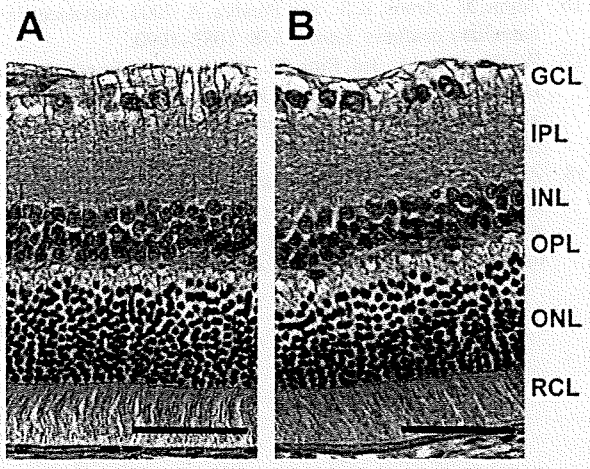

Fig. 5. Retinal histology in *kci/kci* rats. Vertical retinal cross sections from a heterozygous (A) and a homozygous (B) KCI mutant rat at 16 weeks of age. Each layer is labeled: ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), and the layer of rods and cones (LRC). No anatomic defects were noted in the homozygous rat. Bar=50 μm.

KCI rat showed abnormal morphologies or organizations of hair cells in the inner ear. Considering the importance of rats in the development of clinical treatments and diagnoses, and new drugs, the *Pcdh15*-deficit KCI rat has potential to serve as a good animal model for UDH1F and DFNB23.

Acknowledgment(s)

We are grateful to the National Bio Resource Project for the Rat in Japan for providing the KCI rat strain (NBPR#0440). This work was supported in part by Grants-in-aid for Scientific Research from the Japan Society for the Promotion of Science (18300141 to TK and 16200029 to TS) and a Grant-in-aid for Cancer Research from the Ministry of Health, Labour and Welfare. We are grateful to Ms. M. Yokoe for her excellent technical assistance.

References

1. Ahituv, N. and Avraham, K.B. 2002. Mouse models for human deafness: current tools for new fashions. *Trends Mol. Med.* 8: 447–451.
2. Ahmed, Z.M., Riazuddin, S., Bernstein, S.L., Ahmed, Z., Khan, S., Griffith, A.J., Morell, R.J., Friedman, T.B., Riazuddin, S., and Wilcox, E.R. 2001. Mutations of the protocadherin gene PCDH15 cause Usher syndrome type 1F. *Am. J. Hum. Genet.* 69: 25–34.
3. Ahmed, Z.M., Riazuddin, S., Ahmad, J., Bernstein, S.L., Guo, Y., Sabar, M.F., Sieving, P., Riazuddin, S., Griffith, A.J., Friedman, T.B., Belyantseva, I.A., and Wilcox, E.R. 2003. PCDH15 is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23. *Hum. Mol. Genet.* 12: 3215–3223.
4. Alagramam, K.N., Zahorsky-Reeves, J., Wright, C.G., Pawlowski, K.S., Erway, L.C., Stubbs, L., and Woychik, R.P. 2000. Neuroepithelial defects of the inner ear in a new allele of the mouse mutation Ames waltzer. *Hear. Res.* 148: 181–191.
5. Alagramam, K.N., Murcia, C.L., Kwon, H.Y., Pawlowski, K.S., Wright, C.G., and Woychik, R.P. 2001. The mouse Ames waltzer hearing-loss mutant is caused by mutation of Pcdh15, a novel protocadherin gene. *Nat. Genet.* 27: 99–102.
6. Alagramam, K.N., Yuan, H., Kuehn, M.H., Murcia, C.L., Wayne, S., Srisailpathy, C.R., Lowry, R.B., Knaus, R., Van Laer, L., Bernier, F. P., Schwartz, S., Lee, C., Morton, C.C., Mullins, R.F., Ramesh, A., Van Camp, G., Hageman, G.S., Woychik, R.P., and Smith, R.J. 2001. Mutations in the novel protocadherin PCDH15 cause Usher syndrome type 1F. *Hum. Mol. Genet.* 10: 1709–1718.
7. Alagramam, K.N., Stahl, J.S., Jones, S.M., Pawlowski, K.S., and Wright, C.G. 2005. Characterization of vestibular dysfunction in the mouse model for Usher syndrome 1F. *J. Assoc. Res. Otolaryngol.* 6: 106–118.
8. Hampton, L.L., Wright, C.G., Alagramam, K.N., Battey, J.F., and Noben-Trauth, K. 2003. A new spontaneous mutation in the mouse Ames waltzer gene, Pcdh15. *Hear. Res.* 180: 67–75.
9. Haywood-Watson, R.J. 2nd, Ahmed, Z.M., Kjellstrom, S., Bush, R.A., Takada, Y., Hampton, L.L., Battey, J.F., Sieving, P.A., and Friedman, T.B. 2006. Ames waltzer deaf mice have reduced electroretinogram amplitudes and complex alternative splicing of pcdh15 transcripts. *Invest. Ophthalmol. Vis. Sci.* 47: 3074–3084.
10. Kazmierczak, P., Sakaguchi, H., Tokita, J., Wilson-Kubalek, E.M., Milligan, R.A., Muller, U., and Kachar, B. 2007. Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. *Nature* 449: 87–91.
11. Keats, B.J. and Corey, D.P. 1999. The usher syndromes. *Am. J. Med. Genet.* 89: 158–166.
12. Pawlowski, K.S., Kikkawa, Y.S., Wright, C.G., and Alagramam, K.N. 2006. Progression of inner ear pathology in Ames waltzer mice and the role of protocadherin 15 in hair cell development. *J. Assoc. Res. Otolaryngol.* 7: 83–94.
13. Rebibo-Sabbah, A., Nudelman, I., Ahmed, Z.M., Baasov, T., and Ben-Yosef, T. 2007. In vitro and ex vivo suppression by aminoglycosides of PCDH15 nonsense mutations underlying type 1 Usher syndrome. *Hum. Genet.* 122: 373–381.

14. Senften, M., Schwander, M., Kazmierczak, P., Lillo, C., Shin, J.B., Hasson, T., Geleoc, G.S., Gillespie, P.G., Williams, D., Holt, J.R., and Muller, U. 2006. Physical and functional interaction between protocadherin 15 and myosin VIIa in mechanosensory hair cells. *J. Neurosci.* 26: 2060–2071.
15. Taylor, B.A., Navin, A., and Phillips, S.J. 1994. PCR-amplification of simple sequence repeat variants from pooled DNA samples for rapidly mapping new mutations of the mouse. *Genomics* 21: 626–632.
16. Tekin, M., Arnos, K.S., and Pandya, A. 2001. Advances in hereditary deafness. *Lancet* 358: 1082–1090.
17. Washington, J.L. 3rd, Pitts, D., Wright, C.G., Erway, L.C., Davis, R.R., and Alagramam, K. 2005. Characterization of a new allele of Ames waltzer generated by ENU mutagenesis. *Hear. Res.* 202: 161–169.
18. Zheng, Q.Y., Yu, H., Washington, J.L. 3rd, Kisley, L.B., Kikkawa, Y.S., Pawlowski, K.S., Wright, C.G., and Alagramam, K.N. 2006. A new spontaneous mutation in the mouse protocadherin 15 gene. *Hear. Res.* 219: 110–120.

Physiol Genomics 33: 205–211, 2008.
First published January 22, 2008; doi:10.1152/physiolgenomics.00222.2007.

Functional polymorphisms in inbred rat strains and their allele frequencies in commercially available outbred stocks

Takashi Kuramoto, Satoshi Nakanishi, and Tadao Serikawa

Institute of Laboratory Animals, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto, Japan

Submitted 23 September 2007; accepted in final form 16 January 2008

Kuramoto T, Nakanishi S, Serikawa T. Functional polymorphisms in inbred rat strains and their allele frequencies in commercially available outbred stocks. *Physiol Genomics* 33: 205–211, 2008. First published January 22, 2008; doi:10.1152/physiolgenomics.00222.2007.—Polymorphisms that have been proven to influence gene functions are called functional polymorphisms. It is significant to know the distribution of functional polymorphisms in the rat, widely used in animal models for human diseases. In this study, we assessed 16 functional polymorphisms consisting of 3 coat color and 13 disease-associated genes in 136 rat strains, as a part of the genetic profiling program of the National Bio Resource Project for the Rat (NBRP-Rat). Polymorphisms of *Cdkn1a*, *Fcgr3*, *Grp10*, *Lss*, and *Fdft1*, which were proven to function in prostate tumorigenesis, glomerulonephritis, hyperphagia, and cholesterol biosynthesis, were shared among various inbred strains. These findings indicated that most rat strains harbored the disease-associated alleles and suggested that many unidentified functional polymorphisms might exist in inbred rat strains. The functional polymorphisms shared in inbred strains were also observed within outbred stocks available commercially. Therefore, this implies that experimental plans based on either rat inbred strains or outbred stocks need to be carefully designed with a full understanding of the genetic characteristics of the animals. To select the most suitable strains for experiments, the NBRP-Rat will periodically improve and update the genetic profiles of rat strains.

rat resource; National Bio Resource Project; disease model; single nucleotide polymorphism

THE LABORATORY RAT (*Rattus norvegicus*) is an indispensable tool in experimental medicine and drug development. It is used extensively as a model organism for studying normal and disease processes in humans because of our extensive knowledge of rat physiology and the large number of rat models that mimic human diseases. In drug development, rats are central to determination of drug efficacy and toxicity before human clinical trials.

Recently, a large set of genomic tools and information on the rat genome such as the rat genome draft sequences (6), the large-scale single nucleotide polymorphism (SNP) haplotype map (STAR project), and the Phenome Project (11, 13) have been developed, allowing us to identify genes responsible for monogenic as well as polygenic traits in rat models. Because these mutations have been proven to function in the pathogenesis of diseases, they are also called functional polymorphisms. Recently, the assessment of functional polymorphisms by PCR has demonstrated some not to be unique to the strain in which they were initially identified but to be shared by several rat strains (9, 16, 22). Furthermore, some functional polymorphisms were also found in outbred colonies at various allele frequencies (9, 22).

The National Bio Resource Project for the Rat (NBRP-Rat) is one of the largest resource centers for the rat (18). More than 200 inbred strains have been deposited so far in the NBRP-Rat and are maintained as live animals or cryopreserved embryos. The NBRP-Rat offers comprehensive data sets on the strains deposited that are useful for the design of experiments for genetic analysis (14). With the growing number of genes identified as causative in rat disease models, the genetic profiles of the deposited rat strains for these functional polymorphisms need to be updated.

In this study, to further improve the genetic profiles of the strains deposited in the NBRP-Rat, we examined rat strains for known functional polymorphisms such as disease-associated and coat color alleles. Because coat color mutations are involved in organelle trafficking, signal transduction, and differentiation (20), it is also important to know how prevalent coat color mutations are in the established inbred strains. We also investigated how prevalent these polymorphisms are among commercial colonies of outbred rats, and we discuss controversial points in the use of outbred stocks.

MATERIALS AND METHODS

Animals. A total of 136 inbred rat strains were examined. Genomic DNA samples of 114 strains were obtained from the NBRP-Rat in Japan and those of 22 strains from commercial breeders or holders of the strains.

Eleven outbred colonies were examined. Outbred rats from each colony were purchased in bulk from the following commercial breeders. Crj:CD(SD) (n = 31), Crj:CD(SD)IGS (n = 31), Crj:WI(Glx/BRL/Han)IGS (n = 31), and Crj:Wistar (n = 31) outbred rats were from Charles River Laboratories Japan (Yokohama, Japan). BrlHan:WIST (n = 32), Jcl:SD (n = 32), and Jcl:Wistar (n = 32) outbred rats were from CLEA Japan (Kawasaki, Japan). Slc:SD (n = 32) and Slc:Wistar (n = 32) outbred rats were from Japan SLC (Hamamatsu, Japan). Iar:Long-Evans (n = 30) and Kwl:Long-Evans (n = 30) were from the Institute for Animal Reproduction (Kasumigaura, Japan) and Kiwa Laboratory Animals (Wakayama, Japan). Genomic DNA was prepared from tail biopsy specimens.

Animal care and experimental procedures were approved by the Animal Research Committee of Kyoto University and were conducted according to the Regulation on Animal Experimentation at Kyoto University.

Genotyping. Sixteen functional polymorphisms were picked up by searching the literature (Table 1). They comprised 3 coat color genes and 13 disease-associated genes. PCR primer sequences and methods for the detection of functional polymorphisms are listed in Table 2. Genotyping of *Fdft1* or *Ian5* was performed by the direct sequencing of the PCR products. PCR products were treated with ExoSAP-IT (Amersham Biosciences) to digest single-stranded DNA and excess primers. Cycle sequencing was performed with BigDye Terminator Ready Reaction Mix version 3.1 according to the manufacturer's instructions (Applied Biosystems). PCR samples were purified with

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: T. Kuramoto, Institute of Laboratory Animals, Graduate School of Medicine, Kyoto Univ., Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan (e-mail: tkuramot@anim.med.kyoto-u.ac.jp).

Table 1. *Functional polymorphisms examined in this study*

Phenotype	Causative Gene	Type of Mutation	Representative Strains	Origin	Reference
Nonagouti (a)	*A*	19-bp Deletion in exon 2	BN, F344		10
Analbuminemia	*Alb*	7-bp Deletion at splicing donor site in intron H	NAR	Jcl:SD	4
Insulin-dependent diabetes mellitus KDP QTL 1 (Iddm/kdp1)	*Cblb*	Nonsense mutation at codon 455 (R455X)	KDP	Long-Evans	26
Cholecystokinin A receptor	*Cckar1*	6.8-kb Deletion spanning the promoter region, exon 1, and exon 2	OLETF	Long-Evans	21
QTL for insulin resistance, defective fatty acid metabolism and hypertriglyceridemia in SHR	*Cd36*	Deletion spanning the region between exon 4 of *Cd36* and exon 5 of *Cd36-ps2*	SHR/Crl	Kyo:Wistar	7,8
Prostate cancer susceptibility 2 (Pcs2)	*Cdkn1a*	119-bp Insertion in the promoter region	BUF		25
UPL rat cataract (Uca)	*Cox50*	Missense mutation at codon 340 (R340W)	UPL	Crj:SD	23
Crescentic glomerulonephritis 1 (Crgn1)	*Fcgr3*	Gene duplication	LEW, BN, IS		1
Cataract SCR 2 (Cats2)	*Fdft1*	Missense mutation at codon 196 (I196K)	SCR	SHR × fatty	15
Diabetes mellitus OLETF QTL 1 (Dmo1)	*Gpr10*	Point mutation at translation initiation codon	OLETF	Long-Evans	22
Lymphopenia (Lyp)	*Ian5/Gimap5*	1-bp Deletion resulting in frameshift	BBDP	Wistar Furth	12
Cataract SCR 1 lethal (Cats1^l)	*Lss*	12-bp Deletion in exon 15	SCR	SHR × fatty	15
Cataract SCR 1 (Cats1), Q481R	*Lss*	Missense mutation at codon 481 (Q481R)	SCR	SHR × fatty	15
Cataract SCR 1 (Cats1), D139N	*Lss*	Missense mutation at codon 139 (D139N)	SCR	SHR × fatty	15
Pink-eyed dilution (p)	*P/Oca2*	Deletion spanning exon 17 and exon 18	RCS, KHR		9
Ruby (r)	*Rab38*	Point mutation at translation initiation codon	TM, LE/Stm, LEC, FH	Long-Evans	17

Origin, possible rat colony where the mutation arose.

