Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis

Bart M.G. Smits[a], Josine B. Mudde[a], Jose van de Belt[a], Mark Verheul[a], Jocelien Olivier[b], Judith Homberg[a], Victor Guryev[a], Alexander R. Cools[b], Bart A. Ellenbroek[b], Ronald H.A. Plasterk[a] and Edwin Cuppen[a]

Objective **The rat is one of the most important model organisms for biomedical and pharmacological research. However, the generation of novel models for studying specific aspects of human diseases largely depends on selection for specific traits using existing rat strains, thereby solely depending on naturally occurring variation. This study aims to provide the tools to manipulate the rat genome in a more directed way.**

Methods **We developed robust, automated, and scaleable reverse genetic methodology based on ENU (*N*-ethyl-*N*-nitrosourea)-driven target-selected mutagenesis. Optimal mutagenesis conditions have been determined in three different rat strains and a universal, rapid, and cost-effective dideoxy resequencing-based screening setup was established for mutation discovery. The effectiveness of the approach is illustrated by the identification of 120 induced mutations in a set of genes of interest, including six that result in unique rat knockout models due to the introduction of premature stop codons. In addition, 56 mutations were found that change amino acids, including critical residues in transmembrane domains of receptors and channels.**

Conclusions **The approach described here allows for the systematic generation of knockout and protein function altering alleles in the rat. The resulting induced rat models will be powerful tools for studying many aspects of a wide variety of human diseases.**
Pharmacogenetics and Genomics **16:159–169**

Pharmacogenetics and Genomics 2006, **16**:159–169

Keywords: rat disease models, ENU mutagenesis, mutation discovery by high-throughput resequencing, rat knockout, missense mutations

[a]Hubrecht Laboratory, Utrecht, The Netherlands and [b]Department of Psychoneuropharmacology, Radboud University, Nijmegen, The Netherlands

Correspondence and requests for reprints to Dr E. Cuppen, Hubrecht Laboratory, Functional Genomics Group, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
Tel: +31 30 2121969; fax: +31 30 2516554;
e-mail: ecuppen@niob.knaw.nl

Sponsorship: This work was supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics (IGE01017) and NV Organon.

Received 16 August 2005 **Accepted** 27 September 2005

Introduction

The laboratory rat is an important experimental animal model in human medical research, including neural regeneration, psychiatric disorders, behavioral intervention, and addiction, as well as in drug development [1,2]. To mimic neurological disorders in rats, researchers primarily depend on induction by pharmacological agents or surgery [3]. Current genetic rat models for specific biomedical relevant traits originate from selective inbreeding programs. Although hundreds of useful rat models for many aspects of a variety of human disease have been identified, for most of them the underlying gene defect or polymorphism is still unknown. In addition, the existence of a model for a specific trait depends strongly on the degree of natural occurring variation in the strains that are used. In the mouse, this limitation was addressed by chemical mutagenesis using the supermutagen *N*-ethyl-*N*-nitrosourea (ENU) [4]. Many years of exploration on efficiency of ENU as a mutagen have preceded the initiation of systematic large-scale phenotype-driven mouse ENU-mutagenesis screens [5,6] that have now resulted in a wide variety of valuable novel mouse models [7].

An alternative, gene-driven, approach to study gene function and generate disease models is by modifying the function of a gene in a targeted way, i.e. by overexpression or inactivation (knockout). Although current gene-modification and knockout tools for the mouse are extremely versatile and flexible [7], they do depend on homologous recombination in special, pluripotent embryonic stem cells, which, despite many efforts, have not been identified for most of the commonly used model organisms, including the rat. Interestingly, chemical mutagenesis has been used successfully to overcome this limitation in various organisms. ENU or EMS (ethylmethane sulfonate) mutagenesis is used to introduce random mutations, followed by a gene-based strategy for the identification of inactivating mutations, e.g. using CEL I-based enzymatic

The appendix for this paper is available online at the Pharmacogenitics and Genomics website (www.pharmacogeneticsandgenomics.com)

