

---

**CHAPTER 6**

---

# Mouse and Rat Genome Informatics

JUDITH A. BLAKE\*, JANAN EPPIG and CAROL J. BULT

*The Jackson Laboratory*  
600 Main Street  
Bar Harbor, ME 04609, USA

---

- 6.1 Introduction
  - 6.1.1 Bioinformatics for mouse and rat geneticists
  - 6.1.2 Data integration: the challenge and the conundrum
- 6.2 The model organism databases for mouse and rat
  - 6.2.1 The Mouse Genome Database
  - 6.2.2 Mouse Genome Informatics
  - 6.2.3 RatMap
  - 6.2.4 The Rat Genome Database (RGD)
- 6.3 Mouse genetic and physical maps
  - 6.3.1 Mouse DNA mapping panels
  - 6.3.2 Mouse radiation hybrid maps
  - 6.3.3 The Jackson Lab Radiation Hybrid Map
  - 6.3.4 The EBI Radiation Hybrid Database
  - 6.3.5 Mouse physical maps
- 6.4 Rat genetic and physical maps
  - 6.4.1 Rat genetic maps
  - 6.4.2 Rat radiation hybrid maps
  - 6.4.3 Rat physical maps
- 6.5 Genome sequence resources
  - 6.5.1 Mouse genome sequencing initiative
  - 6.5.2 Mouse genome sequence resources
  - 6.5.3 Mouse cDNA clone resources
  - 6.5.4 Rat genome sequencing initiative
- 6.6 Comparative genomics
- 6.7 From genotype to phenotype
  - 6.7.1 Genetic variants
  - 6.7.2 Mouse Single Nucleotide Polymorphism (SNP) databases
  - 6.7.3 Induced mutant resources

---

\* corresponding author

- 6.7.4 Resources for mouse strain characterization
  - 6.7.5 Phenotypic variants
  - 6.7.6 ENU mutagenesis centres
  - 6.8 Functional genomics
    - 6.8.1 Gene ontology
  - 6.9 Rodent disease models
  - 6.10 Summary
    - Acknowledgements
    - References
- 

## 6.1 INTRODUCTION

Mouse and rat genome informatics is grounded in work on mouse and rat genetics and physiology that has been on-going since the early 20th century. The mouse, with its short generation time, small size, and plethora of phenotypic variants excelled as a tool for genetic investigations, especially after the conceptualization and creation of inbred strains, work begun by C. C. Little (Little and Tyzzer, 1916). Genetic crosses between inbred strains led to detailed mapping of genes and phenotypes, the construction of linkage groups, the development of chromosomal mapping techniques and the investigation of genetic components of phenotypes including diseases. Of particular significance was the development of specialized strains for genetic testing and technologies for manipulating the mouse genome. Standard inbred strains, their various derivatives, and ‘boutique’ mice developed through mutagenesis and genetic engineering have become essential tools. Coupled with advances in micro-technologies that are enabling detailed physiological studies in mice, the rich understanding of mouse genetics is accelerating the studies of genotype–phenotype relationships.

The rat, in contrast, was valued especially for its larger size relative to the mouse, and thus better suitability for physiological studies and experimental interventions. For rat, much is known about diseases, component factors in resistance/susceptibility, and specific networks of disease processes. Areas of research have been broad, including immunology, cancer, diseases of specific organ systems (cardiovascular, urogenital, skeletal, behaviour, growth and metabolism), neurological diseases, haematologic disorders, toxicology, histology, endocrinology, pathophysiology, and pharmacology (Gill *et al.*, 1989; James and Lindpaintner, 1997). The genetics of the rat lagged behind until recently, when genomic tools (expressed sequence tags or ESTs, radiation hybrid and physical maps) for rat have rapidly been created and developed.

Today, rat and mouse are both strong animal models for the investigation of biology particularly with regard to human biology and disease. The availability of two rodent animal models is also fortuitous because it permits the examination of genetic and phenotypic variation between two closely related organisms and the ability, then, to contrast that information with knowledge about the biology of humans.

### 6.1.1 Bioinformatics for Mouse and Rat Geneticists

The term ‘bioinformatics’ is used to refer to many aspects of the intersection of computer science, biology, and information science. The term is often equated with the informatics challenges of the genome projects. There are several reasons for this. First, the genome sequencing efforts generate enormous volumes of electronic data that must be organized, stored, and analysed using powerful computers and sophisticated algorithms from the

inception of the project. Second, substantial fiscal resources are being devoted to these projects, so the advancement of the informatics component is both absolutely necessary and well funded. Finally, there is the high visibility of the genome projects, with frequent newswatches about the discovery of new and interesting genes. As a result of these forces, many scientists think of bioinformatics as an endeavour focused solely on the management and analysis of sequence data.

However, all aspects of biological investigation benefit from the ordered assembly of the information and from the use of computer technologies to store, query, sort and manage biological data. Prior to a database implementation, many structured datasets about mouse genetics and heritable mutants were maintained manually. The first gene description catalogue for mouse was published in 1941 by Dr George Snell (Snell, 1941). As early as the 1950s Dr Margaret Green began compiling mouse linkage and mapping data on index cards. Linkage maps were drawn by hand and published annually in the *Mouse Newsletter* from 1965–1994. Compilations of mutant genes and polymorphic loci, chromosome atlases, and lists of synteny homologies between mouse and man were irregularly published in journals (cf. Eppig, 1992; Nadeau *et al.*, 1991; Staats, 1985) in addition to books such as *Genetic Variants and Strains of the Laboratory Mouse* (Green, 1981; Lyon and Searles, 1989; Lyon *et al.*, 1996). During the 1980s many of these resources began to be maintained electronically and resulted in an early publicly accessible mouse database GBASE (Doolittle *et al.*, 1991) and the Encyclopedia of the Mouse Genome software tools (Eppig *et al.*, 1994). During the 1990s, this sweep of information about the genetics and biology of the laboratory mouse was integrated and brought fully into electronic form with the construction of the Mouse Genome Database (<http://www.informatics.jax.org/>; Richardson *et al.*, 1995) and the development of computer programs to manipulate and query the data such as MapManager (Manly, 1993). In addition, large-scale mapping projects redefined the management of genetic data (Dietrich *et al.*, 1992) and led to the construction of additional bioinformatics resources for mouse geneticists.

Compilations for rat information developed in a different way. Billingham and Silvers published the first compilation of rat strain information in 1959 (Billingham and Silvers, 1959). A standard nomenclature for rat strains emerged in 1973 (Festing and Staats, 1973). Rat strain descriptions were catalogued (Greenhouse *et al.*, 1990), and later maintained electronically by M. F. W. Festing and made publicly available in the model organism databases. Gene data was published sporadically and accumulated slowly due to the emphasis of rat researchers on physiology rather than genetics. The pressure for databases and computational tools for rat has been a recent occurrence. Although RatMap, which exclusively curates mapped genes was started in 1993, the need for resources to manage genomic data (simple sequence length polymorphisms or SSLPs, ESTs, comprehensive gene data, genomic sequence, etc.) was not recognized as critical until the joint US–German Rat Genome Project began generating large volumes of data in the mid/late 1990s. This recognition led to the development of the Rat Genome Database (<http://rgd.mcw.edu>) described more fully below.

### 6.1.2 Data Integration: The Challenge and the Conundrum

The advent of the Internet and the development of the www permitted the development of multiple sites committed to the presentation of biological data relative to the mouse and rat. Some, such as the sequence repositories GenBank (<http://www.ncbi.nlm.nih.gov/>; Wheeler *et al.*, 2001) and EMBL (<http://www.ebi.ac.uk/embl/>; Stroesser *et al.*, 2001), include mouse and rat sequences along with sequences from all other species. Others, such as the Whitehead Institute for Biomedical Research/MIT Center for Genome Research site

(<http://www-genome.wi.mit.edu/cgi-bin/mouse/index>) provide specialized mouse datasets such as the pages for the 'Genetic and Physical Maps of the Mouse Genome'. For investigators, the reality is that information about the genetics and genomics of the laboratory mouse and the rat are found throughout cyberspace. Standards for nomenclature or descriptions of experimental data are not uniformly implemented, and it is often difficult to equate information at one site with information at another. Consequently, the investigator spends much time looking for data, collecting the data, and then manipulating the data before being able to explore and mine the data for knowledge. This has not gone unnoticed by data providers, but efforts to standardize and integrate information are often stymied by the variety of data types, the variability in data annotation, and the diversity of needs of the users. This presents a conundrum for bioinformatics professionals. Scientists do not want to be forced to use standard nomenclature or terminologies in the publication of their own work, but they do want to find a suite of information about a set of genes or sequences without having to do the data integration themselves.