CleanSEQ (Agencourt Bioscience) and then loaded into an ABI PRISM 3100 genetic analyzer (Applied Biosystems).

RESULTS

Disease-related functional polymorphisms shared among rat strains. Six disease-related functional polymorphisms were common among rat strains. The distribution of alleles of each functional polymorphism is listed in Table 3.

When a common origin of the polymorphism is postulated, it would be expected that all strains that carry the polymorphism would cluster in the same branch of the rat phylogenic tree. To test this hypothesis, we analyzed the distribution of each disease-related allele with the distribution of the strains on the phylogenic tree made by the NBRP-Rat with 132 inbred rat strains (**http://www.anim.med.kyoto-u.ac.jp/nbr/phylogenetictree.htm**). The SCR-type (K) allele of *Fdft1* only clustered in a branch that contained Wistar-Kyoto (WKY) and spontaneously hypertensive rat (SHR) substrains (Supplemental Fig. S1).[1] The remaining alleles distributed throughout the phylogenic tree (data not shown).

Functional polymorphisms unique to strain in which mutation has been identified. Seven mutant alleles that were unique to the original strain included the 7-bp deletion at the splicing donor site in the *Alb* gene in the NAR strain, the nonsense mutation at codon 455 (R455X) of the *Cblb* gene in the KDP rat, the 6.8-kb deletion spanning the promoter region, exon 1, and exon 2 of the *Cckar1* gene in the OLETF rat, the deletion spanning the region *Cd36* and *Cd36-ps2* in the SHR/Crl strain, the missense mutation at codon 340 (R340W) of the *Cox50* gene in the UPL rat, the frameshift mutation of the *Ian5* gene in the BB rat, and the 12-bp deletion in exon 15 of the *Lss* gene in the SCR strain (2, 4, 12, 15, 21, 23, 26) (Table 4). These mutations might have significant effects so that the mutation itself worked as a selective marker in establishment of the strain and exaggerates the clinical phenotype of these strains.

[1] The online version of this article contains supplemental material.

Coat color mutations. P and *Rab38* are causative genes for the Pink-eyed dilution (*p*) and Ruby (*r*) in the rat (9, 17). Genotyping of *P* and *Rab38* could not detect any strains other than those described to carry these mutations (9, 17). The functional polymorphism at the *A* locus (agouti or nonagouti) was evaluated in the albino rat strains. Among 91 albino rat strains, 11 had the *A/A* genotype, 79 had the *a/a* genotype, and 1 was heterozygous (Table 4).

Survey of functional polymorphisms commonly shared in inbred strains in commercially available outbred stocks. Because most inbred rat strains are descended from outbred colonies, we speculated that functional polymorphisms common among inbred strains could be observed in various outbred rat colonies. We then surveyed 6 disease-related polymorphisms and 3 coat color mutations found in multiple strains in a total of 344 outbred rats from 11 stocks available commercially.

Five functional polymorphisms and two coat color mutations showed different allele frequencies from colony to colony, but any prevalence of the *Fdft1* mutation or the ruby mutation could not be found (Table 5). In the colonies of Slc:Wistar and Kwl:Long-Evans, all functional polymorphisms examined seemed to be genetically fixed to homozygosity for single alleles, even if the sample size was only 32 or 30.

DISCUSSION

Our study revealed that most rat strains harbored the mutant alleles that were already identified and whose effects were relatively large (Table 3). These mutant alleles distributed throughout the strains and were found in outbred rat stocks, except for the *Fdft1* missense mutation. These findings suggested that the mutant alleles examined in this study were already present in the ancestral outbred rats of the modern inbred strains and then fixed during establishment of the strains. The *Fdft1* missense allele, which was clustered on a particular branch of the phylogenic tree but not found in the outbred rats, might occur in the Wistar-related strains kept in Japan and be inherited by the SHR strains. It is likely that

Table 2. *PCR primers for genotyping*

Gene Symbol	Primer Name	Forward Primer (5& → 3′)	Reverse Primer (5′ → 3′)	Methods
A	agouti-M&N	CACAGCCACCTGGTATTTGA	CTCCCTCTCATGGGTCAGAG	Length polymorphism
Alb	rAlb_int9del-F&R	CTGGAGAAGTGCTGTGCTGA	AACATCAGGATTGACCTACGTATC	Length polymorphism
Cblb	Cblb-RFLP-F&R	TGCCCCTTCTGTCGCTGTGA	CCTCGGTTTTGAATCAACAG	RFLP (*Taq*I)
Cckar1	rCckar_DEL-F&R	AACACTCACCATGGCAAGG	GTCACTTGGCAACAGGAAGC	Presence or absence
	rCckar_WT-F&R	AGGAGAGAGACAGGAATGAGC	AAGGTTGAGGTTGATCCAAGC	Presence or absence
Cd36	rCd36_Del-F&R	CTGAAACTATCATATCTCGAT	TTTGGAGATTGGTTTTACCAA	Presence or absence
Cdkn1a	rCdkn1a_5′-F&R	GGGCTGTTATTAGACATGA	AGAGCCACGCACATCTATG	Length polymorphism
Cox50	rCx50-4F&R	GCCAAGCCTTTTAGTCAG	TCACTAGGACAGTGGGTTTA	RFLP (*Aci*I)
Fcgr3	rFcgr_exon5-F&R	GTCCCTAAATTCTGAATTTC	AAAGAAGTCACAGAAAGGAG	Length polymorphism
Fdft1	rFdft1_SNP-F&R	AGGCTGTTTCTGTGTCGTGA	TGAGGCCAAAACTGTCTTCC	Sequencing
Gpr10	rGrp10_SNP-1&2	GGGCTTGTAGAAAGATAGAAAGTG	GTGACTGCTGCAGCTCTGG	RFLP (*Nla*III)
Ian5/Gimap5	rIan5_ex3del-F&R	AAAAGTGCCACAGGGAACAG	GGTCACCAGCAACAACACAT	Sequencing
Lss	rLss_SNP-F&R	ACACTGGACTGTGGCTGGAT	TTAGTCCGCTAAGGCACAGC	Length polymorphism, RFLP(*Ava*I)
Lss	rLss_D139-F&R	AGAGAAGGGCTTGGGTTAGC	CAGGCATGTGTGTGTGTTGT	RFLP (*Bcc*I)
P/Oca2	rP-157&92	AAGGGGTGTTGAATTTTGTCA	AGGAGAGCAGGACTGGATCA	Presence or absence
	rP-91&92	AAAGAGTAAAGCATTGACATTCTTCTC	AGGAGAGCAGGACTGGATCA	Presence or absence
Rab38	Rab38Ex1-F&R	AAGCTCCAGGCTCCGCAAGA	CCTCTCACTGTCCCATTCCAC	RFLP (*Fok*I)

functional polymorphisms can be divided into two categories: those that existed before the establishment of many rat strains and those that occurred after many strains were produced. The later polymorphism might help us to speculate on the genotype of the strains that are in the same branch of the phylogenic tree.

The prevalence of the disease-related polymorphisms among different inbred strains raises concern about using rat strains as models for diseases such as prostate cancer, experimental glomerulonephritis, and obesity. Cyclin-dependent kinase inhibitor 1a (*Cdkn1a*), also called *p21*, is a tumor suppressor gene and considered a good candidate for one of the prostate cancer susceptibility genes. The 119-bp insertion was associated with high *Cdkn1a* expression in the rat prostate, which might imply resistance to prostate cancer in the BUF/NacJcl rat (24, 25). The insertion was shared in 49 strains including the prostate cancer-resistant BUF/NacJcl, but not in 92 strains including the prostate cancer-susceptible ACI/NJcl. Interestingly, BUF/Mna, a substrain of the BUF rat, did not share the insertion. These findings suggested that both insertion-positive and -negative strains should be employed in carcinogenicity tests of a chemical for the prostate.

Fcgr3-related sequence (*Fcgr3-rs*) has been identified as a determinant of macrophage overactivity and glomerulonephritis in WKY. Rats carrying the *Fcgr3-rs* gene in addition to *Fcgr3* showed resistance to macrophage-dependent mesangial cell damage and glomerular necrosis in experimentally induced nephrotoxic nephritis (1). *Fcgr3-rs* was shared in 79 strains including the glomerulonephritis-resistant LEW strains, but not in 56 strains including the glomerulonephritis-susceptible WKY strains. In humans, the copy numbers of *FCGR3B*, a paralog of *Fcgr3*, varied from zero to six and were associated with susceptibility to systemic autoimmunity (5). In this study, we could assess the presence or absence of *Fcgr3-rs*, but not the actual copy number in the rat genome. It would be worth surveying the copy numbers of *Fcgr3-rs* in rat strains and evaluating the association of copy number with macrophage overactivity and glomerulonephritis.

Gpr10 encodes the receptor for prolactin-releasing peptide (PrRP) and has been identified as a determinant of hyperphagia leading to obesity and dyslipidemia in the OLETF rat (22). The translation initiation codon mutation (ATG-to-ATA) was shared in several strains: 45 strains had ATA/ATA, and 95 had ATG/ATG. Among the 136 strains genotyped in this study, 85 were phenotyped for body weight in males at 10 wk of age in the Rat Phenome Project (13). The average body weights of the *Grp10*-mutant strains ($n = 31$) were significantly higher than those of the wild-type strains ($n = 54$): 282.2 ± 50.1 vs. 247.6 ± 45.0 g (means ± SD, $P < 0.001$). Furthermore, it has been described that body weights of outbred rats carrying the *Gpr10* mutation were significantly greater than those of wild-type rats (22). These findings suggested a significant role for the *Grp10* gene in the control of body weight in the rat.

Genotyping of the functional polymorphism could also nominate a potential candidate gene for the existing mutant strain.

Table 3. *Disease-related functional polymorphisms shared among rat stains*

Functional Polymorphism	Genotype	Phenotype	No. of Strains (%)	Representative Strain
119-bp Insertion in the promoter region of *Cdkn1a*	Negative	Low expression of *Cdkn1a*	86 (63)	ACI/NJcl
	Positive	High expression of *Cdkn1a*	49 (36)	BUF/NacJcl
Gene duplication of *Fcgr3* (*Fcgr3-rs*)	Negative	High macrophage activity	53 (39)	WKY
	Positive	Low macrophage activity	79 (58)	LEW
Missense mutation at codon 196 of *Fdft1* (I196K)	ATA (Ile)		107 (79)	ACI
	AAA (Lys)	Low FDFT1 activity	29 (21)	SCR
Point mutation at translation initiation codon of *Gpr10*	ATG (Met)		89 (65)	BN
	ATA (Ile)	Hyperphagia	45 (33)	OLETF
Missense mutation at codon 139 of *Lss* (D139N)	GAT (Asp)		72 (53)	ACI
	AAT (Asn)	Low LSS activity	62 (46)	SCR
Missense mutation at codon 481 of *Lss* (Q481R)	CAA (Gln)		62 (46)	ACI
	CGA (Arg)	Low LSS activity	72 (53)	SCR

Table 4. *Genotypes of functional polymorphisms in 137 rat strains*

Strain Name	NBRP No.	Source of Genomic DNA	*A*	*Cdkn1a*	*Fcgr3-rs*	*Fdft1*	*Grp10*	*Lss D139N*	*Lss Q481R*	*P*	*Rab38*
ACI/NHok	0033	NBRP-Rat	+	A	L	A	B	A	A	+	+
ACI/NJcl	NA	CLEA Japan	+	A	L	A	B	A	A	+	+
ACI/NKyo	0001	NBRP-Rat	+	A	L	A	B	A	A	+	+
ACI/NKyo-*Lyst*$^{bg\text{-}Kyo}$	0002	NBRP-Rat	+	A	L	A	B	A	A	+	+
ACI/NMna	0203	NBRP-Rat	+	A	L	A	B	A	A	+	+
ACI/NSlc	0158	NBRP-Rat	+	A	L	A	B	A	A	+	+
ACIS	0034	NBRP-Rat	+	A	L	A	B	A	A	+	+
AI	0145	NBRP-Rat	a	A	L	A	B	A	A	+	+
ALB	0035	NBRP-Rat	a	B	L	A	B	S	S	+	r
AMI	NA	NBRP-Rat	a	A	W	S	B	S	S	+	+
BB/WorTky	0030	NBRP-Rat	a	A	L	A	B	A	A	+	+
BN/I	0037	NBRP-Rat	a	A	L	A	B	A	A	+	+
BN/2	0038	NBRP-Rat	a	A	L	A	B	A	A	+	+
BN/fKtsSlc	0165	Japan SLC	a	A	L	A	B	A	A	+	+
BN/fMailHok	0036	NBRP-Rat	a	A	L	A	B	A	A	+	+
BN/Katholiek/Kts	0164	NBRP-Rat	a	A	L	A	B	A	A	+	+
BN/Seac	NA	KYUDO	a	A	L	A	B	A	A	+	+
BN/SsNHsd	NA	Harlan Sprague Dawley	a	A	L	A	B	A	A	+	+
BN/SsNSlc	0149	NBRP-Rat	a	A	L	A	B	A	A	+	+
BUF/Mna	0200	NBRP-Rat	a	A	L	A	B	S	S	+	+
BUF/NacJcl	NA	CLEA Japan	a	B	L	A	O	S	S	+	+
DA/Slc	0157	NBRP-Rat	+	A	L	A	B	A	A	+	+
DahlS/Seac	NA	KYUDO	a	B	L	A	het	A	A	+	+
DMY/Kyo	0003	NBRP-Rat	a	B	W	S	O	A	A	+	+
DON/Kyo	0004	NBRP-Rat	a	A	W	A	O	S	S	+	+
DOP	0093	NBRP-Rat	a	het	L	A	B	A	A	+	+
DRH/Seac	0078	NBRP-Rat	a	B	L	A	O	A	A	+	+
EXHC/Seac	0082	NBRP-Rat	a	B	W	A	B	A	A	+	+
F344/DuCrlCrlj	NA	CLEA Japan	a	B	L	A	B	A	A	+	+
F344/NSlc	0156	NBRP-Rat	a	B	L	A	B	A	A	+	+
F344/N-*Whn*$^{rnu/Kyo}$	0083	NBRP-Rat	a	B	L	A	B	A	A	+	+
F344/Snk	NA	Sankyo	a	B	L	A	B	A	A	+	+
F344/Stm	0140	NBRP-Rat	a	B	L	A	B	A	A	+	+
FH/HamSlc	0161	NBRP-Rat	a	B	L	A	B	A	A	+	+
GAERS	0285	NBRP-Rat	a	B	L	A	O	A	A	+	r
GK/Slc	0147	NBRP-Rat	a	A	L	A	B	A	A	+	+
GRY/Kyo	0368	NBRP-Rat	a	A	W	S	O	A	S	+	+
Gunn/Slc-*j/j*	0148	NBRP-Rat	a	A	L	A	O	S	S	+	+
HAA	0242	NBRP-Rat	a	B	L	A	O	S	S	+	+
HOB	0085	NBRP-Rat	a	A	L	A	B	A	A	+	+
HTX/Kyo	0006	NBRP-Rat	a	B	L	A	B	A	A	+	+
HWY/Slc	0152	NBRP-Rat	a	A	L	A	B	S	S	+	+
ICR	0290	NBRP-Rat	+	A	L	A	O	A	A	+	+
IER	0291	NBRP-Rat	a	B	L	S	B	S	S	+	+
IEW	0292	NBRP-Rat	a	B	L	S	B	S	S	+	+
IS/Kyo	0008	NBRP-Rat	a	B	L	A	B	S	S	+	+
IS/Kyo-*Tlk*	0009	NBRP-Rat	+	A	L	A	O	S	S	+	+
KB/Oda	0306	NBRP-Rat	+	A	L	A	O	S	S	+	+
KDP	0026	NBRP-Rat	+	A	L	A	O	S	S	+	+
KHR/Kyo	0007	NBRP-Rat	a	A	L	A	B	S	S	+	+
KMI	0029	NBRP-Rat	a	A	W	A	B	A	A	p	+
KND	0027	NBRP-Rat	a	B	W	A	B	A	A	p	+
KYN	0040	NBRP-Rat	a	A	L	A	B	S	S	+	+
KZC	0028	NBRP-Rat	a	B	W	A	B	S	S	+	+
KZ-*Lepr*^{fa}	0032	NBRP-Rat	a	A	L	A	O	S	S	+	+
LAA	0243	NBRP-Rat	+	A	L	A	O	A	A	+	+
LE/Stm	0139	NBRP-Rat	a	B	nd	A	O	S	S	+	+
LEA/Hkm	0244	NBRP-Rat	a	B	W	A	B	A	A	+	r
LEA/Hok	0041	NBRP-Rat	+	A	W	A	O	A	A	+	+
LEA/Tj	0050	NBRP-Rat	+	A	W	A	O	A	A	+	+
LEC/Crj	NA	Charles River Laboratories Japan	+	A	nd	A	O	A	A	+	+
LEC/Hok	0042	NBRP-Rat	+	A	W	A	B	A	A	+	r
LEJ/Hkm	0245	NBRP-Rat	+	A	W	A	B	A	A	+	r
LEJ/Hok	0043	NBRP-Rat	a	B	L	A	O	A	A	+	+
LETO	NA	Otsuka Pharmaceutical	a	B	L	A	O	A	A	+	+
LEW/Jms	0103	NBRP-Rat	a	A	L	A	B	S	S	+	+
LEW/Seac	NA	KYUDO	a	A	L	A	O	A	A	+	+
LEW/SsNSlc	0150	NBRP-Rat	a	B	L	A	O	A	A	+	+