1744-6872

heteroduplex cleavage [8], denaturing high-performance liquid chromatography (DHPLC)-based heteroduplex detection [9,10], or by dideoxy resequencing [11,12]. These approaches have been proven successful for a variety of species, including *Caenorhabditis elegans* [13], *Arabidopsis* [14], zebrafish [15], mouse [10], maize [16], *Lotus* [17], and most recently for the rat [18,19]. However, the efficiency of this approach primarily depends on the effectiveness of the mutagenesis. For mouse, the ENU mutation frequency was found to be strongly dose and strain dependent [20] and estimated to be roughly one mutation per one million basepairs, based on the analysis of 38×10^6 bp by DHPLC [10,21], and 10×10^6 bp by resequencing [12,22]. For rat, only limited information on the molecular mutation frequency is available. Gould and coworkers reported the first ENU-induced rat knockouts [18], but an estimation of the molecular mutation frequency and the associated effectiveness of target-selected mutagenesis is difficult to deduce since the mutation discovery strategy, a yeast-based truncation assay, only detects nonsense mutations. However, the rate of inherited phenotypic mutants among the F1 animals was found to be strain and dose dependent and suggested a similar mutation frequency as for the mouse. Indeed, in a small-scale study, we independently provided a proof of principle for ENU mutagenesis-based reverse genetics in the rat and showed that the molecular mutation frequency varied between 1 mutation per 0.6 and 4×10^6 bp, for a single rat strain tested [19].

However, to convert this approach into routine technology, two critical issues have to be addressed. First, the effectiveness of the approach depends very strongly on the mutagenesis efficiency. Therefore, we have determined the optimal ENU-induced molecular mutation frequency and spectrum in three different commonly used rat strains, for the first time providing solid numbers that can be used as a basis for calculating the size of a mutant population that is needed for the generation of knockout. Second, the number of knockouts that can be retrieved from a population is limited by the speed of screening. Here, we present automated scaleable technology that is both rapid and cost-effective for high-throughput mutation discovery. Most importantly, however, the effectiveness of the resulting setup is illustrated by the identification 120 ENU-induced mutations, resulting in six rat knockouts caused by the introduction of premature stop codons, as well as 56 potentially valuable amino acid replacements caused by ENU-induced missense mutations.

Methods

Animals and ENU mutagenesis protocol

Rats from three inbred strains, BN/RijHsd, F344/NHsd, and LEW/HanHsd, and one outbred strain, Wistar/Crl were used. Animal experiments were performed in accordance with national and local rules and ethical guidelines. Male animals of 11 weeks of age were given three weekly intraperitoneal injections of ENU. The inbred strains were given 3×20, 30 and 40 mg of ENU/kg bodyweight. The Wistar strain received 3×30, 35 and 40 mg of ENU/kg bodyweight.

Preparation of ENU (Isopac; Sigma, Poole, UK) was done within 1 h prior to injections. One gram of ENU was dissolved in 5 ml 96% ethanol. After dissolving the powder by vigorous shaking, 95 ml of phosphate citrate buffer (0.1 M NaH_2PO_4, 0.05 M citric acid, pH 5.0) was added. The concentration was determined by measuring the optical density (OD) of a $10 \times$ dilution at wavelength of 395 nm and the assumption that 1 OD unit equals a concentration of approximately 1 mg/ml. Final concentrations of the ENU stock typically varied between 6 and 8.5 mg/ml, depending on the batch number.

To monitor fertility after treatment, the injected males were paired with one or two females starting 3 weeks after the last injection. Progeny from these early matings was not analyzed, but counted for fertility measurements. Ten weeks after the last injection, fertile males of the highest-dosed fertile groups were kept on a 3-weekly breeding scheme with two females to produce F1 progeny for mutation analysis.

Tissue sampling and genomic DNA isolation

From 2639 F1 progeny, a tail clip was fetched under isoflurane anesthetics at three to four weeks of age. Tail clips were sampled in a 96 deep well block (2.5 ml Riplate, Ritter) and lysed overnight in 400 µl of lysis buffer, containing 100 mM Tris (pH 8.5), 200 mM of NaCl, 0.2% of sodium dodecyl sulfate (SDS), 5 mM of ethylene diamine tetraacetic acid (EDTA), and 100 µg/ml of freshly added Proteinase K at 55°C. Tissue debris was spun down for 20 min at 6000 *g* and supernatant was transferred to a fresh block. DNA was precipitated by adding an equal volume of isopropanol, mixing and centrifugation for 20 min, 6000 *g* at room temperature. The supernatant was removed by gently inverting the block and the pellets were washed with 70% ethanol and dissolved in 400 µl water. For PCR, 5 µl of a $50 \times$ dilution of the DNA stock was used as template.