The solution is easy to define, but hard to implement. It is dependent more on the sociology of doing science rather than the need for a technological solution. Data integration requires the implementation of standards and structures across multiple information resources (Bult *et al.*, 2000). Key strategies for data integration are the use of accessioned data entities, the application of nomenclature standards for key objects such as genes and strains, and the use of controlled, structured vocabularies and ontologies for functional annotation of biological information. Most of the larger data providers of interest to mouse and rat geneticists are now working to implement shared standards and to provide curated links between the different resources. Much harder is the integration of the scientific literature. As yet, most authors are unaware of and/or are not required to use standard nomenclature for genes, proteins, anatomy or biochemical reactions in the publication of laboratory research results. The result is that it is more difficult than it needs to be to bring experimental data into electronic form and to integrate it with other information. Hopefully, the use of data and nomenclature standards will become more common as scientists of all types recognize the value of bioinformatics resources and consequently appreciate the necessity and the power of data integration.

## 6.2 THE MODEL ORGANISM DATABASES FOR MOUSE AND RAT

One approach to integration of information about mouse and rat has been the construction of model organism databases. Several issues swirl around informatics sites devoted to model organisms. On the one hand, better interoperability among large data providers might obviate the need for an organisms-specific site. On the other hand, for model organisms such as *Saccharomyces*, *Caenorhabditis elegans*, *Drosophila* and others, including mouse and rat, there is a need for a central site that integrates all kinds of information about these well-studied species. Various approaches to shared data structures and standards are continually under discussion and have resulted in the increased similarities and links between the model organisms databases. Will there ultimately be one information system for all biology? Or will there continue to be specialized model organism sites loosely connected with other bioinformatics servers? The interconnectivity and transparency between bioinformatics resources continues to evolve, and it is imprudent to envisage bioinformatics systems just a few years hence. Today there exist model organism databases for the mouse and the rat, the Mouse Genome Database and the Rat Genome Database. Both work to provide comprehensive access to experimental and consensus data about these model organisms.

### 6.2.1 The Mouse Genome Database

The Mouse Genome Database (MGD) (<http://www.informatics.jax.org>) is the original model organism database for the laboratory mouse (Blake *et al.* 2001). Derived from the merger of several small specialized databases in 1994, MGD now focuses on the integrated representation of genotype (sequence) to phenotype data for the mouse with a particular emphasis on information about genes and gene products. MGD provides official gene nomenclature for the research community and works closely with human and rat genome curators to implement common standards for annotation of genes and other genome features. As part of the Mouse Genome Informatics (MGI) system (see below), MGD focuses on data integration through representations of relationships between genes, sequences and phenotypes, the representation of mouse mapping data, the association of genes to the Gene Ontology (GO), the description of targeted mutations and other alleles, and the curation of mammalian orthologies.

### 6.2.2 Mouse Genome Informatics

MGD is one component of the Mouse Genome Informatics (MGI) consortium based at The Jackson Laboratory. Other components of the MGI consortium include the Gene Expression Database (GXD; Ringwald *et al.*, 2001), the Mouse Tumor Biology Database (MTB; Bult *et al.*, 2001) and the Mouse Genome Sequencing Project (MGS). GXD focuses on the presentation of detailed experimental data about time and place of gene expression during development. MTB provides web-based access to mouse models of human cancers including experimental data and genotype-specific information. MGS works with the public mouse genome sequencing coalition to link the emerging genome with the mouse biological information. Overall, then, the MGI project provides the research community with a canonical set of mouse genes, their official names and genome locations, sequences, mammalian homologies, expression and functional information, phenotypic alleles and variants, associated literature and extensive links to other bioinformatics resources. This highly-integrated system is complemented with many cross-links to genetic and genomic resources for other organisms.

### 6.2.3 RatMap

RatMap (<http://ratpmap.gen.gu.se>) focuses on presenting the subset of rat genes, DNA markers, and quantitative trait loci (QTL) that are localized to chromosomes. RatMap maintains a highly-curated set of data, including nomenclature, chromosomal assignment and localization, mapping method statements, human and mouse homologues, references, and links to nucleotide sequences, UniGene and Rat Genome Database (RGD). In addition, RatMap maintains the rat idiograms and current cytogenetic maps. RatMap also provides a 'gene and position predictor' (GAPP) report that presents predicted positions for over 6000 rat genes based on conserved syntenic chromosomal segments between mouse and rat (Helou *et al.*, 2001).

### 6.2.4 The Rat Genome Database (RGD)

The Rat Genome Database (RGD, <http://rgd.mcw.edu>; Twigger *et al.*, 2002) is a collaborative effort between the Bioinformatics Research Center at the Medical College of

Wisconsin, The Jackson Laboratory and the National Center for Biotechnology Information (NCBI) to gather, integrate and make available data generated from ongoing rat genetic and genomic research efforts. Initially released in 2000, RGD includes curated data on rat genes, QTL, ESTs, sequence tagged sites (STSs) and microsatellite markers as well as details of inbred rat strains. RGD also contains detailed information on nomenclature, genetic and RH maps, mouse and human homologies, Gene Ontology data, and includes key literature citations. Research tools that are provided include 'VCMaP', a sequence-based homology tool and gene prediction and RH mapping tools. RGD is introducing disease-based curation for disease processes frequently studied in the rat. Integration of the emerging rat genomic sequence is also planned.

### 6.3 MOUSE GENETIC AND PHYSICAL MAPS

The genetic map of the mouse has been built over time through the contributions of many research groups, using a variety of methods, including, but not limited to, backcross, intercross and complex cross analyses, congenic strain analysis and recombinant inbred and recombinant congenic strain analyses. Chromosomal rearrangements, somatic cell hybrids and *in situ* hybridization are used to supplement these methods. These diverse methods, utilizing a wide variety of laboratory and wild-derived mouse strains, have been used to develop the consensus linkage map for mouse (MGD, [http://www.informatics.jax.org/searches/linkmap\\_form.shtml](http://www.informatics.jax.org/searches/linkmap_form.shtml)). For many purposes, this map is a standard for understanding the overall genomic organization of the mouse and for identifying potential candidate genes for diseases in particular regions.

#### 6.3.1 Mouse DNA Mapping Panels

The development of large interspecific and intersubspecific crosses, for which progeny DNA are stored for cumulative genotyping, provides single-source high-resolution linkage maps containing thousands of markers and with well-defined crossover points (cf. Avner *et al.*, 1988; Copeland and Jenkins, 1991; Dietrich *et al.*, 1992; European Backcross Collaborative Group, 1994; Rowe *et al.*, 1994). Any newly discovered gene for which DNA polymorphism is detectable between the original parental strains can be mapped immediately without setting up a *de novo* cross and the cumulative data can be used to explore questions of recombination distribution across the genome and crossover interference. These DNA backcross panels are, however, not suitable for mapping new genes that are only defined by phenotype.

Genotyping data for individual progeny from many of these DNA mapping panels are available through the Mouse Genome Database ([http://www.informatics.jax.org/searches/crossdata\\_form.shtml](http://www.informatics.jax.org/searches/crossdata_form.shtml)). In addition, maps can be generated using these data via the MGD Map Building tool at [http://www.informatics.jax.org/searches/linkmap\\_form.shtml](http://www.informatics.jax.org/searches/linkmap_form.shtml). Two of these DNA mapping panels are also maintained at specific websites: The Jackson Laboratory DNA Mapping Panels (<http://www.jax.org/resources/documents/cmdata/bkmap>) and the Whitehead Institute for Biomedical Research/MIT DNA Mapping Panels (<http://www-genome.wi.mit.edu/cgi-bin/mouse/index#genetic>).