Continued

Table 4.—*Continued*

Strain Name	NBRP No.	Source of Genomic DNA	*A*	*Cdkn1a*	*Fcgr3-rs*	*Fdft1*	*Grp10*	*Lss D139N*	*Lss Q481R*	*P*	*Rab38*
MD	0143	NBRP-Rat	het	A	het	A	het	het	het	+	+
MES	0160	NBRP-Rat	a	A	L	A	B	A	A	+	+
M-SHRSP/Ta	NA	Kyoto University	a	A	W	S	B	S	S	+	+
NAR/Slc	0153	NBRP-Rat	a	B	L	A	O	S	S	+	+
NE	0286	NBRP-Rat	a	B	L	A	B	A	A	+	+
NER/Kyo	0010	NBRP-Rat	+	B	L	A	O	A	A	+	+
NIG-III	0044	NBRP-Rat	+	A	L	A	B	A	A	+	+
ODS/ShiJcl	NA	CLEA Japan	a	B	L	S	B	A	S	+	+
ODS/ShiJcl-*od/od*	NA	CLEA Japan	a	B	L	S	B	A	S	+	+
OLETF	NA	Otsuka Pharmaceutical	a	A	L	A	O	S	S	+	+
OM/NSlc	0162	NBRP-Rat	a	A	L	A	B	A	A	+	+
Opactas Rat	0101	NBRP-Rat	a	A	L	A	O	A	A	+	+
PVG/Seac	0080	NBRP-Rat	a	B	W	A	B	A	A	+	+
RCS/Kyo	0011	NBRP-Rat	a	A	L	A	B	A	A	p	+
RC-Tg	0166	NBRP-Rat	a	B	L	A	B	A	A	+	+
RICO/Ngs	0220	NBRP-Rat	a	A	W	A	B	S	S	+	+
SCR	NA	Kiwa Laboratory Animals	a	A	W	S	B	S	S	+	+
SDJ	0045	NBRP-Rat	a	A	W	A	B	S	S	+	+
SDT/Jcl	NA	CLEA Japan	a	A	L	A	O	A	A	+	+
SER/Kyo	0012	NBRP-Rat	+	B	W	S	O	S	S	+	+
SHR/Hcm	0257	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHR/Izm	0193	NBRP-Rat	a	A	W	A	B	S	S	+	+
SHR/Kyo	0013	NBRP-Rat	a	A	W	A	B	S	S	+	+
SHR/Kyushu	NA	Kyoto University	a	A	W	A	B	S	S	+	+
SHR/NCrj	NA	Charles River Laboratories Japan	a	A	W	A	B	S	S	+	+
SHR/Nig	0142	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHR/Ta	0236	NBRP-Rat	a	A	W	A	B	S	S	+	+
SHR2/Dmcr	0448	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHR3/Dmcr	0449	NBRP-Rat	a	A	W	A	B	S	S	+	+
SHR4/Dmcr	0450	NBRP-Rat	a	A	W	A	B	S	S	+	+
SHRSP/Ezo	0146	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSP/Hos	NA	Japan SLC	a	A	W	S	B	S	S	+	+
SHRSP/Izm	0194	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSP/Ngsk	0144	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSP/Ta	0237	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSPA1sb/Ta	0238	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSP3/Dmcr	0452	NBRP-Rat	a	A	W	S	B	S	S	+	+
SHRSR/Ta	0239	NBRP-Rat	a	A	W	A	B	S	S	+	+
THE	0270	NBRP-Rat	a	A	W	A	B	S	S	+	+
TLE	0271	NBRP-Rat	a	A	W	S	O	S	S	+	+
TM/Kyo	0014	NBRP-Rat	a	A	L	A	B	S	S	+	r
TO	0046	NBRP-Rat	+	B	W	A	B	S	S	+	+
TO/Hkm	0246	NBRP-Rat	+	B	W	A	B	S	S	+	+
TRM/Kyo	0015	NBRP-Rat	a	B	W	S	O	S	S	+	+
TRMR/Kyo	0016	NBRP-Rat	a	B	W	S	O	S	S	+	+
UPL	0217	NBRP-Rat	a	A	nd	A	B	A	A	+	nd
VF/Kyo	0022	NBRP-Rat	a	B	W	S	O	S	S	+	+
W/Kyo	0017	NBRP-Rat	+	A	L	A	B	A	S	+	+
WAG/RijKyo	0018	NBRP-Rat	a	A	L	A	B	A	A	+	+
WBN/KobSlc	0155	NBRP-Rat	a	B	L	A	B	A	A	+	+
WF/Kop	0259	NBRP-Rat	a	A	L	A	O	A	A	+	+
WIAR	0284	NBRP-Rat	a	A	L	A	O	A	S	+	+
WKA/Seac	0079	NBRP-Rat	a	B	W	A	B	S	S	+	+
WKAH/HkmSlc	0154	NBRP-Rat	+	A	L	A	B	A	S	+	+
WKAH/Hok	0047	NBRP-Rat	+	A	L	A	B	A	S	+	+
WKY/Ezo	0378	NBRP-Rat	+	B	W	A	B	A	S	+	+
WKY/Hcm	0258	NBRP-Rat	a	B	W	A	O	S	S	+	+
WKY/Izm	0195	NBRP-Rat	a	B	W	A	O	S	S	+	+
WKY/NCrj	NA	Charles River Laboratories Japan	a	B	W	S	O	A	S	+	+
WKY/NMna	0203	NBRP-Rat	a	B	W	S	O	A	S	+	+
WKY/Ta	0240	NBRP-Rat	a	B	W	S	O	S	S	+	+
WKYO/Kyo	0019	NBRP-Rat	a	B	W	S	O	S	S	+	+
WNA/Num	0092	NBRP-Rat	a	A	L	A	B	S	S	p	+
WS/Slc	NA	Japan SLC	a	A	L	A	B	het	het	+	+
WT	0219	NBRP-Rat	a	A	L	A	O	A	A	+	+

Continued

Table 4.—*Continued*

Strain Name	NBRP No.	Source of Genomic DNA	*A*	*Cdkn1a*	*Fcgr3-rs*	*Fdft1*	*Grp10*	*Lss D139N*	*Lss Q481R*	*P*	*Rab38*
WTC/Kyo	0020	NBRP-Rat	a	B	W	S	O	S	S	+	+
ZI/Kyo	0024	NBRP-Rat	+	A	L	A	O	S	S	+	+
ZIMY/Kyo	0025	NBRP-Rat	+	B	L	A	O	S	S	+	+
		Subtotal of strains carrying wild-type allele	+; 29	A; 86	W; 53	A; 107	B; 89	A; 72	A; 62	+; 132	+; 129
		Subtotal of strains carrying mutant allele	a; 106	B; 49	L; 79	S; 29	O; 45	S; 62	S; 72	p; 4	r; 6
		Subtotal of strains heterozygous	het; 1	het; 1	het; 1	het; 0	het; 2	het; 2	het; 2	het; 0	het; 0

NA, not assigned; nd, not determined owing to the low quality of the genome; het, heterozygous.

Lss (lanosterol synthase) and *Fdft1* (farnesyldiphosphate farnesyltransferase) function in the cholesterol biosynthesis pathway and are determinants for hereditary cataracts in the Syumiya Cataract Rat (SCR) (15). Both functional polymorphisms of *Lss*, D139K and Q481R, reduce LSS activity and comprise the major susceptibility allele for cataracts, $Cats1^S$. The functional polymorphism of *Fdft1*, I196K, reduces FDFT1 activity and is likely to be a strong candidate for the second major susceptibility allele for cataracts, $Cats2^S$. In this study, we demonstrated that the *Lss* mutations and the *Fdft1* mutation are shared by 23 inbred strains that include not only the SCR but also the Ihara Cataract Rat (ICR) strain. The ICR strain displays bilateral cataracts by 4 mo of age, and the two causative loci were identified as $Cati^1$ on chromosome 8 and $Cati^2$ on chromosome 15 (27). Interestingly, $Cati^2$ modified the timing of the onset and was mapped between *D15Rat52* (32.8 Mb in the Celera Rat Map) and *D15Rat20* (52.4 Mb), where *Fdft1* was also mapped (37.1 Mb). Thus it is reasonable to consider *Fdft1* as a potential candidate for $Cati^2$.

We showed that disease-related alleles have penetrated different outbred stocks available commercially (Table 5). Because these alleles have been proven to function in the pathogenesis of prostate cancer, nephritis, obesity, or cholesterol biosynthesis, one should pay attention to the usage of outbred rats especially in such experiments. Although the *p* mutation was found only in four strains, several outbred colonies harbored the *p* mutation at allele frequencies of 0.05–1.00 (Table 5). Besides having a critical role in controlling tyrosinase processing and melanosome biogenesis, P protein might modulate arsenic sensitivity and intracellular glutathione metabolism when expressed in the yeast (19). Considering the extensive utility of the rat in toxicological testing and the prevalence of the *p* mutation in commercial stocks, it would be worth examining sensitivity to arsenicals and other metalloids in rats with or without the *p* mutation.

The genetic makeup of outbred colonies can be affected by historical events such as directional selection and bottlenecks, leading to reduced variation, and genetic drift, mutation, and genetic contamination, which result in genomic differences in individual colonies (3). Thus it is important to recognize that genetic characteristics of outbred colonies tend to change over time. Our genotyping results suggested the occurrences of such genetic alterations in the main outbred colonies that are currently in use. For example, Crj:CD(SD), Jcl:SD, and Slc:SD rat colonies share a common origin, the SD rat colony of Charles River Laboratories of the United States (data sheets from Charles River Laboratories Japan, CLEA Japan, Japan SLC), but they showed different allele frequencies for several functional polymorphisms (Table 5). Although the years when the founders of each colony were imported were different (Crj:SD in 1975, Slc:SD in 1968, and Jcl:SD in 1964), genetic characteristics seemed to have altered during the more than 30 years since their separation from the original stock. Considering the possible genetic alterations of the outbred stock, excellent reproducibility of animal experiments with outbred stocks, for example, for determining sensitivities to substances or for examining physiological parameters, would not be expected. Thus experimental plans based on outbred stocks need to be fully justified to avoid wasting animals and funding.

In this study, we inventoried 16 functional polymorphisms in 136 rat strains and also examined their distribution in several outbred stocks. This study has resulted in a useful catalog of the distribution of disease-related alleles in commonly used rat strains. The finding that 6 of the 13 disease-associated poly-

Table 5. *Allele frequencies of functional polymorphisms in commercially available outbred stocks*

Outbred Stock	No. of Rats	*A*		*Cdkn1a*		*Fcgr3-rs**		*Fdft1*		*Grp10*		*Lss D139N*		*Lss Q481R*		*P*		*Rab38*	
		+	*a*	*A*	*B*	*b*	*ab*	*A*	*S*	*B*	*O*	*A*	*S*	*A*	*S*	+	*p*	+	*r*
Crj:WI(Glx/BRL/Han)IGS	31	0.84	0.16	0.82	0.18	0.13	0.87	1.00	0.00	0.23	0.77	0.02	0.98	0.02	0.98	0.42	0.58	1.00	0.00
Crj:Wistar	31	0.61	0.39	0.61	0.39	0.03	0.97	1.00	0.00	0.21	0.79	0.79	0.21	0.79	0.21	1.00	0.00	1.00	0.00
Jcl:Wistar	32	0.06	0.94	0.72	0.28	0.78	0.22	1.00	0.00	0.94	0.06	0.61	0.39	0.61	0.39	0.00	1.00	1.00	0.00
Slc:Wistar	32	0.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
BrlHan:WIST	32	0.64	0.36	0.72	0.28	0.09	0.91	1.00	0.00	0.45	0.55	0.00	1.00	0.00	1.00	0.67	0.33	1.00	0.00
Crj:CD(SD)	31	0.00	1.00	0.31	0.69	0.00	1.00	1.00	0.00	0.31	0.69	0.32	0.68	0.32	0.68	1.00	0.00	1.00	0.00
Crj:CD(SD)IGS	31	0.10	0.90	0.60	0.40	0.00	1.00	1.00	0.00	0.63	0.37	0.58	0.42	0.58	0.42	1.00	0.00	1.00	0.00
Jcl:SD	32	0.00	1.00	0.48	0.52	0.00	1.00	1.00	0.00	0.02	0.98	0.44	0.56	0.31	0.69	0.95	0.05	1.00	0.00
Slc:SD	32	0.00	1.00	0.48	0.52	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Iar:Long-Evans	30	0.00	1.00	1.00	0.00	0.23	0.77	1.00	0.00	0.00	1.00	0.37	0.63	0.33	0.67	1.00	0.00	1.00	0.00
Kwl:Long-Evans	30	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Crj, Charles River Laboratories Japan; Jcl, CLEA Japan; BrlHan, BRL Hannover; Slc, Japan SLC; Iar, Institute for Animal Reproduction; Kwl, Kiwa Laboratory Animals. *Allele frequencies of the Fcgr3-rs locus could not be determined, because the homozygote for the insertion could not be distinguished from the heterozygote by the standard PCRs.

morphisms are shared between different strains is important for experimental design, because this principle is likely to hold true for many other uncharacterized functional polymorphisms. To help research interests to select the most suitable strains for their experiments, the NBRP-Rat will periodically improve and update the genetic profiles of rat strains.

