meet this need RGD has been creating specialized web sites (or 'portals') which provide an overview of disease-related genes, QTL and strains from rat, mouse and human in a single web page¹¹. The information on these sites is developed through targeted literature curation and ontology annotation and is focused on rat but also includes mouse data provided by MGI and human QTL data curated by RGD. The focus is on diseases with significant research in the rat and which are of high clinical relevance. The portals currently include Cardiovascular (Fig. 1), Neurological and Obesity/Metabolic Syndrome. The Cancer/Neoplasms portal is in development with released schedule for mid-2008. Links for these three portals are provided in Supplementary Table 8 and screenshots of the Cardiovascular portal are shown in Supplementary Figures 10a and 10b.

Bioinformatic Tools for Rat Genomics

In addition to the curation and integration of rat genomic data, RGD provides a variety of tools to navigate, analyze and visualize this data. **GViewer** provides a broad, ontology-based search engine that can be used to find rat, mouse and human genes and other objects related to such things as pathways, diseases and phenotypes within the database. It provides a unique graphical view of the search results showing the distribution of these objects across the rat genome. RGD maintains a **BioMart** data warehouse in collaboration with the MCW Proteomics Center that can be used to create ad hoc datasets. BioMart can be linked to other BioMarts (such as EURAMart, described above) to enable complex queries between databases. **SNPlotyper** is a new tool supporting analysis and visualization of the emerging rat SNP datasets, particularly the most recent release from the STAR consortium. Links to each of these tools and others are provided in Supplementary Table 9.

Rat Strains - specialized model systems for basic and translational research

There are large numbers of different rat strains in use by the research community. Each has unique characteristics in terms of genotype and phenotype and as such embodies a unique model system. RGD curates strain data from the literature and currently has over 1400 records including inbred, outbred, congenic and consomic strains as well as heterogeneous stock and recombinant inbred lines. Many have been annotated according to their observed phenotypes or applicability to disease studies, and by using this information it is possible to identify strains that may provide a model system for subsequent research projects.

Discussion

As evidenced by the ongoing enhancements to the rat genome sequence and the comprehensive bioinformatics resources described above, the rat and its genome has a broad base of support. This is being strengthened by closer coordination between the genome assembly and annotation groups beginning with the impending genome upgrade. Plans are in development for the creation of a consensus gene set for the rat. This will be based on comparisons of results from the Ensembl, UCSC and RefSeq pipelines in collaboration with RGD to manage nomenclature and functional annotations.

There are many exciting developments coming to fruition in rat genomics through the efforts of groups worldwide. An overview of these developments and a vision for the future of the rat are covered in more detail in the accompanying perspective²⁵. However, it is indisputable that modern biology rests heavily on the support of bioinformatics and public databases and in this, and many other areas, the portents look very promising for the year of the rat.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the Intramural Research Program of the NIH, National Library of Medicine. The Rat Genome Database is supported in part by NIH grants HL-64541 and HG-002273. EURATools and the STAR consortium are supported via the Sixth Framework Programme of the European Union, action line LSH-2003- 1.1.0-1. The UCSC Genome Browser project is funded by grants from the National Human Genome Research Institute (NHGRI), the Howard Hughes Medical Institute (HHMI) and the National Cancer Institute (NCI). The Phase 2 genome project and SNP discovery at Baylor College of Medicine Human Genome Sequencing Center is funded by NHGRI HG-003273. We would like to thank Tim Aitman, Norbert Hübner and Anne Kwitek for helpful comments during the preparation of this manuscript.

Bibliography

1. Gibbs RA, et al. *Nature* 2004;428:493–521. [PubMed: 15057822]
2. Worley KC, Weinstock, George M, Gibbs, Richard A. *Physiol. Genomics* 2007;32:273–282. [PubMed: 18029439]
3. STAR Consortium. SNP and Haplotype mapping for genetic analysis in the Rat *Nat Genet.* 2008This Issue
4. Zimdahl H, et al. *Science* 2004;303:807. [PubMed: 14764869]
5. Hubbard TJ, et al. *Nucleic Acids Res* 2007;35:D610–7. [PubMed: 17148474]
6. Curwen V, et al. *Genome Res* 2004;14:942–50. [PubMed: 15123590]
7. Fernandez-Suarez, XM.; Searle, S. In *Silico Genomics and Proteomics: Functional Annotation of Genomes and Proteins*. Nova Science Publishers; 2006. p. 109-113.E., B.N., M. & R., A.
8. Eyraas E, Caccamo M, Curwen V, Clamp M. *Genome Res* 2004;14:976–87. [PubMed: 15123595]
9. Cunningham F, et al. *Nat Genet* 2006;38:853. [PubMed: 16874317]
10. Mueller M, et al. *Bioinformatics* 2006;22:509–11. [PubMed: 16357031]
11. Kasprzyk A, et al. *Genome Res* 2004;14:160–9. [PubMed: 14707178]
12. Twigger SN, Shimoyama M, Bromberg S, Kwitek AE, Jacob HJ. *Nucleic Acids Res* 2007;35:D658–62. [PubMed: 17151068]
13. Walker JR, et al. *Genome Res* 2004;14:742–9. [PubMed: 15060018]
14. Pruitt KD, Tatusova T, Maglott DR. *Nucleic Acids Res* 2007;35:D61–5. [PubMed: 17130148]
15. Maglott D, Ostell J, Pruitt KD, Tatusova T. *Nucleic Acids Res* 2007;35:D26–31. [PubMed: 17148475]
16. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. *Proc Natl Acad Sci U S A* 2003;100:11484–9. [PubMed: 14500911]
17. Kent WJ, et al. *Genome Res* 2002;12:996–1006. [PubMed: 12045153]

18. Karolchik D, et al. *Nucleic Acids Res.* 2007
19. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA. *Nucleic Acids Res.* 2007
20. Apweiler R. *Nucleic Acids Res* 2007;35:D193–7. [PubMed: 17142230]
21. Bodenreider O, Stevens R. *Brief Bioinform.* 2006
22. Ashburner M, et al. *Nat Genet* 2000;25:25–9. [PubMed: 10802651]
23. Smith CL, Goldsmith CA, Eppig JT. *Genome Biol* 2005;6:R7. [PubMed: 15642099]
24. Voigt B, et al. *Physiol Genomics* 2008;32:335–42. [PubMed: 18042833]
25. Aitman T, et al. 'A community view of progress and prospects in rat genetics' *Nat Genet.* 2008This Issue

Box 1 – Current Rat Genome statistics compared to Mouse and Human based on queries at NCBI (March 25, 2008)

	Rat (RGSC v3.4)	Mouse (NCBI m37)	Human (NCBI 36)
Genome Size	2.51 Gbp	3.42 Gbp	3.25 Gbp
Assembly status	Draft	Finished	Finished
Protein Coding Genes with RefSeqs	24948	29171	24291
Homologene	17868	19504	19436
Complete mRNA	80510	110870	97822
EST	869431	4923946	8251867
UniGene Clusters	81726	106980	157342

Figure 1.

GViewer visualization of rat genes and QTL associated with cardiovascular disease. The genes and are QTL are shown aligned with the rat genome which has been colored according to the corresponding syntenic regions in Human. 586 rat genes, 388 rat QTL and 127 rat strains have been associated with various types of cardiovascular disease. These are all available through the RGD Cardiovascular Disease Portal along with 588 Human and 559 Mouse genes and 49 Human QTL.