#### 6.3.2 Mouse Radiation Hybrid Maps

Recombination maps from DNA mapping panels provide unambiguous placement of gene order. However, for very closely linked genes, these maps may not be able to resolve

locus order. For mouse, a radiation hybrid (RH) panel (T31) of 100 cell lines developed from a 3000-rad irradiated primary cell line from mouse embryo fused with hamster fibroblast has been developed (McCarthy *et al.*, 1997). Radiation hybrids can be used for high throughput mapping and high resolution of locus order because each hybrid cell line contains a highly fragmented subset of the mouse genome. The co-retention of mouse genes across the 100-cell panel is indicative of their relative distance apart, assuming random chromosomal breakage and leads to the construction of RH maps (cf. Van Etten *et al.*, 1999). Two complementary databases serve as community resources for gathering, distributing and analysing the T31 RH data.

### 6.3.3 The Jackson Lab Radiation Hybrid Map

The JAX RHmap provides web-based access to a comprehensive, integrated database that includes all typing data, retention frequency and log of the odds (LOD) scores for markers typed on the T31 panel, as well as RH framework maps for many of the chromosomes (<http://www.jax.org/resources/documents/cmdata/rhmap/>). All publicly available T31 data from large genome centres at the Whitehead Mouse RH Database ([http://www-genome.wi.mit.edu/mouse\\_rh/index.html](http://www-genome.wi.mit.edu/mouse_rh/index.html)), the UK Mouse Genome Centre ([http://www.mgc.har.mrc.ac.uk/physical/est\\_mapping/est.html](http://www.mgc.har.mrc.ac.uk/physical/est_mapping/est.html)) and Genoscope–CNS ([http://www.genoscope.cns.fr/externe/English/Projets/Projet\\_ZZZ/rhmap.html](http://www.genoscope.cns.fr/externe/English/Projets/Projet_ZZZ/rhmap.html)), as well as from many individual laboratories are included.

The website includes an electronic submission interface for depositing RH typing data from users, data error checking and quality control, technical support, data analysis and the development of RH maps. All data, with references and experimental notes can be viewed or downloaded. Data are shared with the Mouse Genome Database (MGD) and the EBI data repository (RHdb, below).

### 6.3.4 The EBI Radiation Hybrid Database

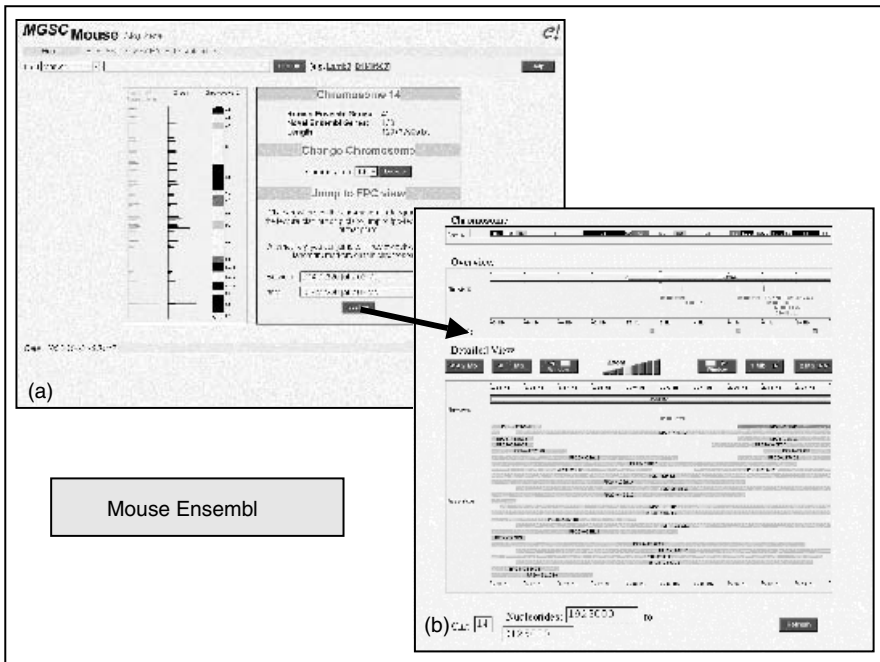
The European Bioinformatics Institute (EBI) Radiation Hybrid Database (RHdb) is a repository for the raw data for constructing radiation hybrid maps, STS data, scores and experimental conditions (Rodriguez-Tomé and Lijnzaad, 2001; <http://www.ebi.ac.uk/RHdb/index.html>). The EBI RHdb is designed to be a species-neutral database, and currently contains human, mouse, and rat RH data. Data content relies entirely on submissions from data providers and research groups. Maps are not assembled from the accumulating data, but maps may be submitted by data developers.

### 6.3.5 Mouse Physical Maps

Two genome centres have produced physical maps for mouse that are accessible via the Internet: the Whitehead Institute/MIT (<http://www-genome.wi.mit.edu/cgi-bin/mouse/index#phys>) and the UK Mouse Genome Centre at Harwell (<http://www.mgc.har.mrc.ac.uk/physical/phys.html>). Whitehead Institute/MIT data include contigs and STS content mapping across the entire mouse genome and utilizes existing SSLP markers that characterize the MIT genetic map of the mouse to tie the physical and recombination maps together. The UK Mouse Genome Centre data consists of physical maps of selected regions of the genome that are being developed in association with individual research interests, notably regions of chromosomes 13 and X. Data from these sites are integrated into MGD, as well as being available from the originator's site.

A physical map of the genome of the C57BL/6J strain of laboratory mouse has been constructed using Bacterial Artificial Chromosome (BAC) clones (Gregory *et al.*, 2002). This map serves as the framework for the Mouse Genome Sequencing initiative (described below). The current BAC map for the mouse was derived from 305,768 BAC clones from two libraries: RPCI23 (female) and RPCI24 (male) (Osoegawa *et al.*, 2000). These libraries are available for distribution to the scientific community through the BACPAC Resource at the Children's Hospital Oakland Research Institute (<http://www.chori.org/bacpac>). The RPCI23 library is also available through Research Genetics (<http://www.resgen.com/products/RPCI23MBAC.php3>).

The clones from the RPCI BAC libraries were fingerprint mapped at the Genome Sequencing Centre in Vancouver, British Columbia (Marra *et al.*, 1997; [http://www.bcgs.bc.ca/projects/mouse\\_mapping/](http://www.bcgs.bc.ca/projects/mouse_mapping/)). The fingerprint data were combined with BAC end sequence data (Zhao *et al.*, 2001; [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)) to produce a mouse physical map that contains 296 contigs and covers an estimated 2,739 Mb (Gregory *et al.*, 2002). The average length of the contigs is 9.3 Mb. Of the 296 contigs, 228 can be localized to a chromosome. Approximately 97% of the total clone coverage for the mouse genome (2,658 Mb in 211 contigs) can be aligned to the human genome sequence.



**Figure 6.1** Mouse Ensembl. A graphical representation of the clone-based physical map for the proximal end of mouse chromosome 14 from Ensembl. This browser allows users to search for regions of a chromosome between two STS markers and to view the current clone coverage in the selected area. Because the browser is web-based, users do not have to download and install special software to view the BAC map (See Colour Plates).

There are three ways to view the current status of the mouse BAC physical map. Researchers can download and install a software product called FPC from the Sanger Institute (<http://www.sanger.ac.uk/Software/fpc/>) (Soderlund *et al.*, 2000) and use this software to graphically display the BAC clone fingerprint data generated by the Genome Sequence Centre in Vancouver. A similar display tool called the internet Contig Explorer (iCE) is available from the Genome Sequence Centre in Vancouver (<http://ice.bcgsc.bc.ca/>). An option for viewing the map that does not require the installation of software is to view the physical map using the Ensembl mouse browser at the Sanger Institute (<http://mouse.ensembl.org/>); or the mouse MapViewer of NCBI (<http://www.ncbi.nlm.nih.gov>).

The ultimate physical map, of course, is the genome sequence itself. Despite the expectation that the mouse genome will be available soon, the need for genetic maps and other physical maps will not disappear. The mouse sequence will continue to be built, reassembled and re-annotated for many years to come, making the physical contig map an important resource for anchoring this new information as it develops. Genetic maps will be needed indefinitely, for the mapping of QTLs, spontaneous mutations and other phenotypes with undetermined molecular defects. In addition, genetic maps are essential for studying chromosome structure and function, and recombination itself.