ACKNOWLEDGMENTS

We are grateful to M. Yokoe and R. Okajima for excellent technical assistance. OLETF and LETO rats were provided by Otsuka Pharmaceutical Co. Ltd. (Tokushima, Japan). F344/Snk was provided by Medicinal Safety Research Laboratories, Sankyo Co., Ltd. (Shizuoka, Japan).

This work was supported by the National Bio Resource Project for the Rat and carried out as a part of the genetic profiling program of the Project.

GRANTS

This work was supported in part by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (18300141 to T. Kuramoto and 16200029 to T. Serikawa) and a Grant-in-aid for Cancer Research from the Ministry of Health, Labour and Welfare.

REFERENCES

1. **Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT.** Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439: 851–855, 2006.
2. **Aitman TJ, Glazier AM, Wallace CA, Cooper LD, Norsworthy PJ, Wahid FN, Al-Majali KM, Trembling PM, Mann CJ, Shoulders CC, Graf D, St Lezin E, Kurtz TW, Kren V, Pravenec M, Ibrahimi A, Abumrad NA, Stanton LW, Scott J.** Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat Genet* 21: 76–83, 1999.
3. **Chia R, Achilli F, Festing MF, Fisher EM.** The origins and uses of mouse outbred stocks. *Nat Genet* 37: 1181–1186, 2005.
4. **Esumi H, Takahashi Y, Sato S, Nagase S, Sugimura T.** A seven-base-pair deletion in an intron of the albumin gene of analbuminemic rats. *Proc Natl Acad Sci USA* 80: 95–99, 1983.
5. **Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, Heward JM, Gough SC, de Smith A, Blakemore AI, Froguel P, Owen CJ, Pearce SH, Teixeira L, Guillevin L, Graham DS, Pusey CD, Cook HT, Vyse TJ, Aitman TJ.** *FCGR3B* copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39: 721–723, 2007.
6. **Gibbs RA et al.** Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493–521, 2004.
7. **Glazier AM, Scott J, Aitman TJ.** Molecular basis of the *Cd36* chromosomal deletion underlying SHR defects in insulin action and fatty acid metabolism. *Mamm Genome* 13: 108–113, 2002.
8. **Gotoda T, Iizuka Y, Kato N, Osuga J, Bihoreau MT, Murakami T, Yamori Y, Shimano H, Ishibashi S, Yamada N.** Absence of *Cd36* mutation in the original spontaneously hypertensive rats with insulin resistance. *Nat Genet* 22: 226–228, 1999.
9. **Kuramoto T, Gohma H, Kimura K, Wedekind D, Hedrich HJ, Serikawa T.** The rat pink-eyed dilution (*p*) mutation: an identical intragenic deletion in pink-eye dilute-coat strains and several Wistar-derived albino strains. *Mamm Genome* 16: 712–719, 2005.
10. **Kuramoto T, Nomoto T, Sugimura T, Ushijima T.** Cloning of the rat agouti gene and identification of the rat nonagouti mutation. *Mamm Genome* 12: 469–471, 2001.
11. **Kwitek AE, Jacob HJ, Baker JE, Dwinell MR, Forster HV, Greene AS, Kunert MP, Lombard JH, Mattson DL, Pritchard KA Jr, Roman RJ, Tonellato PJ, Cowley AW Jr.** BN phenome: detailed characterization of the cardiovascular, renal, and pulmonary systems of the sequenced rat. *Physiol Genomics* 25: 303–313, 2006.
12. **MacMurray AJ, Moralejo DH, Kwitek AE, Rutledge EA, Van Yserloo B, Gohlke P, Speros SJ, Snyder B, Schaefer J, Bieg S, Jiang J, Ettinger RA, Fuller J, Daniels TL, Pettersson A, Orlebeke K, Birren B, Jacob HJ, Lander ES, Lernmark A.** Lymphopenia in the BB rat model of type 1 diabetes is due to a mutation in a novel immune-associated nucleotide (*Ian*)-related gene. *Genome Res* 12: 1029–1039, 2002.
13. **Mashimo T, Voigt B, Kuramoto T, Serikawa T.** Rat Phenome Project: the untapped potential of existing rat strains. *J Appl Physiol* 98: 371–379, 2005.
14. **Mashimo T, Voigt B, Tsurumi T, Naoi K, Nakanishi S, Yamasaki K, Kuramoto T, Serikawa T.** A set of highly informative rat simple sequence length polymorphism (SSLP) markers and genetically defined rat strains. *BMC Genet* 7: 19, 2006.
15. **Mori M, Li G, Abe I, Nakayama J, Guo Z, Sawashita J, Ugawa T, Nishizono S, Serikawa T, Higuchi K, Shumiya S.** Lanosterol synthase mutations cause cholesterol deficiency-associated cataracts in the Shumiya cataract rat. *J Clin Invest* 116: 395–404, 2006.
16. **Mori M, Sawashita J, Higuchi K.** Functional polymorphisms of the *Lss* and *Fdft1* genes in laboratory rats. *Exp Anim* 56: 93–101, 2007.
17. **Oiso N, Riddle SR, Serikawa T, Kuramoto T, Spritz RA.** The rat Ruby (*R*) locus is *Rab38*: identical mutations in Fawn-hooded and Tester-Moriyama rats derived from an ancestral Long Evans rat sub-strain. *Mamm Genome* 15: 307–314, 2004.
18. **Serikawa T.** Colourful history of Japan's rat resources. *Nature* 429: 15, 2004.
19. **Staleva L, Manga P, Orlow SJ.** Pink-eyed dilution protein modulates arsenic sensitivity and intracellular glutathione metabolism. *Mol Biol Cell* 13: 4206–4220, 2002.
20. **Steingrimsson E, Copeland NG, Jenkins NA.** Mouse coat color mutations: from fancy mice to functional genomics. *Dev Dyn* 235: 2401–2411, 2006.
21. **Takiguchi S, Takata Y, Funakoshi A, Miyasaka K, Kataoka K, Fujimura Y, Goto T, Kono A.** Disrupted cholecystokinin type-A receptor (*CCKAR*) gene in OLETF rats. *Gene* 197: 169–175, 1997.
22. **Watanabe TK, Suzuki M, Yamasaki Y, Okuno S, Hishigaki H, Ono T, Oga K, Mizoguchi-Miyakita A, Tsuji A, Kanemoto N, Wakitani S, Takagi T, Nakamura Y, Tanigami A.** Mutated G-protein-coupled receptor *GPR10* is responsible for the hyperphagia/dyslipidaemia/obesity locus of *Dmo1* in the OLETF rat. *Clin Exp Pharmacol Physiol* 32: 355–366, 2005.
23. **Yamashita S, Furumoto K, Nobukiyo A, Kamohara M, Ushijima T, Furukawa T.** Mapping of A gene responsible for cataract formation and its modifier in the UPL rat. *Invest Ophthalmol Vis Sci* 43: 3153–3159, 2002.
24. **Yamashita S, Suzuki S, Nomoto T, Kondo Y, Wakazono K, Tsujino Y, Sugimura T, Shirai T, Homma Y, Ushijima T.** Linkage and microarray analyses of susceptibility genes in ACI/Seg rats: a model for prostate cancers in the aged. *Cancer Res* 65: 2610–2616, 2005.
25. **Yamashita S, Wakazono K, Nomoto T, Tsujino Y, Kuramoto T, Ushijima T.** Expression quantitative trait loci analysis of 13 genes in the rat prostate. *Genetics* 171: 1231–1238, 2005.
26. **Yokoi N, Komeda K, Wang HY, Yano H, Kitada K, Saitoh Y, Seino Y, Yasuda K, Serikawa T, Seino S.** *Cblb* is a major susceptibility gene for rat type 1 diabetes mellitus. *Nat Genet* 31: 391–394, 2002.
27. **Yokoyama M, Amano S, Tsuji A, Sasahara M, Serikawa T, Ihara N, Matsuda M, Hazama F, Handa J.** Genetic analysis of cataract in Ihara epileptic rat. *Mamm Genome* 12: 207–211, 2001.

TECHNICAL REPORTS

nature genetics

SNP and haplotype mapping for genetic analysis in the rat

The STAR Consortium*

The laboratory rat is one of the most extensively studied model organisms. Inbred laboratory rat strains originated from limited *Rattus norvegicus* founder populations, and the inherited genetic variation provides an excellent resource for the correlation of genotype to phenotype. Here, we report a survey of genetic variation based on almost 3 million newly identified SNPs. We obtained accurate and complete genotypes for a subset of 20,238 SNPs across 167 distinct inbred rat strains, two rat recombinant inbred panels and an F_2 intercross. Using 81% of these SNPs, we constructed high-density genetic maps, creating a large dataset of fully characterized SNPs for disease gene mapping. Our data characterize the population structure and illustrate the degree of linkage disequilibrium. We provide a detailed SNP map and demonstrate its utility for mapping of quantitative trait loci. This community resource is openly available and augments the genetic tools for this workhorse of physiological studies.

The unique power of the laboratory rat resides in the extensive biological characterization of a wide range of inbred strains representing models for common human diseases (see URL 1 in URLs section of Methods below). Although the rat is primarily known as a physiological model, there has been a steady increase in the use of the rat in genetic and genomic studies over the last decade[1]. The genome of the Brown Norway rat (BN/NHsdMcwi) has been sequenced, but, as a sequence of a single inbred rat strain, it provided little insight into the genetic variation that is responsible for the wide range of disease phenotypes, drug resistance or variability in toxicology responses in the different rat strains. At present, genetic variability in the rat genome is usually assayed using a limited set of microsatellite markers[1–4]. A dense set of polymorphic markers, for which SNPs provide the most cost-effective solution, would transform the genetic toolkit available for rat biologists.

The breeding history of rat strains, in common with that of many other laboratory animals, is known to have a complex genesis[5], with a number of unknown relationships in the formation of the laboratory strains. In addition, strains often carry the same designation but substrains are not necessarily identical because in a number of cases breeding stocks were distributed before the line became inbred, with varying physiological consequences[6,7]. Thus it is important to have detailed marker information available on any specific substrain. The presence of a number of recombinant inbred lines (RI)—in particular the HXB-BXH sets, derived from the BN-Lx/Cub and SHR/Ola parental strains and the FXLE-LEXF sets, derived from the F344/Stm and the LE/Stm strains—and the presence of both congenics and consomics provides a rich set of renewable genetic resources available for rat biologists to examine the variation of phenotypes between different genotypes.

To study genome-wide genetic variation, we initiated the genetic dissection of the ancestral segments making up the most commonly used rat inbred lines, and we developed a comprehensive open resource of validated SNP markers for essentially any strain combination. We provide extensive maps of strain distribution patterns (SDPs) for the two largest rat recombinant inbred (RI) strains, estimations on linkage disequilibrium, and haplotype structure in the rat genome and evaluate the use of correlation between phenotype and ancestral sequence origin across many inbred strains required to facilitate the identification of underlying alleles. This study provides a set of permanent resources for rat genetics (SNPs, SDPs and genetic maps), immediately facilitates more statistically powerful analysis on the RI strains and provides insight in the genesis of the different rat strains available to researchers today.

RESULTS

Polymorphisms in the rat genome and generation of a SNP map

We generated a SNP map of the rat genome containing about 3 million distinct SNPs mapped to the draft genome sequence, at an average density of approximately one SNP per 800 bps. Three distinct sources of DNA sequencing reads were used for automated computational SNP discovery (**Table 1**): (i) shotgun sequence generated from the four strains SS/Jr, WKY/Bbb, GK/Ox and SHRSP/Bbb, all widely used in disease mapping experiments in crosses and congenic strains[8]; (ii) genome-wide shotgun sequence at ×1.5 coverage of the outbred Sprague-Dawley rat generated by Celera; and (iii) BAC end sequences from the F344/Stm rat, which is also widely used in genetic studies. For the Sprague-Dawley rat, approximately 14% of the SNPs were heterozygous on the basis of evidence from overlapping aligned sequencing reads. Because of the relatively low coverage, this is an underestimate of the heterozygosity in this strain. As expected by the relatively low sequencing coverage, we observed that homozygous SNPs were supported on average by fewer aligned reads than

*The complete lists of participants and affiliations appear at the end of the article. Correspondence should be addressed to N.H. (nhuebner@mdc-berlin.de).

Received 30 October 2007; accepted 27 February 2008; published online 28 April 2008; doi:10.1038/ng.124

Table 1 Number of sequencing reads and detected SNPs per strain

Source	Number of reads	Non-redundant SNPs
STAR shotgun sequence	249,525	128,976 (SS/Jr: 56,639)[a] (WKY/Bbb: 32,601)[a] (GK/Ox: 16,838)[a] (SHRSP/Bbb: 28,332)[a]
Celera Sprague-Dawley genome-wide shotgun sequence	7,990,225	2,650,525
F344 BAC end sequence	344,064	196,812
Total		2,976,313

[a]Number of SNPs before removal of redundancy.

heterozygous SNPs, which leads to a systematic underestimation of the true heterozygosity in this strain.

To assess the utility of the SNPs for rat genetics we genotyped a subset ($n = 20,283$) in 167 inbred strains and two RI panels of 31 and 33 strains and tested a subset ($n = 9,691$) in 89 F_2 animals. The allele frequencies of SNPs across the inbred strains were approximately evenly distributed between 3% and 50%. We assayed seven 1,536-plex assays that could be run on the Illumina BeadLab station (10,752 SNPs). In parallel, we developed a 9,691-SNP rat targeted genotyping panel (called the '10K panel') to be run on an Affymetrix platform (see **Supplementary Methods** online). The rationale behind this separation was to evaluate the appropriate technology for efficient SNP genotyping in the rat. These genotyping tools now are commercially available from Illumina and Affymetrix, respectively. The two panels together yielded 20,283 validated SNPs. We genotyped 1,057 SNPs with both platforms in 231 rats as a control, with a concordance of 99.8%. Furthermore, we constructed separate phylogenetic trees using data from each genotyping platform and found that the clustering was very stable, even for those nodes that are not supported by a high bootstrap value (**Supplementary Fig. 1** online).

Annotation of putative functional SNPs from inbred strains

We predicted the functional effects of 325,788 SNPs. This analysis was restricted to SNPs derived from the five inbred strains used in the SNP discovery (SS/Jr, GK/Ox, SHRSP/Bbb, WKY/Bbb and F344/Stm). We excluded the large set of SNPs identified from the outbred Sprague-Dawley rat. This ensures that all predicted functional consequences can be tested experimentally in stable inbred strains, eliminating the uncertainty that a particular animal may not carry the described allele because of incomplete inbreeding of the colony. We estimated the selective pressure on amino acid replacement mutation for all residues with nonsynonymous coding SNPs ($n = 1,160$) by calculating the omega value, or ratio of nonsynonymous/synonymous substitution rates[9]. We identified 56 SNPs with omega values lower than 0.1 (**Supplementary Table 1** online), which is indicative of a likely effect in protein function; seven of these lie in rat orthologs of human disease genes involved in hereditary diseases or cancer. These include the gene *ALDH2*, involved in acute alcohol intolerance, the gene *PCCB*, involved in propionic acidemia, the cancer gene *AFF4* (AF4/FMR2 family member 4) and the proto-oncogene tyrosine kinase receptor ret precursor (*RET*).

Further, 324 SNPs are predicted to disrupt the normal pattern of splicing, and 57 SNPs and 63 SNPs, respectively, create a potential donor splice site (GT) or potential acceptor splice site (AT). Finally, we assessed the potential effect of the SNPs on transcriptional and post-transcriptional regulation. One thousand nineteen SNPs in promoter regions map into conserved transcription factor binding sites and 568 SNPs map into DNA triplexes[10]. One hundred thirty-two SNPs in 3′ UTR regions affect microRNA targets (**Supplementary Table 1**).

Phylogenetic relationships among rat strains

It is unclear from documented information how ancestral subspecies, strains and individual rats have contributed to shaping the genomes of the modern laboratory rat strains[11]. Considering the many rounds of inbreeding and interbreeding that most likely took place, a complex evolutionary history can be expected. Therefore, we chose to visualize the interstrain relationship and genetic proximity in a phylogenetic network[12] rather than a tree (**Fig. 1**). The observed strain relationships agreed very well with the known history of rat strains and supported all significant clusters of strains previously recognized in analysis of microsatellite markers[5]. The reticulation around the center of the network may reflect extensive genetic heterogeneity of the ancestral rat population. The network revealed a complex breeding history of WKY-related rat strains, corroborating the notion that breeding stocks of several strains (WKY, SHR, SHRSP) were distributed before the line became inbred[6,7]. The Brown Norway (BN) rats were placed as the most diverged strain, which might be explained by the SNP ascertainment bias, because the BN genome sequence was used as a reference sequence for SNP discovery. However, the separation of BN is also supported by microsatellite data[5,13] and a study that used different SNP panels[14]. The inclusion of wild and outbred rats into the phylogenetic

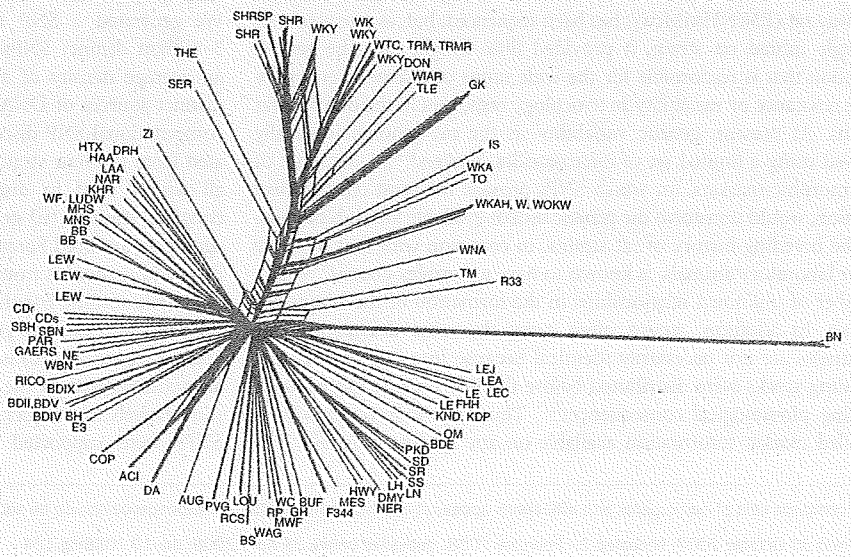

Figure 1 Phylogenetic neighbor-net network constructed from 20,283 polymorphic positions genotyped in 167 laboratory rats. Reticulation in the center of the network likely reflects genetic heterogeneity of the ancestral rat population. The group of WKY-related strains shows a complex pattern of relationships, due probably to incomplete inbreeding of stocks before they were disseminated to various laboratories and subsequently inbred to completion. The network also shows presence of residual inter-isolate variation within SHR, LEW, BB, WKY, LE, GK and BN inbred strains.

analysis resulted in their branching from the reticulate center and not outgrouping of BN or non-BN rats. Within non-BN strains, we defined ten clusters of strains that evidently shared breeding history; these clusters were highly supported as monophyletic groups on a traditional phylogenetic tree (**Supplementary Fig. 1**).

An important issue for studies that involve inbred rats is their degree of inbreeding and the variation between substrains that are maintained at different laboratories. Previously it has been shown that BN and DA inbred substrains obtained from different locations harbor genetic differences[14]. Our dataset included multiple substrain samples and confirmed the presence of variation between substrains. This effect was most pronounced in LE (29% of genotyped variable sites were different in the pairwise substrain comparison), WKY (up to 19%), LEW (13%), SHR (11%), BB (10%), PKD (5%) and, to a lesser degree, in GK (1%) and BN (0.6%) inbred strains. These observations indicate that, at least for some inbred strains, the use of different substrains may have a significant effect on the outcome and reproducibility of experimental results.

Estimations on linkage disequilibrium and haplotype structure in the rat genome

Although the genotyped panel of markers is not dense enough to support conclusive evaluation of linkage disequilibrium (LD) structure and a complete haplotype map, it can be used for obtaining the estimates on haplotype length and its comparison with that of other organisms. Using 15,901 SNPs with minor allele frequency of >5% among all genotyped samples, we defined the haplotype blocks as adjacent SNPs lacking historical recombination[15]. Using Haploview[16], a total of 837 blocks were detected, encompassing 19% of SNPs and covering about 12% of the rat genome, with an average block size of 411 kb. These included 323 blocks of only two SNPs, but these blocks were relatively small and in total covered less than 2 Mb. In contrast, the average size of the 514 blocks that contained three or more SNPs was 665 kb. The observed block structure completely disappeared upon permutation of SNP positions and was relatively stable when the number of substrains or density of markers were randomly reduced by 10% (**Supplementary Fig. 2** online). If we extrapolate from the current data, we estimate that it will require another 50,000 to 75,000 SNPs to define the remaining haplotype structure comprehensively.

We compared rat haplotype structure with unpublished mouse haplotype data from the Broad Institute (see URL 2 below). We balanced the mouse and rat sets to contain the same number of strains ($n = 38$), SNP density (6.4 kb^{-1}) and inter-SNP distance distribution. Under the same criteria for haplotype block partitioning, the extent of LD was larger in laboratory mice, where haplotype blocks covered a larger fraction of the genome (35% compared to 12% in the rat), contained a higher proportion of informative markers (56% versus 21%) and had a greater average size (648 kb versus 388 kb). The direct comparison of LD decay profiles in rat and mouse (**Supplementary Fig. 3** online) further substantiated this notion. On the other hand, linkage disequilibrium in the rat was larger than that in cow, for which haplotype blocks cover only 2.2% of autosomes[17]. Although LD in rats was less pronounced than that in mice, it still extended over hundreds of kilobases, unlike LD in humans or across dog breeds, for which the correlation coefficient r^2 drops below 0.1 at 100 kb[18–20]. These results suggest that the breeding histories of laboratory rats and mice are qualitatively different. However, there are also considerable differences in extent of LD and complexity of phylogenetic relationships when rat and mouse laboratory populations are compared. In the mouse, large LD blocks can be recognized that reflect ancestral contributions from different subspecies[17–19], whereas there is no such evidence in the rat. At the same time, the phylogenetic relationships among groups of rats are hard to deduce (**Supplementary Fig. 1**), reflecting more divergent genetic background of a rat founder population. Comparison of LD between rat and mouse showed that the size of haplotype blocks in syntenic regions showed small, but significant, correlation ($r^2 = 0.18$), consistent with the previously observed correlation between murine (mouse and rat) recombination rate and fine LD structure[21,22]. Olfactory genes were the only gene class overrepresented in rat LD blocks ($P < 10^{-6}$). Twenty-one distinct gene clusters harboring about one-third ($n = 325$) of all olfactory genes were located in LD blocks longer than 500 kb. The same phenomenon was observed in mice (130 genes, $P < 10^{-7}$ for blocks exceeding 2 Mb), suggesting an increased selective pressure on rodent genes involved in sensory perception of smell.

Notably, we identified 939 interchromosomal SNP pairs in full linkage disequilibrium. These SNPs were heavily shifted toward low minor-allele frequencies, with only 38 and 4 of them having minor-allele frequencies larger than 0.1 and 0.15, respectively. More detailed inspection revealed that, besides being rare, the vast majority of these variants were private to branches of the phylogenetic tree (for example, many of them were restricted to the SHR + WKY + GK cluster). Thus, perfectly correlated SNPs on different chromosomes are unlikely to result from epistatic effects or genome assembly errors, but are more likely to represent a shared physical genomic structure (that is, genetic background). It should be mentioned that imperfect but significant pairwise correlation ($r^2 \geq 0.5$) was observed among about 0.2% of the interchromosomal SNP pairs. The highly correlated subset disappeared almost completely when SNP alleles were randomized and could thus reflect epistatic interactions as well as ancestral relationships.

Prospects for genome-wide association mapping using inbred rat strains

One hundred of the most diverse inbred rat strains were evaluated by simulation for their potential for genome-wide association mapping of quantitative trait loci (QTL). The method originated in the mouse genetics community[23], where it has generated much discussion[24]. From our simulations, we found that the threshold for genome-wide significance varied depending on the extent and nature of the genetic component of the phenotypic variance. For example, the genome-wide threshold for significance when there was no genetic component to the phenotypic variance was $\log P_{max} = 4.1$, close to the Bonferroni estimate of 4.3. In contrast, for an infinitesimal model in which many small-effect QTL, in total accounting for 50% of the total variance, were distributed uniformly across the genome, there was a median $\log P_{max} = 21.5$; that is, much higher. For a single major QTL explaining 50% of the variance, the genome-wide maximum coincided with the true position in 31% of simulations, and there was a local maximum exceeding the genome-wide null threshold of significance at the true QTL in 51% of simulations and within 1 Mb of the true position in 91% of cases. However, there was a median $\log P_{max} = 14.3$, which lies between the two thresholds above, and, on average, 72 putative QTL exceeded the null threshold of 4.1. Finally, for a realistic complex trait scenario of ten 5% QTL, there was a median $\log P_{max} = 9.96$, and the median number of putative QTL exceeding 4.1 was 1,412. Thus, there were a very large number of false positive QTL, and consequently each true QTL was close to a putative QTL. **Supplementary Table 2** online gives the numbers of putative QTL identified at different thresholds and illustrates the problem of balancing true and false positive rates. For example, a threshold of 8 limits the number of putative QTL to only twice the number of true QTL, but the majority of putative QTL locations do not coincide with the true locations at this threshold.

Construction of a rat genetic map using an F_2 cross and recombinant inbred lines

We typed a total of 20,283 SNPs in two independent panels of recombinant inbred (RI) strains derived from SHR and BN-Lx rats (HXB-BXH) ($n = 31$) and from F344/Stm and LE/Stm rats (FXLE-LEXF) ($n = 33$), and we typed 9,691 SNPs in 89 progeny of an F_2 cross between BN/Par and GK/Ox rats (GK × BN). These populations have been used for mapping complex phenotypes for metabolic syndrome[25], expression QTL (eQTL)[26] and metabonomic traits[27], and extensive phenotype characterizations as part of the Japanese phenome project (see URL 3 below and ref. 28). In addition, genotype analysis in F_2 rats enabled assessment of the reliability of heterozygous genotype calls.

Genetic map construction was initially carried out in the GK × BN F_2 cross with the JoinMap program as previously described[2]. This approach has the advantage that prior knowledge of markers' physical order is not required for calculating genetic distances. More than 8,400 microsatellite and SNP markers have now been mapped in this cross and SNP typing has significantly improved the resolution of the genetic maps (**Supplementary Table 3** online). Following genotype verifications, we confirmed the existence of strong distortion of segregation previously reported in chromosomes 3, 4, 9 and 13 (ref. 2). Alignment of genetic and physical maps showed the general agreement of marker order and distance (**Supplementary Fig. 4** online). However, we identified regions (from 1 SNP to 8 Mb) where SNP-based genetic maps were inconsistent with the current rat genome assembly draft.

We then repeated genetic mapping in this cross and both panels of RI strains using the R and R/QTL software packages[29,30] integrating SNP genotype and physical map data, resulting in 16,543 SNPs mapped. Data were initially filtered to remove markers containing genotyping errors (for example, absence of segregation in the cohort despite apparent allele variation in the parental strains) and blocks of adjacent SNPs with identical segregation patterns were collapsed into SDPs. Markers that generated suspiciously large map distances were removed, using criteria derived from the approximately linear relationship of genetic and physical distances. Details of the typed markers and mapped positions are given in **Supplementary Table 4** online and URL 4 below and the resulting maps in **Supplementary Figure 5** online. We found strong evidence of discrepancies between the genetic map and the draft genome assembly (**Fig. 2**). In particular, genetic mapping in all three panels identified a p11-centromeric segment of chromosome 1 that has been wrongly assembled in the p14-telomeric region of chromosome 17. Genetic mapping data suggested further intra- and interchromosomal relocations in regions of chromosomes 2, 4, 11, 12, 14, 17. Known conflicts between rat genome assemblies, provided by BCM and Celera (see URL 5 below), indicated the relocation in the p14 region of chromosome 17, supporting the Celera assembly, and one conflict on chromosome 9 was resolved favoring the BCM assembly (data not shown). The other conflicting mapping results require further independent verifications.