## 6.4 RAT GENETIC AND PHYSICAL MAPS

### 6.4.1 Rat Genetic Maps

The early development of rat genetics paralleled that of the mouse, with the establishment of genetic linkage between albino and pink-eyed dilution in both mouse and rat (Castle and Wachter, 1924; Dunn, 1920). Haldane (Haldane, 1927) recognized that, if these genes were homologues, they represented conserved synteny over evolutionary time. Subsequently, research geneticists focused on mouse and the rat became the major tool for physiologists. Thus, the development of the rat genetic map began to lag behind that of the mouse. As of 1991 there were 214 genes mapped in rat (Levan *et al.*, 1991) in contrast to nearly 3000 genes mapped in mouse (Hillyard *et al.*, 1991). This disparity in the number of genes mapped has continued to this day, with 1576 genes currently mapped in rat (RatMap, 2002) versus 18,983 in mouse (MGD, 2002). Maps of rat genes are largely cytogenetic rather than recombination maps and are maintained by RatMap (<http://ratmap.gen.gu.se>). After a century of concentrated use of rat by physiologists, rat genetics is now undergoing a revival as genomic tools are developed and its genome is finally being sequenced.

The resurgence of interest in the rat map has paralleled the development of genomic resources for rat. In the 1990s the first rat genome projects were begun to generate ESTs, YAC and BAC libraries, and SSLP maps. There were a number of backcrosses and intercrosses made among rat strains that were used to develop SSLP maps with several hundred to a few thousands markers (cf. Bihoreau *et al.*, 1997; Brown *et al.*, 1998; Dracheva *et al.*, 2000; Watanabe *et al.*, 2000; Wei *et al.*, 1998). Most of these SSLP maps are not yet integrated, although SSLP data and maps for some of the crosses are available through RGD. Data from two F2 intercrosses have been integrated and the resulting map containing 4786 SSLP markers can be found at the Whitehead Institute (<http://www-genome.wi.mit.edu/rat/public/>). In parallel, a large collaborative Allele Characterization Project was begun to establish allele sizes of 8000 SSLPs among 48 genetically and physiologically important inbred rat strains ([http://www.brc.mcw.edu/LGR/research/lgr\\_acp](http://www.brc.mcw.edu/LGR/research/lgr_acp)).

html). Data generated from this project will provide investigators with a means of quickly selecting informative markers for new and existing mapping crosses.

### 6.4.2 Rat Radiation Hybrid Maps

A rat whole genome radiation hybrid panel (T55) generated by Linda McCarthy in Peter Goodfellow's laboratory has been used to construct high-resolution maps of the rat genome ([http://www.well.ox.ac.uk/rat\\_mapping\\_resources/rat\\_radiation\\_hybrid\\_maps.html](http://www.well.ox.ac.uk/rat_mapping_resources/rat_radiation_hybrid_maps.html)). The first radiation hybrid map was based on 5255 markers and included both microsatellites and known genes (Watanabe *et al.*, 1999). Another map using the same panel was constructed as a framework map using 2000 evenly spaced markers (<http://rgd.mcw.edu/RHMAPSERVER/>; Steen *et al.*, 1999). Both sites provide RH map web servers for users to map their markers — users submit data to the Rat RH Map Server and a map placement with a summary report is returned.

### 6.4.3 Rat Physical Maps

In contrast to the mouse, the rat has no genome-wide clone-based physical maps, only a few for specific regions such as the MHC locus (Gunther and Walter, 2001; Ioannidu *et al.*, 2001). Most of the 'physical map' for the rat genome consists of the cytogenetic maps that are maintained in RatMap and include a fair amount of FISH data. A physical BAC map of the rat is in preparation, as part of the NHGRI-sponsored rat genome sequencing initiative. A BAC library (CHORI-230) from the BN/SsNHsd/MCW (Brown Norway) strain of laboratory rat has been prepared using the same methods as were used for the mouse BAC libraries (<http://www.chori.org/bacpac/>) (Osoegawa *et al.*, 2000). The BAC clones from this library are being fingerprint mapped by the Genome Sequencing Centre in Vancouver, Canada ([http://www.bcgsc.bc.ca/projects/rat\\_mapping/](http://www.bcgsc.bc.ca/projects/rat_mapping/)). There are currently (late 2001) 136,195 clones in their database. The BAC ends for this library are being sequenced at The Institute for Genomic Research (TIGR; [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)).

## 6.5 GENOME SEQUENCE RESOURCES

### 6.5.1 Mouse Genome Sequencing Initiative

The initiative to sequence the genome of the laboratory mouse was announced by the National Human Genome Research Institute (NHGRI) of NIH in September 1999 as part of an overall 'action plan' for mouse genomics (Battay *et al.*, 1999). The goals of the initiative were to have a working draft of the genome of the C57BL/6J strain of mouse completed by 2003 and the finished genome sequence by 2005. The initial strategy for obtaining the mouse genome sequence was to build a physical BAC map of the genome as the BAC clones were sequenced (<http://www.nhgri.nih.gov/NEWS/MouseRelease.htm>).

In October of 2001 the strategy for obtaining the mouse sequence changed to include a whole genome shotgun approach. Part of the rationale for this change in sequencing strategy was that the shotgun sequences for the mouse genome could be used to assist in the identification of genes in the working draft of the human genome. The sequencing

centres of the Sanger Institute, Washington University Medical Centre and the Whitehead Institute for Biomedical Research were funded to generate whole genome shotgun data for the mouse (<http://www.nih.gov/science/models/mouse/>).

Simultaneously with this shift in sequencing strategy, NIH launched a program to sequence mouse BAC clones that covered genomic regions of high biological interest. Individual investigators were invited to submit applications requesting specific BACs to be sequenced. Several sequencing centres, including the Cold Spring Harbor Laboratory, Harvard University Medical School and the University of Oklahoma were funded to sequence these BACs (<http://www.nih.gov/science/models/bacsequencing/>). The NIH BAC sequencing program was initially restricted to clones from specific BAC libraries for the mouse. However the program now accepts applications for the sequencing of clones from any BAC library and also from organisms other than mouse.

Several other sequencing centres around the world are using their sequencing capacity for regional and/or comparative sequencing of the mouse genome. For example, the DOE-funded Joint Genome Institute focused on sequencing segments of the mouse genome that are homologous to human chromosome 19 (<http://bahama.jgi-psf.org/pub/ch19/>; Dehal *et al.*, 2001). The Medical Research Council (MRC) is focusing on sequencing of mouse chromosomes 2, 4, 13 and mouse–human comparative sequencing for chromosome X (<http://mrcseq.har.mrc.ac.uk/>). Although the primary focus of the Baylor College of Medicine genome centre is now on sequencing the rat genome, it originally focused on sequencing BACs across mouse chromosome 11.

The NCBI maintains a status report of the progress of the mouse genome sequence project (<http://www.ncbi.nlm.nih.gov/genome/seq/MmHome.html>) as well as a registry of BAC clones that are being sequenced under the auspices of the Trans-NIH BAC Sequencing Program (<http://www.ncbi.nlm.nih.gov/genome/clone/cstatus.html>).

## 6.5.2 Mouse Genome Sequence Resources

There are several ways to access mouse genome sequence (here we focus on freely-accessible public resources). The whole genome shotgun data for the mouse can be found in a ‘Trace Archive’ maintained by the NCBI and can be searched via BLAST (<http://www.ncbi.nlm.nih.gov/blast/mmtrace.html>). A similar resource is maintained at the European Bioinformatics Institute (EBI; <http://www.ebi.ac.uk/blast2>). As of December 2001, there were over 31 million sequencing reads available in these archives; greater than six times the coverage of the mouse genome.

The Mouse Genome Sequencing Consortium has released an annotated draft assembly of the mouse genome to the research community (Mouse Genome Sequencing Consortium, 2002). The current draft assembly covers over 96% of the genome; a complete genome sequence for the laboratory mouse is anticipated by 2005. The draft genome and the associated annotations can be accessed using the Ensembl genome browser (<http://www.ensembl.org>), NCBI’s Map Viewer (<http://www.ncbi.nlm.nih.gov>), and the University of Santa Cruz’s genome browser (<http://genome.ucsc.edu>).