When we set out to construct a genetic map for the X chromosome based on the physical order of markers, we detected several unlinked markers, which rendered the mapping impossible. In-depth investigation of these linkage breaks revealed that they occurred on contig boundaries (**Supplementary Table 5** online). We rearranged the fragments of the chromosome resulting from splitting the contigs that were not linked (lod score < 2) in the order that generated the smallest average recombination fraction in the three populations (**Supplementary Fig. 6a** online). Using the resulting marker positions, we constructed three genetic maps (**Supplementary Fig. 6b**), summarized in **Supplementary Table 4**.

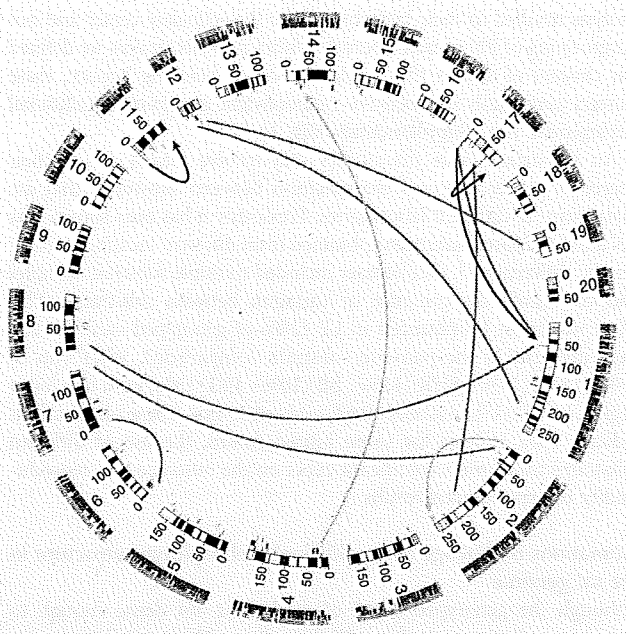

Figure 2 Identified discrepancies between rat genome assembly and genetic maps. Rearrangement of the physical map according to genetic mapping information. Data from each cohort is color coded (red, FXLE-LEXF; green, HXB-BXH; blue, GK × BN). Outer circle, positions of informative SNPs for each cohort. Bars in the inner circle, conflicts in the genetic map. Arrows, relocation, according to minimal recombination fraction, of SNP markers that had extreme genetic distances compared to their physical distance from adjacent markers. For black lines, all crosses support the rearrangement; lime green, HXB-BXH and F_2 cross support the rearrangement; orange, unresolved genomic conflicts.

SDPs for mapping quantitative traits in rat recombinant inbred strains

We carried out quantitative trait mapping for a subset of 74 traits that were measured in the FXLE-LEXF panel of 33 RI strains[31] and their parental progenitors F344/Stm and LE/Stm as part of the Japanese Rat Phenome Project[28]. Of the 20,283 SNPs tested, 28.5% (5,778) were polymorphic between the parental strains, with 1,033 distinct SDPs across the 33 RI strains. In total, we identified 250 significant QTL (false discovery rate < 0.05) for 74 phenotypic parameters (see **Supplementary Table 6** online and URL 2 below). Although we detected loci previously reported for a number of traits (for example, cholesterol levels; **Supplementary Fig. 7** online), most of the significant linkages were new, as most of the phenotypes assessed here have not been mapped in the rat previously (**Supplementary Table 6**). The number of SDP identified by the current SNP map increased more than threefold over the existing microsatellite-based map[13] and determined 3,766 recombination events for the 33 RI strains. This resulted in a marked improvement of genome wide coverage and greater QTL mapping resolution. Our results demonstrate that this RI resource, historically generated to study genes involved in tumorigenesis, is applicable to the detection of physiological and behavioral traits and risk factors for complex diseases.

Accessibility of the data

The complete set of newly reported SNPs and the entire set of genotypes across all rat strains are publicly accessible, through Ensembl (see URL 7 below) and other web sites (**Supplementary**

Note online). The BioMart tool provides a particularly flexible interface to this data, whereby both genomic position queries and gene list queries can be used to select a set of SNPs that are polymorphic between two strain combinations. The new SNPs are fully integrated with SNPs from other sources on Ensembl overview displays such as ContigView. A tool to select subsets of SNPs is available (see URL 8 below), as is visualization of polymorphisms along chromosomes (see URL 9). To facilitate access to the data in the context of further information, the data represented here has been integrated in other databases, namely the Rat Genome Database (RGD; see URL 1) and GeneNetworks (see URL 10). Finally, data presented for functional assessment of the SNP consequence type is comprehensively available (see URL 11).

DISCUSSION

We present a comprehensive study of genetic variation in the laboratory rat in order to accelerate its use as a model of human complex diseases. To this end, we have identified approximately 3 million SNPs and predicted the functional effects of 325,788 of them. This brings a far richer genetic toolkit to this common toxicology and physiology model mammal, and the presence of pre-typed renewable genetic resources, such as RI lines, provides a resource in which any phenotypic assay available on rat can be augmented by a genotype scan, provided the assay can be performed on the RI panel. The functional analysis of SNPs is in its infancy, but already provides a useful priority list of potential functional variants for further testing, in particular when combined with other data, such as eQTL.

We genotyped 20,283 selected SNPs that were distributed evenly across the genome in 167 inbred strains and 64 recombinant inbred lines, resulting in a community resource of validated polymorphic markers for any strain combination. These strains represent founders of crosses with more than 90% of the rat QTL reported in the literature, and thus this resource will serve as a valuable tool for functional genomics and facilitate positional cloning of QTL and the identification of causal variants.

Our analysis of the evolutionary history of the rat based on these data showed that there were ten clusters of strains sharing breeding history and that the Brown Norway strain separated phylogenetically from all other strains. Our first-generation haplotype map of the laboratory rat showed the genomic history of genomic segments and may allow for the imputation of genotypes in other strains with sparser genotype and sequence data. Notably, our study showed different extents of LD in populations of laboratory mice and rats. Moreover, the phylogenetic relationships inferred from the genotype data suggested a more complex origin and relationships between rat strains than between mouse strains. Theoretically, applying correlation between phenotype and ancestral sequence origin across many inbred strains could enable the identification of genomic regions that are likely to contain the responsible genes. However, our simulations showed that genome-wide association mapping using 100 inbred rat strains is only practicable for single large-effect QTL, and even in these contexts it is not guaranteed to identify the QTL location. Nevertheless, the method does show promise for single large-effect eQTL. Moreover, knowledge of phylogenetic relationships between strains may help in the selection of informative strains for further phenotypic characterization.

The genetic maps that were generated from RI panels and an F_2 cross showed that the draft genome sequence is largely correct, but did also reveal several regions that need further investigation. In addition, we provided a high-resolution map of the contribution of ancestral genomic segments for every individual strain in two rat recombinant inbred panels. The utility of such information was illustrated by mapping QTL for 74 phenotypic parameters in one of these RI panels (FXLE-LEXF).

The availability of robustly assayed SNPs and renewable genetic resources provided here constitutes the next step for the genetic toolkit for the rat. The rat is extensively used in many biological assays, and lowering the cost and other barriers for the application of genetic tools to this organism provides many synergies between the vast range of existing working assays on this organism and a powerful genetic toolkit. We expect that this resource will lead in the future to the resequencing of key strains, the discovery of more genetic associations, and their final resolution to a molecular variant, leading to a new avenue to research human disease.

METHODS

Animals. We used 167 inbred rat strains that covered the diversity of the most commonly used strains in research. Tissue was provided by researchers from the rat community and DNA extraction was performed at the MDC. The list of strains with designation and ILAR code can be found at the web page of the STAR consortium (see URL 4 below). Also, most of the strains are listed in RGD (see URL 1). For most of these strains, QTL data are available. In our analysis we captured strains that encompass about 90% of the rat QTL reported in the RGD (see URL 1). The sets of recombinant inbred strains and the F_2 cross are described in the **Supplementary Methods.**

Genomic shotgun fragment sequencing. Shotgun libraries of a single male rat for each of strains SS/Jr, WKY/Bbb, GK/Ox and SHRSP/Bbb were constructed by sonication of 15 μg of genomic DNA. Fragments between 800–2,000 bp in length were subcloned into pUC18 and clones were sequenced from both ends using BigDye terminator chemistry (v3.1) and ABI3730 sequencers (Applied Biosystems). Further information on base calling methods is available in the **Supplementary Methods.**

BAC library construction and end sequencing. The RNB1 rat BAC library was produced by cloning partially SacI-digested genomic DNA isolated from peripheral lymphocytes of a male rat of strain F344/Stm into the pKS145 vector[32]. The BAC library, consisting of 172,800 clones, was used for BAC-end sequencing. BAC DNA extractions were performed using the PI-1100 plasmid isolator (Kurabo), and BAC clones were sequenced using BigDye terminator (v3.1) sequencing kits and ABI 3730 sequencers (Applied Biosystems). Raw sequence data were base-called by KB Basecaller. All sequences were submitted to the DNA Databank of Japan. BAC clones are available from the RIKEN BioResource Center DNA bank (see URL 12 below).

SNP calling. SNP discovery used the SSAHAsnp algorithm[33]. Briefly, this procedure aligned the sequencing reads above to version 3.4 of the rat genome assembly. We apply several filters on alignment quality and neighborhood quality standard, which is defined by the PHRED score of the variant base and surrounding bases.

SNP selection and genotyping. For Illumina GoldenGate, genotyping was carried out using the GoldenGate protocol in a fully automated BeadLab[34]. Samples were processed in 96-well plates. For Affymetrix Targeted Genotyping, genotyping was carried out using the GeneChip Scanner 3000 Targeted Genotyping System protocol from Affymetrix, originally described as MIP technology[35,36].

Data was subjected to stringent quality control procedures eliminating samples and SNPs that did not reach sufficiently high call rates. All SNPs with more than 10% heterozygous genotypes in the inbred strains were removed from the analysis in the final dataset. Also, SNPs with a call rate below 90% were dropped. Our conclusive dataset of 20,283 SNPs comprised 99.2% of all SNPs genotyped, with an overall success rate of 98.7%, covering the genome with an average distance of 130 kb. More information about the genotyping design is given in the **Supplementary Methods.**

Computing functional predictions. We estimated selective pressure by calculating the omega value as previously described[9]. Transcription factor binding

sites (TFBSs) in the upstream region of the genes were identified by scanning the promoter region with the MatScan and JASPARS[37] collection of matrices. Next, we identified the TFBSs conserved between species (human and rat) using meta-alignments[38]. Finally, we identified the SNPs that mapped into the conserved TFBSs (1,019). These SNPs are considered to have a putative effect in the expression of the gene.

Phylogenetic structure predictions. We used the genotype information, encompassing 20,283 genome positions from 167 inbred strains, to determine the phylogenetic relationships among the strains. For building a split network, we used the NeighborNet method with uncorrected p-distances implemented in Splitstree4.8 software[12]. We also produced more traditional tree-like structure calculated by MEGA4 package[39] using the neighbor-joining method with uncorrected p-distances and bootstrap test with 1,000 replicates.

Linkage disequilibrium and haplotype structure. We used Haploview 3.32 software[16] to estimate haplotype block structure in rat and mouse laboratory strains. Custom Perl scripts were written to allow selection of SNPs and identification of the most genetically divergent rat strains, to facilitate SNPs/strains randomization or random removal and to calculate LD decay profiles. These scripts are available from authors upon request. Functional analysis of overrepresentation of gene ontology terms for genes located in high LD regions was done with gProfiler web-server[40]. Further information on haplotype analysis is given in the **Supplementary Methods.**

Genetic map constructions and QTL analysis. Genetic mapping in the GK × BN F_2 cross was carried out with JoinMap as previously described[2]. We then used an automatic construction procedure for the genetic map of the HXB-BXH and FXLE-LEXF RI populations and the GK × BN cross from SNP markers and the physical map positions of the SNPs. Next we used the empirical observation of larger number of recombinations between markers with increasing physical distance for an automated reconstruction of the map. The procedure involved systematically evaluating the removal of markers that generate suspiciously large distances in the map. The criterion to call an interval suspicious was defined by a linear model. The model was defined by a user-specified intercept, which is the minimal genetic distance at which distances are considered for removal, and a slope that was computed chromosome-wise from the data. We set this threshold to 3 cM. In order to determine the slope, an initial genetic map was estimated for all markers using the order defined by the physical map. Then all map distances greater than the 95% quantile were removed and the slope was defined as the sum of the remaining genetic distances over the sum of physical distances between markers. The algorithm performed these steps for each chromosome: (i) compute the initial map based on the physical order of markers; (ii) estimate the linear model; (iii) while the size of the genetic map is reduced, evaluate the size of the genetic map when removing candidate markers and select the marker leading to the minimal map size.

For QTL mapping in LEXF-FXLE RI strains, calculations were performed with WinQTL Cart version 2.5 (see URL 13 below). Composite interval mapping was used as QTL mapping strategy. A detailed description of the QTL mapping strategy is given in the **Supplementary Methods**, as is our analysis of simulations on genome-wide association.