Other genome resources include MouseBLAST (Figure 6.2), a server maintained by the MGS group at The Jackson Laboratory that allows researchers to connect mouse sequence data with the wealth of biological knowledge about the mouse available in the MGI. Finally, the Mouse Genome Resources pages at NCBI (<http://www.ncbi.nlm.nih.gov/>

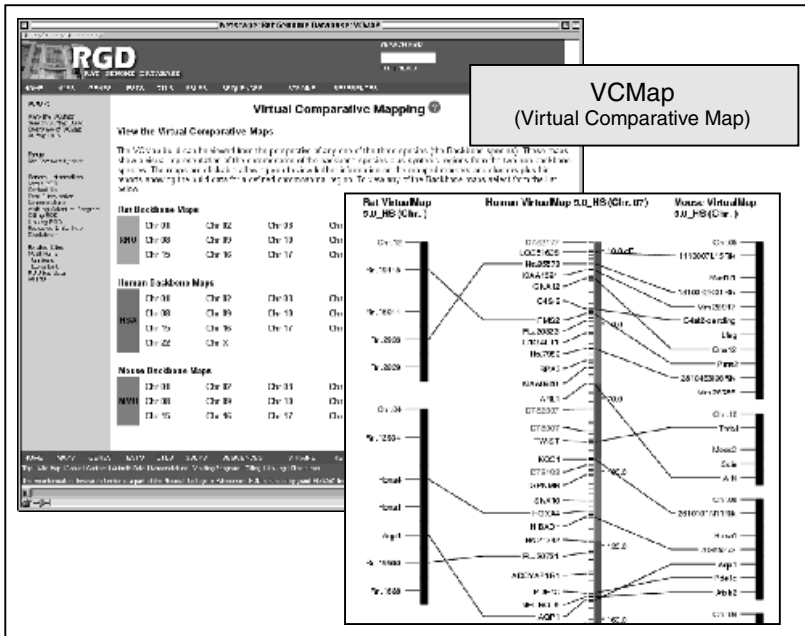


### 6.5.4 Rat Genome Sequencing Initiative

In February 2001, the National Heart, Lung, and Blood Institute (NHLBI) announced funding support for the sequencing of the rat genome ([http://www.nhgri.nih.gov/NEWS/nih\\_expands\\_programs.html](http://www.nhgri.nih.gov/NEWS/nih_expands_programs.html)). Three sequencing centres have been funded to produce enough genome sequence data to have a working draft of the rat genome by 2004: Celera Genomics, Baylor College of Medicine Genome Sequencing Centre and Genome Therapeutics, Inc.

## 6.6 COMPARATIVE GENOMICS

The sequencing of both the mouse and rat genomes promises to stimulate research based on comparative genome organization and comparative analysis between the human, mouse



**Figure 6.3** Virtual Comparative Map. The Virtual Comparative Map is generated using sequence-based algorithms that predict syntenic regions inferred from homology among mapped sequences. Sequence comparisons between ESTs and cDNAs from human, mouse and rat are combined with Radiation Hybrid map locations to define regions of synteny. Locations for unmapped markers in a species are then predicted based on the map location of the orthologous marker in a syntenic region of another species. The forepanel shows a virtual comparative map using human as the backbone map (centre) and syntenic regions of rat (left) and mouse (right). Mapped genes, UniGenes and STSs are shown, with lines connecting predicted homologues among the species. Data sources for the virtual maps are RGD, NCBI and MGD. The virtual comparative maps are available at <http://rgd.mcw.edu/VCMAP/> (See Colour Plates).

and rat. Research papers based on comparison of large conserved segments between mouse and human are being published (Dehal *et al.*, 2001; Glusman *et al.*, 2001). Another approach is to use genome comparisons for elucidation of a suite of comparable genome features such as transcription factors (Wasserman *et al.*, 2000). Computational approaches to uncovering conserved regions such as exons or regulatory sites facilitate the discovery of new important genome features (Oeltjen *et al.*, 1997).

The direct comparison of genomic sequence from conserved linkage groups between mouse and human (and other organisms) has proven to be an effective strategy for identifying biologically relevant regions (coding and non-coding) in genomes. Two of the most commonly used tools for this effort are VISTA (<http://www-gsd.lbl.gov/vista/>; Mayor *et al.*, 2001) and PIPMAKER (<http://bio.cse.psu.edu/pipmaker/>; Schwartz *et al.*, 2000). These resources allow researchers to submit large genomic sequence regions to be aligned and analysed for the presence of conserved sequence elements. The VISTA group provides a set of pre-aligned sequences of mouse and human from finished genomic data in GenBank (<http://pipeline.lbl.gov/>). Applications include determining all of the protein-coding segments in both species, locating regulatory signals, understanding the mechanisms and history of genome evolution and deducing the similarities and differences in gene organization between the species of interest.

Other comparative map viewers incorporate information about the rat. One resource is the Gene and Position Predictor (GAPP) produced by RatMap which provides predicted comparative maps using known gene orthologues and zoo-FISH data (<http://gapp.gen.gu.se/Description.html>; Nilsson *et al.*, 2001). A different type of predictive map is the Virtual Comparative Map (VCMAP) (<http://rgd.mcw.edu/VCMAP/>; Figure 6.3). These maps are generated using sequence-based algorithms that predict syntenic regions inferred from homology among mapped sequences. The Otsuka GEN Research Institute posts a genome-wide comparative map of the rat based primarily on extensive RH mapping data ([http://ratmap.ims.u-tokyo.ac.jp/cgi-bin/comparative\\_home.pl](http://ratmap.ims.u-tokyo.ac.jp/cgi-bin/comparative_home.pl)). Finally, maps of curated orthologues for mouse/rat/human are available from MGD ([http://www.informatics.jax.org/menus/homology\\_menu.shtml](http://www.informatics.jax.org/menus/homology_menu.shtml)).

## 6.7 FROM GENOTYPE TO PHENOTYPE

Beyond a generalized representation of the mouse and rat are the intricacies of differences due to differing genetic backgrounds that can be revealed by comparisons between strains, among the rodent species, between rodents and other mammals and even between more distantly related organisms. The publication of the mouse genome sequence and the promise of the rat genome sequence in the near future will facilitate systematic genome-wide approaches to investigate normal and disordered cellular and physiological states. Genome-wide surveys of gene expression or genotype variation will enhance the gene-by-gene approach to the assessment of gene function. Scientists have long known of the importance of genetic background in the analysis of gene function or dysfunction due to the phenotypic variability resulting from epistatic interactions. Now, it may be possible to precisely assess the effect of genotype variability on the expression, function and interaction of gene products. As ever, the challenge for bioinformaticians will be to integrate the data from various experimental approaches into a coherent representation of the model organism. Ideally, one would like to query for a set of gene products expressed at the same time/state, evaluate the effect of genotype on the function and phenotypic presentation of variant gene products or compare ‘snapshots’ of cellular component sets between tissues or strains of rodents.

### 6.7.1 Genetic Variants

Genetically-engineered strains of mice including mice altered by gene transfer (transgenics), homologous recombination (gene targeting) and chemical mutagenesis provide powerful new tools for biomedical research. The use of these strains has become critical for basic research and for investigating causes of and potential treatments for human disease. The number of genes in mice that have enough characterization to be given descriptive names now exceeds 12,000, perhaps one-third or one-quarter of the estimated total number of genes. Genome manipulation techniques that target specific genes (e.g. knock-outs, knock-ins, and conditional mutations) or that identify sequence variants (e.g. microsatellites and single nucleotide polymorphisms or SNPs), are providing new alleles for biological analysis. Although many factors can contribute to a phenotype, a widely used research approach focuses on the isolated effects of single genes and their mutant alleles on biological systems. An alternative approach is to study quantitative traits where multiple genes contribute to the observed phenotypes. Here a one-to-one relationship between gene and phenotype does not exist and, as in humans, the discovery of the genes underlying complex traits such as obesity and hypertension continues to be challenging, but should become more tractable as new mapping resources are developed.

### 6.7.2 Mouse Single Nucleotide Polymorphism (SNP) Databases

SNP technologies are being exploited for the investigation of human syndromes and diseases (Schork *et al.*, 2000). Human SNP resources such as dbSNP (see Chapter 3) provide access to high-density SNP maps for humans. Large-scale discovery and genotyping of SNPs in mice is underway (Lindblad-Toh *et al.*, 2000) and a limited quantity of mouse SNP data is already available in the Roche mouse SNP database (<http://mousesnp.roche.com/>) and the Whitehead/MIT SNP database (<http://www-genome.wi.mit.edu/snp/mouse/>). With the sequencing of large genomic regions of multiple mouse inbred strains, further SNP sets for mouse will be defined and could facilitate computer-based identification of QTL loci between inbred strains; one group has already reported some success using this method, but the approach is controversial at present (Chesler *et al.*, 2001; Darvasi, 2001; Grupe *et al.*, 2001).

### 6.7.3 Induced Mutant Resources

The rapid generation of many induced mutants of the mouse through the use of technologies such as homologous recombination and targeted knock-outs has created the need for a central facility to collect and distribute them to the scientific community. The Induced Mutant Resource (IMR) (<http://www.jax.org/resources/documents/imr/>) at The Jackson Laboratory is an example of a national clearing-house for the collection and distribution of a subset of genetically-engineered mice. The IMR maintains an on-line database to provide information about these strains. This information includes a description of the mutant phenotype, husbandry requirements and links to related resources. Another resource providing mouse mutants to the community is the Mutant Mouse Regional Resource Centres (<http://www.mmrrc.org/>). The MMRRC strive to enhance the availability of genetically-engineered mice for the study of human biology and disease. The European Mouse Mutant Resource (EMMA) (<http://emma.rm.cnr.it/>) is another repository for mouse mutant stocks.

### 6.7.4 Resources for Mouse Strain Characterization

Inbred strains in mouse have been specifically generated to facilitate the study of the genetic component of phenotypes including disease phenotypes by being able to isolate

the impact of the mutant gene on a standard genetic background. With the advent of many new technologies, molecular information about the whole genome is becoming available for different inbred strains, and the need for standard evaluation of differences between inbred strains is apparent. New initiatives to study strain characteristics in mice and rats are underway with the attendant development of bioinformatics resources.

The Mouse Phenome Database (MPD; <http://www.jax.org/phenome/>) was established to provide a collection of baseline phenotypic data on commonly used and genetically diverse inbred mouse strains. Many institutions and investigators are involved in this effort to provide standard sets of strain characteristics for the most commonly used strains of mice. The MPD will enable investigators to identify appropriate strains for physiological testing and disease onset and susceptibility.

### 6.7.5 Phenotypic Variants

In contrast with the reliance on the gene-by-gene approach to discovery of functions and roles for genes and for the investigation of diseases and disorders, a recent development has been the use of systematic large-scale phenotype-driven mutagenesis studies in the mouse. This approach uses chemical or physical disruption of the genome followed by identification of putative mutants using a series of phenotypic screens for particular traits. This phenotype-driven approach to genome characterization has an important role to play in linking gene identification with gene function. This approach will allow researchers to better understand the molecular basis of diseases through the identification of mutants that develop the same or similar phenotypes but that have mutations in different genes. Furthermore, a full appreciation of the genetic basis of a disease requires that the phenotypes associated with multiple alleles of the same gene be studied to identify hypomorphs, alleles that confer gain of function, etc. Although it is unclear how much of the genome can be saturated with this approach, these projects will provide the community with a vast array of new phenotypes for biological analysis.

### 6.7.6 ENU Mutagenesis Centres

Several public large-scale ENU mutagenesis projects are already underway and are providing new models for the study of disease and gene function to the community (Brown and Nolan, 1998; De Angelis *et al.*, 2000; Justice *et al.*, 1999; Nolan *et al.*, 2000) (Table 6.1). Some of the mutagenesis centres are working in several disease areas to identify new mutants including the ENU Mutagenesis Programme at Harwell (<http://www.mgu.har.mrc.ac.uk/mutabase/>), the RIKEN Mouse Functional Genomics Group (<http://www.gsc.riken.go.jp/Mouse/>), and the GSF ENU Mouse Mutagenesis Screen Project (<http://www.gsf.de/isg/groups/enu-mouse.html>). Several of these mutagenesis centres are focusing on the identification of new mutant mice to serve as models for neurological disorders (Moldin *et al.*, 2001) including the Neuroscience Mutagenesis Facility at The Jackson Laboratory (<http://www.jax.org/nmf/>), the Neurogenomics Centre at Northwestern University (<http://genome.northwestern.edu/>), the Tennessee Mouse Genome Consortium (<http://www.tnmouse.org/>) and the McLaughlin Research Institute (<http://www.montana.edu/wwwmri/enump.html>). The mutagenesis facility at the Baylor College of Medicine (<http://www.mouse-genome.bcm.tmc.edu/ENU/MutagenesisProj.asp>) is focusing on developmental defects. The Medical Genome Centre in Australia focuses on cancer-related phenotypes ([http://jcsmr.anu.edu.au/group\\_pages/mgc/CancerGenLab.html](http://jcsmr.anu.edu.au/group_pages/mgc/CancerGenLab.html)). The Mouse Heart, Lung,

**TABLE 6.1 Mouse Mutagenesis Centres and Databases**

Mutagenesis Centre	Disease Focus	URL
ENU Mutagenesis Programme (Harwell)	General	<a href="http://www.mgu.har.mrc.ac.uk/mutabase/">http://www.mgu.har.mrc.ac.uk/mutabase/</a>
RIKEN Mouse Functional Genomics Group	General	<a href="http://www.gsc.riken.go.jp/Mouse/">http://www.gsc.riken.go.jp/Mouse/</a>
GSF ENU Mouse Mutagenesis Screen Project	General	<a href="http://www.gsf.de/isg/groups/enu-mouse.html">http://www.gsf.de/isg/groups/enu-mouse.html</a>
Neuroscience Mutagenesis Facility at The Jackson Laboratory	Neurological	<a href="http://www.jax.org/nmf/">http://www.jax.org/nmf/</a>
Neurogenomics Centre at Northwestern University	Neurological	<a href="http://genome.northwestern.edu/">http://genome.northwestern.edu/</a>
Tennessee Mouse Genome Consortium	Neurological	<a href="http://www.tnmouse.org/">http://www.tnmouse.org/</a>
McLaughlin Research Institute	Neurological	<a href="http://www.montana.edu/wwwmri/enump.html">http://www.montana.edu/wwwmri/enump.html</a>
Baylor College of Medicine	Developmental disorders	<a href="http://www.mouse-genome.bcm.tmc.edu/ENU/MutagenesisProj.asp">http://www.mouse-genome.bcm.tmc.edu/ENU/MutagenesisProj.asp</a>
Medical Genome Centre (Australia)	Cancer	<a href="http://jcsmr.anu.edu.au/group_pages/mgc/CancerGenLab.html">http://jcsmr.anu.edu.au/group_pages/mgc/CancerGenLab.html</a>
The Mouse Heart, Lung, Blood, and Sleep Disorders Centre (JAX)	Cardiovascular	<a href="http://www.jax.org/hlbs/index.html">http://www.jax.org/hlbs/index.html</a>

Blood, and Sleep Disorders Centre at The Jackson Laboratory is focusing on the identification of new mutants for cardiovascular diseases (<http://www.jax.org/hlbs/index.html>).

## 6.8 FUNCTIONAL GENOMICS

In the post-genome world, mouse and rat models will be heavily used for investigation of gene function and disease pathogenesis (Schimenti and Bucan, 1998; Temple *et al.*, 2001; Zheng *et al.*, 1999). With the completion of the mouse genome, attention can move to genome-wide screens for gene expression and systematic investigation of gene function. The inclusion of functional information with gene annotations first appeared in the sequence data repositories. From the start, issues of quality control for data associations were evident. Evaluation of sequence similarities often led to the transfer of function information from one gene annotation report to another without experimental verification or any statement about the basis for the function assertion. The first detailed functional classification was developed to catalogue the genes of *Escherichia coli* (Riley, 1993). Since then, functional annotation schemes have been developed for single organisms, multi-organism databases, and for pathway-related systems (see Rison *et al.*, 2000 for review).