URLs. URL 1, http://rgd.mcw.edu/strains/; URL 2, http://www.broad.mit.edu/~claire/MouseHapMap; URL 3, http://www.anim.med.kyoto-u.ac.jp/nbr; URL 4, http://www.snp-star.eu; URL 5, http://rgd.mcw.edu/gbreport/gbrowser_error_conflicts.shtml; URL 6, http://www.anim.med.kyoto-u.ac.jp/nbr/RI_SNPs.html; URL 7, http://www.ensembl.org/; URL 8, http://gscan.well.ox.ac.uk/rats/rat.snp.selector.cgi; URL 9, http://www.well.ox.ac.uk/rat_mapping_resources/SNPbased_maps.html; URL 10, http://www.genenetwork.org/; URL 11, http://bg.upf.edu/funcSTAR/; URL 12, http://www.brc.riken.jp/lab/dna/en/index.html; URL 13, http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Accession codes. GenBank nucleotide: AAXN01000001–AAXN01072867, AAXL01000001–AAXL01031928, AAXP01000001–AAXP01073497, AAXM01000001–AAXM01023012. DNA Databank of Japan: DH508174–DH839445. dbSNP Submitter Method: RAT_STRAIN-READS_SNPS_200712, RAT_STRAIN-GENOTYPES_200712.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by European Union grants LSGH-2004-005235 and LSHG-CT-2005-019015. We acknowledge funding from the National Genome Research Network of the German Ministry of Science and Education. We thank all of the technical staff of the Sequencing Technology Team at the RIKEN Genomic Sciences Center for their assistance. Part of this work was supported by the National BioResource Project of the Ministry of Education, Culture, Sports, Science and Technology of Japan. D.G. is supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Science (057733/Z/99/A). M.-T.B. and D.G. acknowledge support from the Wellcome Cardiovascular Functional Genomics Initiative (066780/Z/01/Z). M. Pravenec is supported by the Howard Hughes Medical Institute and by the Grant Agency of the Czech Republic. M. Pravenec and V.K. are supported by grants from the Ministry of Education of the Czech Republic.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Jacob, H.J. & Kwitek, A.E. Rat genetics: attaching physiology and pharmacology to the genome. *Nat. Rev. Genet.* **3**, 33–42 (2002).
2. Bihoreau, M.T. *et al.* A linkage map of the rat genome derived from three F_2 crosses. *Genome Res.* **7**, 434–440 (1997).
3. Guryev, V., Berezikov, E., Malik, R., Plasterk, R.H. & Cuppen, E. Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res.* **14**, 1438–1443 (2004).
4. Zimdahl, H. *et al.* A SNP map of the rat genome generated from cDNA sequences. *Science* **303**, 807 (2004).
5. Thomas, M.A., Chen, C.F., Jensen-Seaman, M.I., Tonellato, P.J. & Twigger, S.N. Phylogenetics of rat inbred strains. *Mamm. Genome* **14**, 61–64 (2003).
6. Kurtz, T.W. & Morris, R.C. Jr. Biological variability in Wistar-Kyoto rats. Implications for research with the spontaneously hypertensive rat. *Hypertension* **10**, 127–131 (1987).
7. Kurtz, T.W., Montano, M., Chan, L. & Kabra, P. Molecular evidence of genetic heterogeneity in Wistar-Kyoto rats: implications for research with the spontaneously hypertensive rat. *Hypertension* **13**, 188–192 (1989).
8. Gauguier, D. The rat as a model physiological system. In *Encyclopedia of Genetics* vol. 3 (eds. Jorde, L.B., Little, P., Dunn, M. & Subramaniam, S.) 1154–1171 (Wiley, London, 2006).
9. Arbiza, L. *et al.* Selective pressures at a codon-level predict deleterious mutations in human disease genes. *J. Mol. Biol.* **358**, 1390–1404 (2006).
10. Goñi, J.R., de la Cruz, X. & Orozco, M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res.* **32**, 354–360 (2004).
11. Hedrich, H.J. (ed.) *Genetic Monitoring of Inbred Strains of Rat* (Gustav Fischer, Stuttgart, New York, 1990).
12. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
13. Mashimo, T. *et al.* A set of highly informative rat simple sequence length polymorphism (SSLP) markers and genetically defined rat strains. *BMC Genet.* **7**, 19 (2006).
14. Smits, B.M. *et al.* Efficient single nucleotide polymorphism discovery in laboratory rat strains using wild rat-derived SNP candidates. *BMC Genomics* **6**, 170 (2005).
15. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
16. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
17. Wade, C.M. *et al.* The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**, 574–578 (2002).
18. Frazer, K.A. *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053 (2007).
19. Yang, H., Bell, T.A., Churchill, G.A. & Pardo-Manuel de Villena, F. On the subspecific origin of the laboratory mouse. *Nat. Genet.* **39**, 1100–1107 (2007).
20. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
21. Guryev, V. *et al.* Haplotype block structure is conserved across mammals. *PLoS Genet.* **2**, e121 (2006).
22. Jensen-Seaman, M.I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14**, 528–538 (2004).
23. Grupe, A. *et al.* In silico mapping of complex disease-related traits in mice. *Science* **292**, 1915–1918 (2001).
24. Payseur, B.A. & Place, M. Prospects for association mapping in classical inbred mouse strains. *Genetics* **175**, 1999–2008 (2007).
25. Gauguier, D. *et al.* Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. *Nat. Genet.* **12**, 38–43 (1996).
26. Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* **37**, 243–253 (2005).
27. Dumas, M.E. *et al.* Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. *Nat. Genet.* **39**, 666–672 (2007).

28. Mashimo, T., Voigt, B., Kuramoto, T. & Serikawa, T. Rat Phenome Project: the untapped potential of existing rat strains. *J. Appl. Physiol.* **98**, 371–379 (2005).
29. Ihaka, R. & Gentleman, R.R. A language for data analysis and graphics. *J. Comput. Graph. Statist.* **5**, 299–314 (1996).
30. Broman, K.W. The genomes of recombinant inbred lines. *Genetics* **169**, 1133–1146 (2005).
31. Shisa, H. *et al.* The LEXF: a new set of rat recombinant inbred strains between LE/Stm and F344. *Mamm. Genome* **8**, 324–327 (1997).
32. Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).
33. Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
34. Oliphant, A., Barker, D.L., Stuelpnagel, J.R. & Chee, M.S. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* **32** (suppl.), 56–58, 60–61 (2002).
35. Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* **21**, 673–678 (2003).
36. Hardenbol, P. *et al.* Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* **15**, 269–275 (2005).
37. Vlieghe, D. *et al.* A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* **34**, D95–D97 (2006).
38. Blanco, E., Messeguer, X., Smith, T.F. & Guigo, R. Transcription factor map alignment of promoter regions. *PLOS Comput. Biol.* **2**, e49 (2006).
39. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
40. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).

The complete list of authors is as follows:

The STAR Consortium: Kathrin Saar[1], Alfred Beck[2], Marie-Thérèse Bihoreau[3], Ewan Birney[4], Denise Brocklebank[3], Yuan Chen[4], Edwin Cuppen[5], Stephanie Demonchy[6], Joaquin Dopazo[7], Paul Flicek[4], Mario Foglio[6], Asao Fujiyama[8,9], Ivo G Gut[6], Dominique Gauguier[3], Roderic Guigo[10,11], Victor Guryev[5], Matthias Heinig[1], Oliver Hummel[1], Niels Jahn[12], Sven Klages[2], Vladimir Kren[13,14], Michael Kube[2], Heiner Kuhl[2], Takashi Kuramoto[15], Yoko Kuroki[8], Doris Lechner[6], Young-Ae Lee[1,16], Nuria Lopez-Bigas[10,11], G Mark Lathrop[6], Tomoji Mashimo[15], Ignacio Medina[7], Richard Mott[3], Giannino Patone[1], Jeanne-Antide Perrier-Cornet[6], Matthias Platzer[12], Michal Pravenec[13,14], Richard Reinhardt[2], Yoshiyuki Sakaki[8], Markus Schilhabel[12], Herbert Schulz[1], Tadao Serikawa[15], Medya Shikhagaie[11], Shouji Tatsumoto[8], Stefan Taudien[12], Atsushi Toyoda[8], Birger Voigt[15], Diana Zelenika[6], Heike Zimdahl[1] & Norbert Hubner[1]

Affiliations for participants: [1]Max-Delbrück Center for Molecular Medicine, Robert-Rossle-Straße 10, 13125 Berlin, Germany. [2]Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford, OX3 7BN, UK. [4]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [5]Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [6]Commissariat à l'Énergie Atomique, Institut de Génomique, Centre National de Génotypage, 2 rue Gaston Crémieux CP 5721, 91 057 Evry Cedex, France. [7]Department of Bioinformatics, and Functional Genomics Node, Centro de Investigación Príncipe Felipe, Avenida Autopista del Saler 16, 46012 Valencia, Spain. [8]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [9]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan. [10]Center for Genomic Regulation, C/Dr. Aiguader 88, Barcelona Biomedical Research Park Building, 08003 Barcelona, Catalonia, Spain. [11]Experimental and Health Science Department, Universitat Pompeu Fabra, C/Dr. Aiguader 88, Barcelona Biomedical Research Park Building, 08003 Barcelona, Catalonia, Spain. [12]Genome Analysis, Leibniz Institute for Age Research – Fritz Lipmann Institute, Beutenbergstraße 11, 07745 Jena, Germany. [13]Institute of Physiology, Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic. [14]Institute of Biology and Medical Genetics, First Medical Faculty, Charles University, Albertov 4, 12800 Prague 2, Czech Republic. [15]Institute of Laboratory Animals, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [16]Pediatric Pneumology and Immunology, Charite, Campus Virchow Klinikum, Augustenburger Platz 1, 13353 Berlin, Germany.

An ENU-induced mutant archive for gene targeting in rats

To the editor:

Although the laboratory rat is increasingly being used as a mammalian model in biomedical research, no technology exists thus far for the production of *in vivo* genetically engineered mutations equivalent to knockout or knock-in mice because of the lack of development of functional embryonic stem cells in this species. Rat spermatogonial stem cells may possibly have greater potential use for genetic engineering to produce gene-targeted rats[1,2]. The use of somatic cell nuclear transfer to develop cloned rats as an alternative to using embryonic stem cells has also been reported[3], but it seems to be a very difficult technique to perform. At present, *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis followed by a screening method to detect single-nucleotide substitutions within the targeted gene seems to be the most promising technology in rats, as previously reported by several groups[4,5,6]. The screening protocol, however, either a yeast-based screening assay[4,5] or a high-throughput resequencing-based screening technique[7], is expensive. In addition, gene-driven ENU mutagenesis approaches in rats are not a very efficient use of resources because most of the rats generated are usually discarded within a few weeks or months after the targeted genes have been screened owing to a lack of reliable sperm conservation and rederivation technologies.

We have developed a new, efficient approach that combines two methods: a high-throughput, low-

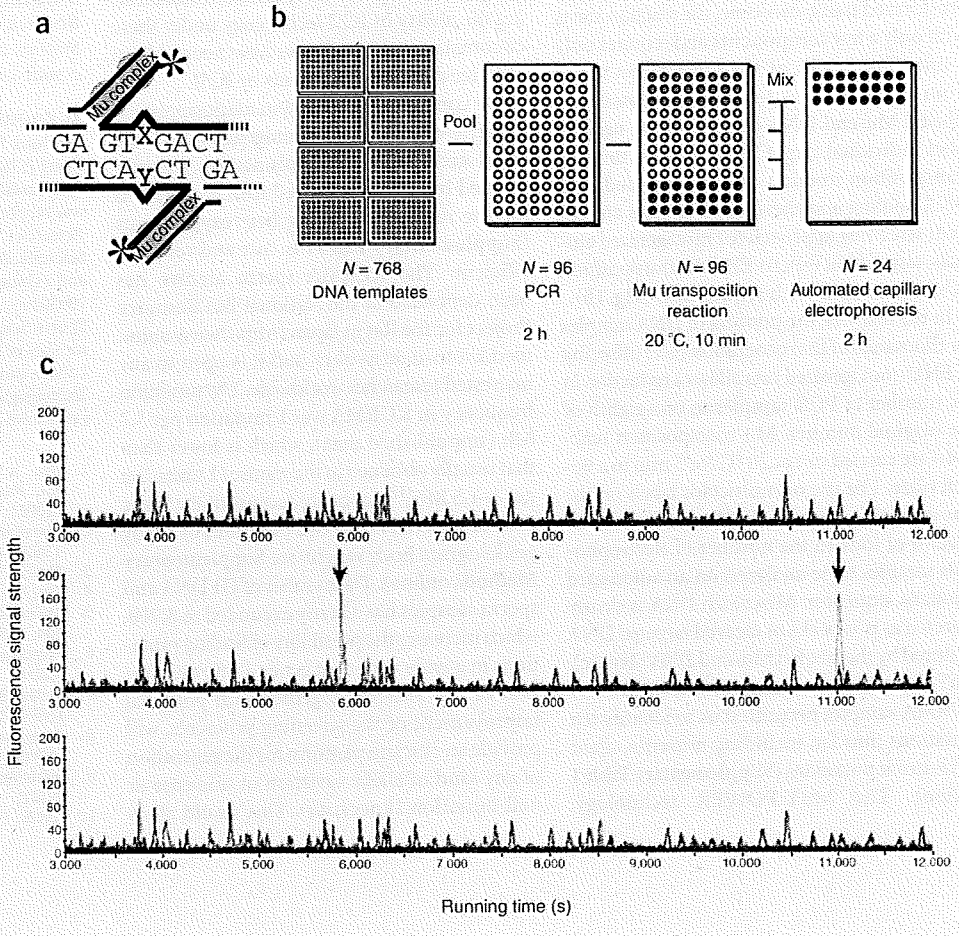

Figure 1 Screening assay for ENU-mutagenized rats. (**a**) Detection of single-nucleotide mismatches by Mu transposition. The Mu transpososome is a complex of Mu-end DNA and MuA transposase (blue circles). The Mu-end DNA fragments are labeled at the 5′ ends (red asterisks). The mismatched DNAs (X and Y) are targeted by the Mu transpososome to generate two reaction products (thick lines). (**b**) Schematic of the MuT-POWER high-throughput DNA screening method for ENU-mutagenized G1 rats. Genomic DNA of G1 rats is extracted and pooled by a factor of eight for PCR. Targeted genomic regions are amplified by PCR for 2–3 h. PCR products are then mixed with Mu transposase labeled with four different fluorophores (blue, green, yellow, red), and incubated at 20 °C for 5 min. Thereafter, Mu reaction products are mixed to form one sample. The resulting reaction mixtures are separated by capillary electrophoresis on an automated sequencer. (**c**) Transposition reaction products are easily visualized as two positive peaks by Gene-Mapper software. The mutations are then confirmed by sequencing each DNA from the eight G1 rats in the pool.

cost screening assay that uses the phage Mu transposition reaction and intracytoplasmic sperm injection (ICSI) for the recovery of the rare heterozygous genotypes from our newly generated frozen sperm repository, the Kyoto University Rat Mutant Archive (KURMA) (**Supplementary Fig. 1** online). Phage Mu transposition shows a strong target site preference for all single-nucleotide mismatches[8]. The process involves *in vitro* cleavage of the mismatch target site by the Mu transpososome, which is a complex of Mu-end DNA and MuA transposase, followed by rejoining of the 3′ ends of the Mu-end DNA to the target site (**Fig. 1a**). Despite the unknown mechanism responsible for Mu transposition, this method has several advantages when used for detecting single-nucleotide substitutions. Using the pre-formed transpososome, the reaction starts with only a few components and requires only 5 min. The length of the DNA target varies from a dozen bases to a few kilobases, and labeled Mu-end DNA eliminates the need to label individual target DNA. Finally, this Mu transposition reaction can be combined with DNA pooling and therefore facilitates an efficient screening approach for mutagenized animals, termed MuT-POWER (Mu Transposition POoling method With sequencER) (**Fig. 1b**).

In the analysis, the pooling of DNA samples for thousands of first generation (G1) offspring of ENU-mutagenized rats allows a reduction in the number of PCR samples to one-eighth of the original number. Mu transposition reactions are carried out at 20 °C for 5 min in a 96-well plate. The use of multicolor labels on the Mu-end DNA allows a further reduction, by a factor of four, in the number of electrophoresis samples to be analyzed on an automated capillary sequencer. Mismatch DNA is clearly detected as two peaks of cleaved targeted DNA captured by the labeled Mu-end DNA (**Fig. 1c**). MuT-POWER screening of the 1,735 G1 KURMA samples permitted us to identify ten mutations thus far, including missense, silent and intronic mutations (**Supplementary Table 1** online). The MuT-POWER technology, therefore, is a powerful tool for detecting ENU-induced mutations, owing to its cost effectiveness, time efficiency and ease of application compared to other protocols[4,5,7].