## 6.9 RODENT DISEASE MODELS

The experimental manipulation of mice and rats for the purpose of creating animal models for human disease is implicit in the scientific endeavours detailed here. Mouse and rat models will continue to be the best models for experimental manipulation of the mammalian genome for the foreseeable future (Bedell, *et al.*, 1997). The fact that inbred strains exist, providing consistent homogenous genetic backgrounds for experimentation, allows the genesis of diseases characteristic of particular inbred strains to be studied, as well as the development and testing of therapeutic interventions. The occurrence of spontaneous or induced single gene mutations in these strains allows precise detailed studies of the multiple effects of that particular mutation. Targeted mutations that produce knock-out or conditional mutations permit researchers to mimic the molecular defect of human diseases. Comparative studies have uncovered many rodent mutations that reflect their counterpart human disease. Multigenic diseases and quantitative trait loci can be dissected in mice and rats using controlled crosses and through creation of specialized strains, such as congenics and consomics, which place particular parts of the genome from one strain onto the background of another strain (cf. Kwitek-Black and Jacob, 2001; Sugiyama *et al.*, 2001).

In addition to the discovery or creation of models that reflect the underlying genetics of particular disease states, researchers may also find it useful to study animal models that reflect phenotypic similarity alone. That is to say, there is a phenotypic similarity in the animal model to a human disease condition, and the animal model is useful for studying that phenotype, even though we do not know that the underlying genetic dysfunction is exactly the same. For example, many cancers have unknown genetic aetiology, but particular strains or mutants prone to the development of particular cancers can serve as effective animal models (Hann and Balmain, 2001).

The strength of using rat and mouse models clearly lies in the physiological research that has gone before and that is accelerating with micro-technology developments. A fuller understanding of the genetics of these organisms, coupled with the imminent availability of their genome sequences, will enhance our ability to analyse the functions of gene products and to dissect the molecular basis of phenotypes.

## 6.10 SUMMARY

With new technologies and methods, the pace of data acquisition only quickens. Simultaneously, there are now intense efforts underway to improve data integration and to support rapid access to and interactive use of molecular and related biological information. Biological databases and information resources existed long before the advent of computers and the internet. We are, however, yet developing and realizing the capacity that computers give us to use the databases not just as archives, but also as research tools. The future of computerized scientific databases and information resources will be in their ability to rapidly retrieve and manipulate data in response to complex queries. The full value of the information they contain can then be exploited to address outstanding scientific inquiries.

## ACKNOWLEDGEMENTS

This work is supported in part by NIH/NHGRI grant HG00330, DOE grant FG99ER 62850, and HG02273. We thank Greg Cox, Cathy Lutz and Tim O'Brien for their critical

reading of this paper; W. John Boddy for preparation of figures and Janice Ormsby for technical review.

## REFERENCES

- Avner P, Amar L, Dandolo L, Guenet JL. (1988). Genetic analysis of the mouse using interspecific crosses. *Trends Genet* **4**: 18–23.
- Batley J, Jordan E, Cox D, Dove W. (1999). An action plan for mouse genomics. *Nature Genet* **21**: 73–75.
- Bedell MA, Largaespada DA, Jenkins NA, Copeland NG. (1997). Mouse models of human disease. Part 11: Recent progress and future directions. *Genes Develop* **11**: 11–43.
- Billingham RE, Silvers WK. (1959). Inbred animals and tissue transplantation immunity. *Transplant Bull* **6**: 399–406.
- Bihoreau M-T, Gaugier D, Kato N, Hyne G, Lindpainter K, Rapp JP, *et al.* (1997). A linkage map of the rat genome derived from three F2 crosses. *Genome Res* **7**: 434–440.
- Blake JA, Richardson JE, Bult CJ, Kadin JA, Eppig JT and the Mouse Genome Database Group (2001). The Mouse Genome Database (MGD): the model organism database for the laboratory mouse. *Nucleic Acids Res* **30**: 1–3.
- Brown DM, Matisse TC, Koike G, Simon JS, Winer ES, Zangen S, *et al.* (1998). An integrated genetic linkage map of the laboratory rat. *Mammal Genome* **9**: 521–530.
- Brown SD, Nolan PM. (1998). Mouse mutagenesis — systematic studies of mammalian gene function. *Hum Mol Genet* **7**: 1627–1633.
- Bult CJ, Krupke DM, Näf D, Sundberg JP, Eppig JT. (2001). Web-based access to mouse models of human cancers: the Mouse Tumor Biology (MTB) Database. *Nucleic Acids Res* **29**: 95–97.
- Bult CJ, Richardson JE, Blake JA, Kadin JA, Ringwald M, Eppig JT and the Mouse Genome Informatics Group (2000). Mouse genome informatics in a new age of biological inquiry. In *Proceedings of the IEEE International Symposium on Bio-Informatics and Biomedical Engineering*, pp. 29–32.
- Castle WE, Wachter WL. (1924). Variations of linkage in rats and mice. *Genetics* **9**: 1–12.
- Chesler EJ, Rodriguez-Zas SL, Mogil JS. (2001). *In silico* mapping of mouse quantitative trait loci. *Science* **294**: 2423.
- Copeland NG, Jenkins NA. (1991). Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet* **7**: 113–118.
- Darvasi A. (2001). *In silico* mapping of mouse quantitative trait loci. *Science* **294**: 2423.
- De Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto B, Marschall S, *et al.* (2000). Genome-wide, large-scale production of mutant mice by ENU Mutagenesis. *Nature Gene* **25**: 444–447.
- Dehal P, Predki P, Olsen AS, Kobayashi A, Folta P, Lucas S, *et al.* (2001). Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* **293**: 104–111.
- Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC, *et al.* (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**: 423–447.
- Doolittle DP, Hillyard AL, Davisson MT, Roderick TH, Guidi JN. (1991). GBASE — The genomic database of the mouse, In *Fifth International Workshop on Mouse Genome Mapping, Lunteren, Netherlands*, p. 27.

- Dracheva SV, Remmers EF, Chen S, Chang L, Gulko PS, Kawahito Y, *et al.* (2000). An integrated genetic linkage map with 1,137 markers constructed from five F2 crosses of autoimmune disease-prone and -resistant inbred rat strains. *Genomics* **63**: 202–226.
- Dunn LC. (1920). Linkage in mice and rats. *Genetics* **5**: 325–343.
- Eppig JT. (1992). Mouse DNA clones and probes. *Mammal Genome* **3**: 300–330.
- Eppig JT, Blackburn RE, Bradt DW, Corbani LE, Davisson MT, Doolittle DP, *et al.* (1994). *The Encyclopedia of the Mouse Genome, an update. Third International Conference on Bioinformatics and Genome Research, Tallahassee*, p. 73.
- European Backcross Collaborative Group (1994). Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. *Hum Mol Genet* **3**: 621–627.
- Festing MFW, Staats J. (1973). Standardized nomenclature for inbred strains of rats, fourth listing. *Transplantation* **16**: 221–245.
- Gill TJ, Smith GJ, Wissler RW, Kunz HW. (1989). The rat as an experimental animal. *Science* **245**: 269–276.
- Glusman G, Rowen L, Lee I, Boysen C, Roach JC, Smit AF, *et al.* (2001). Comparative genomics of the human and mouse T Cell receptor loci. *Immunity* **15**: 337–349.
- GO Consortium. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genet* **25**: 25–29.
- GO Consortium. (2001). Creating the gene ontology resources: design and implementation. *Genome Res* **11**: 1425–1433.
- Green MC. (Ed.) (1981). *Genetic Variants and Strains of the Laboratory Mouse*, 1st edn. Fischer Verlag: Stuttgart.
- Greenhouse DD, Festing MFW, Hasan S, Cohen AL. (1990). Catalogue of inbred strains of rats. In *Genetic Monitoring of Inbred Strains of Rats*, Hedrich HJ (Ed.), Gustav Fischer: Stuttgart, pp. 120–132.
- Gregory SG, *et al.* (2002). A Physical map of the mouse genome. *Nature* **418**: 743–50.
- Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, *et al.* (2001). *In silico* mapping of complex disease-related traits in mice. *Science* **292**: 1915–1918.
- Gunther E, Walter L. (2001). The major histocompatibility complex of the rat (*Rattus norvegicus*). *Immunogenetics* **53**: 520–542.
- Haldane JBS. (1927). The comparative genetics of colour in rodents and carnivora. *Biol Rev Cambridge Phil Soc (London)* **2**: 199–212.
- Hann B, Balmain A. (2001). Building ‘validated’ mouse models of human cancer. *Curr Opin Cell Biol* **13**: 778–784.
- Helou K, Walentinsson A, Levan G, Stahl F. (2001). Between rat and mouse zoo-FISH reveals 49 chromosomal segments that have been conserved in evolution. *Mammal Genome* **12**: 765–771.
- Hillyard AL, Doolittle DP, Davisson MT, Roderick TH. (1991). Locus map of mouse. *Mouse Genome* **89**: 16–30.
- Ioannidu S, Walter L, Dressel R, Gunther E. (2001). Physical map and expression profile of genes of the telomeric class I gene region of the rat MHC. *J Immunol* **166**: 3957–3965.
- James MR, Lindpainter K. (1997). Why map the rat? *Trends Genet* **13**: 171–173.
- Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. (1999). Mouse ENU mutagenesis. *Hum Mol Genet* **8**: 1955–1963.
- Kwittek-Black AE, Jacob HJ. (2001). The use of designer rats in the genetic dissection of hypertension. *Curr Hyperten Rep* **3**: 12–18.