ICSI technology[9] is very useful for producing an archive of cryopreserved ENU-mutagenized sperm that allows the screening of any targeted gene at any time in the future. Our sperm freezing protocol is very simple to perform, even for a large number of G1 samples. The mutant sperm archive is stored in a −80 °C freezer or a liquid nitrogen tank. Possible fertility problems of ENU-mutagenized rats are alleviated by ICSI. A preliminary injection experiment of such frozen sperm into 234 oocytes from F344 females resulted in the successful recovery of 56 live rats (23.9%) (**Supplementary Table 2** and **Supplementary Video 1** online), demonstrating the ability of ICSI technology under these circumstances; thus far, we have been able to rederive all detected mutants by ICSI.

We conclude that our systematic approach of gene-driven ENU mutagenesis and the rapid screening for DNA mismatches, complemented by ICSI of archived frozen rat sperm for the recovery of living heterozygous rats (**Supplementary Fig. 1**) is cost and resource efficient. The KURMA sperm archive has been deposited at the National Bio Resource Project for the Rat in Japan (http://www.anim.med.kyoto-u.ac.jp/nbr), and it is open to any interested researcher worldwide. The mutation frequency in KURMA was 1 mutation per 3.7 Mb (unpublished data), which is lower than that usually obtained in the mouse (1 mutation per 1–1.5 Mb)[10], probably because of the lower dose of injected ENU (two injections of 40 mg per kilogram body weight; see **Supplementary Methods** online). The number of G1 DNA and sperm samples has already expanded to 5,000, which increases the possibility of finding mutations in a wide variety of genes. Future expansion of the archives, together with potential improvement of the screening protocols, will provide an efficient platform for the generation of any kind of allelic variation or missense or null mutation (knockout). This would allow the production of rat models of human diseases, including cancer, hypertension, diabetes and various diseases for which the mouse has proven to be less useful.

Tomoji Mashimo[1], Katsuhiko Yanagihara[2,4], Satoko Tokuda[1], Birger Voigt[1], Akiko Takizawa[1], Reiko Nakajima[2,4], Megumi Kato[3], Masumi Hirabayashi[3], Takashi Kuramoto[1] & Tadao Serikawa[1]

[1]Institute of Laboratory Animals, Kyoto University Graduate School of Medicine, Yoshidakonoe-cho, Sakyo-ku, 606-8501 Kyoto, Japan. [2]Horizontal Medical Research Organization, Kyoto University Graduate School of Medicine, 606-8501 Kyoto, Japan. [3]National Institute for Physiological Sciences, 444-8787 Okazaki, Japan. [4]Present address: Transdisciplinary Research Integration Center, Research Organization of Information and Systems, 411-8540 Mishima, Japan.
e-mail: serikawa@anim.med.kyoto-u.ac.jp

ACKNOWLEDGMENTS
We thank J.L. Guénet for helpful discussion and also thank F. Tagami, T. Yoshida, Y. Kunihiro and S. Nakanishi for experimental support. This study was supported in part by a Grant-in-Aid for Scientific Research (No. 16200029), a Grant-in-Aid for Scientific Research on Priority Areas (No. 18058009) and a Grant-in-Aid for Exploratory Research (No. 19650106) from the Ministry of Education, Culture, Sports, Science and Technology; by a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare; by JST in Research for Promoting Technological Seeds (No. 10-037); and by a research grant from the Japan Epilepsy Research Foundation.

1. Ryu, B.Y., Kubota, H., Avarbock, M.R. & Brinster, R.L. *Proc. Natl. Acad. Sci. USA* **102**, 14302–14307 (2005).
2. Shinohara, T. *et al. Proc. Natl. Acad. Sci. USA* **103**, 13624–13628 (2006).
3. Zhou, Q. *et al. Science* **302**, 1179 (2003).
4. Zan, Y. *et al. Nat. Biotechnol.* **21**, 645–651 (2003).
5. Amos-Landgraf, J.M. *et al. Proc. Natl. Acad. Sci. USA* **104**, 4036–4041 (2007).
6. Homberg, J.R. *et al. Neuroscience* **146**, 1662–1676 (2007).
7. Smits, B.M. *et al. Pharmacogenet. Genomics* **16**, 159–169 (2006).
8. Yanagihara, K. & Mizuuchi, K. *Proc. Natl. Acad. Sci. USA* **99**, 11317–11321 (2002).
9. Hirabayashi, M. *et al. Transgenic Res.* **11**, 221–228 (2002).
10. Coghill, E.L. *et al. Nat. Genet.* **30**, 255–256 (2002).

PERSPECTIVE

nature genetics

Progress and prospects in rat genetics: a community view

Timothy J Aitman[1], John K Critser[2], Edwin Cuppen[3], Anna Dominiczak[4], Xose M Fernandez-Suarez[5], Jonathan Flint[6], Dominique Gauguier[6], Aron M Geurts[7], Michael Gould[8], Peter C Harris[9], Rikard Holmdahl[10], Norbert Hubner[11], Zsuzsanna Izsvák[11], Howard J Jacob[7], Takashi Kuramoto[12], Anne E Kwitek[13], Anna Marrone[1], Tomoji Mashimo[12], Carol Moreno[7], John Mullins[14], Linda Mullins[14], Tomas Olsson[15], Michal Pravenec[16,17], Lela Riley[2], Kathrin Saar[11], Tadao Serikawa[12], James D Shull[18], Claude Szpirer[19], Simon N Twigger[7], Birger Voigt[12] & Kim Worley[20]

The rat is an important system for modeling human disease. Four years ago, the rich 150-year history of rat research was transformed by the sequencing of the rat genome, ushering in an era of exceptional opportunity for identifying genes and pathways underlying disease phenotypes. Genome-wide association studies in human populations have recently provided a direct approach for finding robust genetic associations in common diseases, but identifying the precise genes and their mechanisms of action remains problematic. In the context of significant progress in rat genomic resources over the past decade, we outline achievements in rat gene discovery to date, show how these findings have been translated to human disease, and document an increasing pace of discovery of new disease genes, pathways and mechanisms. Finally, we present a set of principles that justify continuing and strengthening genetic studies in the rat model, and further development of genomic infrastructure for rat research.

[1]*Medical Research Council Clinical Sciences Centre and Imperial College London, Du Cane Road, London W12 0NN, UK.* [2]*University of Missouri, 1600 E. Rollins Road, Columbia, Missouri 65211, USA.* [3]*The Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.* [4]*University of Glasgow, 126 University Place, Glasgow G12 8TA, UK.* [5]*European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK.* [6]*Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.* [7]*Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA.* [8]*University of Wisconsin-Madison, 1400 University Avenue, Madison, Wisconsin 53706, USA.* [9]*Mayo Clinic, 200 First Street, Rochester, Minnesota 55905, USA.* [10]*Karolinska Institutet, 17177 Stockholm, Sweden.* [11]*Max-Delbrück Center for Molecular Medicine, Robert-Rossle-Strasse 10, 13092 Berlin, Germany.* [12]*Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan.* [13]*University of Iowa, 375 Newton Road, Iowa City, Iowa 52242, USA.* [14]*University of Edinburgh, Edinburgh, Scotland EH16 4TJ, UK.* [15]*Karolinska Hospital, Karolinska Institutet, 17176 Stockholm, Sweden.* [16]*Institute of Physiology, Czech Academy of Sciences and Centre for Applied Genomics, Vídeňská 1083, 142 20 Prague 4, Czech Republic.* [17]*Institute of Biology and Medical Genetics, Charles University, Albertov 4, 128 00 Prague 2, Czech Republic.* [18]*Center for Molecular Genetics and Genomics, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA.* [19]*Université libre de Bruxelles, Rue Profs Jeener & Brachet, BE-6041 Gosselies (Charleroi), Belgium.* [20]*Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA.*
e-mail: T.J.A. (t.aitman@csc.mrc.ac.uk) or H.J.J. (jacob@mcw.edu).

Published online 28 April 2008; doi:10.1038/ng.147

The laboratory rat (*Rattus norvegicus*) has been used as an animal model for physiology, pharmacology, toxicology, nutrition, behavior, immunology and neoplasia for over 150 years (reviewed in ref. 1). Because of its size, ease of manipulation and breeding characteristics, it remained the preferred choice for most of these fields throughout the twentieth century, while the mouse became the leading mammal for experimental genetics. However, since the development of the first inbred rat strain by King in 1909 (ref. 2), over 500 inbred rat strains have been developed for a wide range of biochemical and physiological phenotypes and different disease models.

The US National Institutes of Health (NIH), realizing the potential of rat models in understanding basic biology and human health and disease, launched the Rat Genome Project in 1995, followed by the Rat Expressed Sequence Tag (EST) Project in 1997. These two programs, funded by 13 institutes and centers at NIH, produced a variety of basic genomic resources and provided powerful tools to link to and capitalize upon the data and resources in other model systems and humans. In 1999, the US National Heart, Lung and Blood Institute convened a meeting to discuss the opportunities needed to take full advantage of rat models and to define priorities. The meeting recognized the need to build the rat research community and create a rat genome database, and also made four major recommendations, listed in order of priority as shown in **Box 1**.

Since 1999, the community of investigators using rats as genetic models has grown and matured through its coordinated activities and the collective investments leveraged from funding agencies in the European Union, Germany, Great Britain, Japan and the United States. All of the recommendations of the 1999 meeting have been exceeded, with the exception of site-specific gene targeting, which is the ongoing focus of intensive investigation. The resources and infra-

structure generated have been instrumental in accelerating the pace of biological research in the rat, in driving disease gene discovery in the rat model and in translating these discoveries to related diseases in humans.

In this article, we highlight the successes since the 1999 meeting and note where more infrastructure is needed. We then discuss the impact of advances in human genetics for continuing genetic studies in the rat and conclude with a vision for the future of rat research in the immediate post-genome era.

New genome resources

In the last decade, there has been an extraordinary increase in rat genomic resources, including the generation of a comprehensive 7.5× sequence of the Brown Norway (BN) rat genome; an increase from 5,000 genetic markers to millions of SNPs between different rat strains and from 93,000 ESTs to more than 800,000 ESTs and 5,000 annotated rat gene sequences; development of new methodologies for high-throughput genotyping and expression profiling; and construction of extensive database services in the Rat Genome Database (RGD).

Rat genome sequence. The goal of the genomic sequencing project was to develop a high-quality draft of a single strain. The Brown Norway strain (BN/NHsdMcwi) was selected as the reference strain because it is a well characterized inbred strain and is the founder strain for several important genetic panels, including the BXH/HXB recombinant inbred panel[3,4] and the SS.BN and FHH.BN consomic panels[5]. The Brown Norway genome was sequenced with what was, at the time, a new strategy that combined a variety of sequencing and mapping resources (bacterial artificial chromosome (BAC) end sequences, whole genome shotgun (WGS) sequence and BAC fingerprint map) along with a new genome assembly program, Atlas. These combined approaches resulted in a draft sequence of the rat genome in which individual contigs are mostly of finished quality[6]. The draft sequence enabled the first detailed three-way sequence comparison between mouse, rat and human, providing new information regarding gene and genome evolution[7].

Since the initial publication, the genome sequence has been upgraded once by replacing the draft with the available finished sequence from targeted BACs (~55 Mb). By using newer versions of the Atlas assembly software developed since the original assembly[8] and combining unique sequence data from a WGS-only assembly with the data from the WGS plus BAC assembly, including additional finished BACs, a new assembly of the rat genome will be completed in 2008.

Given the new and increasingly inexpensive sequencing technologies available and the need in the community to sequence other commonly used inbred strains, further shotgun sequencing of more strains at the 2× to 4× level would provide a substantial increase in the number of known SNPs and copy number variants (CNVs). A pilot 2× sequencing of the spontaneously hypertensive rat (SHR) genome, using massively parallel paired-end sequencing, is at present being carried out by European investigators in the European Union–funded EURATools consortium. When combined with other techniques such as optical mapping[9], and applied to other commonly used strains, these resources will significantly enhance and further accelerate progress of genetics research in the rat.

Genetic markers, linkage maps and microarray resources. Since the publication of the first microsatellite maps of the rat[10,11], the collection of genetic markers has progressively increased[12–15]. In this issue, the STAR consortium reports an additional survey of genetic variation, based on three million SNPs identified from a variety of sources and strains[16]. This consortium also selected a subset of 20,000 SNPs that were used to genotype a panel of over 300 inbred strains and hybrid animals. This has led to the creation of a large dataset of well characterized SNPs for quantitative trait locus (QTL) and disease gene mapping, and has allowed the construction of new high density genetic maps. As discussed in the accompanying manuscript[16], the data further characterize the population structure, illustrate the degree of linkage disequilibrium and present an initial haplotype structure of the rat genome. These community resources are openly accessible and significantly augment the tools available for genetic studies of physiological and disease phenotypes in the rat. While the SNP discovery program is continuing within the United States and within the European Union EURATools consortium, the studies also demonstrate the clear need to extend further the collection of SNPs and markers for saturation genotyping and generation of a detailed haplotype map of the rat genome.

In a similar vein to advances with genetic maps, a recently updated cytogenetic map was constructed from G-banded prometaphase chromosomes. This map contains 535 individual bands, considerably improving the previous ideogram and enabling its better use for interpretation of biological data[17].

Commercial rat microarray platforms have been built by Affymetrix, Agilent and Illumina that query expression of up to 31,000 genes at the whole transcript level, or 850,000 exon clusters on the Affymetrix exon arrays. Agilent and NimbleGen have built tiling microarrays for comparative genome hybridization or expression analysis, with up to 350,000 sequences represented, and commercial platforms for rat microRNA analysis have been developed by Ambion, Solexa and Exiqon. Upgrades to these technology platforms will be needed if the benefits of improvements in the genome assembly and annotation, and sequences from additional strains, are to be realized fully.

Database development. RGD (http://rgd.mcw.edu) was established in September 1999 to focus on the rat genome and its resources in conjunction with the principal genome databases: the US National Center for Biotechnology Information (NCBI); University of California, Santa Cruz (UCSC); Ensembl; and Mouse Genome Informatics (MGI). RGD supports the rat research community through comprehensive curation and nomenclature assignment for the complete rat gene catalog (position, protein identifiers, ontologies for function and pathways, links to expression and phenotypes), microsatellite markers, QTLs and rat strains. RGD also provides tools for data mining, analysis and presentation. As described in the accompanying paper by Twigger *et al.*[18], all rat QTLs and associated QTLs from mouse and human—including

BOX 1 Recommendations of the 1999 Rat Genome Priorities meeting

1. **Germ-line modifications in the rat:** to develop new technologies for site-specific gene targeting and nuclear transfer, cryopreservation of zygotes and sperm, and generation of *in vitro* fertilization techniques, including intracytoplasmic sperm injection.
2. **Additional genomic resources:** to include expanding the Rat EST Project, development of a SNP map, increasing BAC clone resources, pilot genome sequencing, and development of a Rat Genome Database.
3. **National Rat Genetic Resource Center:** to maintain and distribute current and future rat models as recommended in the NIH Rat Model Repository Workshop report.
4. **Interaction with the NIH Mouse Mutagenesis and Phenotyping Program:** to yield significant interaction and collaboration on developing new rodent models of human disease.