- Levan G, Szpirer J, Szpirer C, Klinga K, Hanson C, Islam MQ. (1991). The gene map of the Norway rat (*Rattus norvegicus*) and comparative mapping with mouse and man. *Genomics* **10**: 699–718.
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, *et al.* (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet* **24**: 381–386.
- Little CC, Tyzzer EE. (1916). Further studies on inheritance of susceptibility to a transplantable tumor of Japanese waltzing mice. *J Med Res* **33**: 393–398.
- Lyon MF, Searle AG. (Eds) (1989). *Genetic Variants and Strains of the Laboratory Mouse*, 2nd edn. Oxford University Press: Oxford.
- Lyon MF, Rastan S, Brown SDM. (Eds) (1996). *Genetic Variants and Strains of the Laboratory Mouse*, 3rd edn. Oxford University Press: New York.
- Manly KF. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. *Mammal Genome* **4**: 303–313.
- Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, *et al.* (1997). High throughput fingerprint analysis of large-insert clones. *Genome Res* **7**: 1072–1084.
- Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, *et al.* (2001). VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**: 1046–1047.
- McCarthy LC, Terrett J, Davis ME, Knights CJ, Smith AL, Critcher R, *et al.* (1997). A first-generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res* **7**: 1153–1161.
- MGD. (2002). Statistics for number of localized genes. [ftp://ftp.informatics.jax.org/pub/informatics/reports/MGD\\_Stats.sql.rpt](ftp://ftp.informatics.jax.org/pub/informatics/reports/MGD_Stats.sql.rpt) [1 January 2002].
- Moldin SO, Farmer ME, Chin HR, Battey JF Jr. (2001). Trans-NIH neuroscience initiatives on mouse phenotyping and mutagenesis. *Mammal Genome* **12**: 575–581.
- Mouse Genome Sequencing Consortium (2002). Initial Sequencing and Comparative analysis of the Mouse genome. *Nature*, in press.
- Nadeau JH, Grant P, Kosowsky M. (1991). Mouse on human homology map. *Mouse Genome* **89**: 31–36.
- Nilsson S, Helou K, Walentinsson A, Szpirer C, Nerman I, Stahl F. (2001). Rat–mouse and rat–human comparative maps based on gene homology and high-resolution zoo-FISH. *Genomics* **74**: 287–298.
- Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, *et al.* (2000). A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nature Genet* **25**: 440–443.
- Oeltjen JC, Malley TM, Muzny DM, Miller W, Gibbs RA, Belmont JW. (1997). Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res* **7**: 315–329.
- Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, *et al.* (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* **10**: 116–128.
- RatMap (2002). Statistics for number of localized genes. <http://ratmap.gen.gu.se> [1 January 2002].
- Richardson JE, Eppig JT, Nadeau JH. (1995). Building an integrated mouse genome database. *IEEE Eng Med Biol* **14**: 718–724.
- Riley M. (1993). Functions of the gene products of *Escherichia coli*. *Microbiol Rev* **57**: 862–952.

- Ringwald M, Eppig JT, Begley DA, Corradi JP, McCright IJ, Hayamizu TF, *et al.* (2001). The Mouse Gene Expression Database (GXD). *Nucleic Acids Res* **29**: 98–101.
- Rison SCG, Hodgman TC, Thornton JM. (2000). Comparison of functional annotation schemes for genomes. *Funct Integ Genomics* **1**: 56–69.
- Rodriguez-Tomé P, Lijnzaad P. (2001). RHdb: the Radiation Hybrid database. *Nucleic Acids Res* **29**: 165–166.
- Rowe LB, Nadeau JH, Turner R, Frankel WN, Letts VA, Eppig JT, *et al.* (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mammal Genome* **5**: 253–274.
- Schimenti J, Bucan M. (1998). Functional genomics in the mouse: phenotype based on mutagenesis screens. *Genome Res* **8**: 698–710.
- Schork NJ, Fallin D, Lanchbury JS. (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* **58**: 250–264.
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, *et al.* (2000). PipMaker: A web server for aligning two genomic DNA sequences. *Genome Res* **10**: 577–586.
- Snell GD. (1941). Genes and chromosome mutation. In *Biology of the Laboratory Mouse*, 1st edn, Snell GD. (Ed.). McGraw-Hill: New York, pp. 234–247.
- Soderlund C, Humphrey S, Dunhum A, French L. (2000). Contigs built with fingerprints, markers and FPC V4.7. *Genome Res* **10**: 1772–1787.
- Staats J. (1985). Standardized nomenclature for inbred strains of mice: eighth listing. *Cancer Res* **45**: 945–977.
- Steen RG, Kwitek-Black AE, Glenn C, Gullings-Handley J, Van Etten W, Atkinson OS, *et al.* (1999). A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res* **9**, AP1–8, insert.
- Strausberg RL, Feingold EA, Klausner RD, Collins FC. (2000). The mammalian gene collection. *Science* **286**: 455–457.
- Stroesser G, Baker W, van den Broek A, Camon E, Garcia-Pastor M, Kanz C, *et al.* (2001). The EMBL nucleotide sequence database. *Nucleic Acids Res* **29**: 17–21.
- Sugiyama F, Yagami K, Paigen B. (2001). Mouse models of blood pressure regulation and hypertension. *Curr Hyperten Rep* **3**: 41–48.
- Temple LKF, McLeod RS, Gallinger S, Wright JG. (2001). Defining disease in the genomics era. *Science* **293**: 807–808.
- The RIKEN Exploration Research Group Phase II Team and the FANTOM Consortium. (2001). Functional annotation of a full-length mouse cDNA collection. *Nature* **409**: 685–690.
- Twigger S, Lu J, Shimoyama M, Chen D, Pasko D, Long H, *et al.* (2002). Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res* **30**: 125–128.
- Van Etten WJ, Steen RG, Nguyen H, Castle AB, Slonim DK, Ge B, *et al.* (1999). Radiation hybrid map of the mouse genome. *Nature Genet* **22**: 384–387.
- Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence CE. (2000). Human–mouse genome comparisons to locate regulatory sites. *Nature Genet* **26**: 225–228.
- Watanabe TK, Bihoreau MT, McCarthy LC, Kiguwa SL, Hishigaki H, Tsuji A, *et al.* (1999). A radiation hybrid map of the rat genome containing 5,255 markers. *Nature Genet* **22**: 27–36.
- Watanabe TK, Ono T, Okuno S, Mizoguchi-Miyakita A, Yamasaki Y, Kanemoto N, *et al.* (2000). Characterization of newly developed SSLP markers for the rat. *Mammal Genome* **11**: 300–305.

- Wei S, Wei K, Moralejo DH, Yamada T, Izumi K, Matsumoto K. (1998). An integrated genetic map of the rat with 562 markers from different sources. *Mammal Genome* **9**: 1002–1007.
- Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, *et al.* (2001). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **29**: 11–16.
- Zhao S, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Krol M, *et al.* (2001). Mouse BAC ends quality assessment and sequence analyses. *Genome Res* **11**: 1736–1745.
- Zheng BJ, Mills AA, Bradley A. (1999). A system for rapid generation of coat color-tagged knockouts and defined chromosomal rearrangements in mice. *Nucleic Acids Res* **27**: 2354–2360.