Nested PCR conditions

The first PCR was carried out using a touchdown thermocycling program (92°C for 60 s; 12 cycles of 92°C for 20 s, 65°C for 20 s with a decrement of 0.6°C per cycle, 72°C for 30 s; followed by 20 cycles of 92°C for 20 s, 58°C for 20 s and 72°C for 30 s; 72°C for 180 s; GeneAmp9700, Applied Biosystems, Foster City, California, USA). The multiplex PCR reaction using eight primer combinations, contained 5 µl genomic DNA, 0.2 µM

of each forward primer and 0.2 μM of each reverse primer, 400 μM of each dinucleotide triphosphate (dNTP), 25 mM Tricine, 7.0% glycerol (w/v), 1.6% dimethyl sulfoxide (DMSO, w/v), 2 mM $MgCl_2$, 85 mM ammonium acetate pH 8.7 and 0.2 U Taq Polymerase in a total volume of 10 μl. After thermocycling, the PCR1 reactions were diluted with 25 μl water, mixed by pipetting, and 1 μl was used as template for the second round of PCR. The second PCR was done using a standard thermocycling program (92°C for 60 s; 30 cycles of 92°C for 20 s, 58°C for 20 s and 72°C for 30 s; 72°C for 180 s; GeneAmp9700, Applied Biosystems). PCR2 mixes contained 1 μl diluted PCR1 template, 0.1 μM forward primer, 0.1 μM reverse primer, 100 μM of each dNTP, 25 mM Tricine, 7.0% glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM ammonium acetate pH 8.7 and 0.1 U Taq polymerase in a total volume of 5 μl. Several samples of each amplicon were tested on a 1% agarose gel containing ethidium bromide for the presence of the proper PCR fragment. The sequence of the universal tails and M13 oligonucleotides are, M13F: TGTAAAACGACGGCCAGT, M13R: AGGAAACAGCTATGACCAT.

Sequencing reactions, purification, and analysis

PCR2 products were diluted with 20 μl water and 1 μl was directly used as template for the sequencing reactions. Sequencing reactions, containing 0.12 μl BigDYE (v3.1; Applied Biosystems), 1.88 μl 2.5 × dilution buffer (Applied Biosystems) and 0.4 μM universal M13 primer in a total volume of 5 μl, were performed using cycling conditions recommended by the manufacturer (40 cycles of 92°C for 10 s, 50°C for 5 s and 60°C for 120 s). Sequencing products were purified by ethanol precipitation in the presence of 40 mM sodium-acetate and analyzed on a 96-capillary 3730XL DNA analyzer (Applied Biosystems), using the standard RapidSeq protocol on 36-cm array. Sequences were analyzed for presence of heterozygous mutations using PolyPhred [23] and manual inspection of the mutated positions.

Automation

All PCR and sequencing reactions were set up on a Tecan Genesis RSP200 liquid handling workstation, with a robotic and an eight-channel pipetting arm, an integrated 96-channel pipetting head (TEMO96, Tecan), and four integrated dual-384 well PCR blocks (Applied Biosystems). This setup allows the production of up to 25 000 PCRs in an unattended overnight run.

Project management and primer design using LIMSTILL

All resequencing projects were managed using LIMSTILL, LIMS for Induced Mutations by Sequencing and TILLing (V.G., E.C., unpublished). This web-based publicly accessible information system (http://limstill.niob.knaw.nl) was used to generate projects and visualize gene structures based on Ensembl genome data, the design of PCR primers, entry, archiving, and primary interpretation of mutations. The primer design application within LIMSTILL is Primer3-based and parameters are set to design primers with an optimal melting temperature of 58°C.

SIFT and PolyPhen

For predictions on the effect of amino acid changing mutations, the stand-alone versions of two prediction programs were used: SIFT (v2.0; Sorting Intolerant From Tolerant) [24]; and PolyPhen (command-line version) [25]; in combination with the SwissProt/TrEMBL + PIR databases (downloaded October 25, 2004 from ftp.expasy.org), and BLAST parameters with an expectation cut-off of 1E-04. We sorted the outcome of the two analyses in three categories: category 1 is 'Affected' by SIFT and 'Possibly/Probably Damaging' by Polyphen; category 2 is 'Tolerated' by SIFT and 'Possibly/Probably Damaging' by Polyphen or 'Affected' by SIFT and 'Benign' by Polyphen; category 3 is 'Tolerated' by SIFT and 'Benign' by Polyphen.

Knockout frequency calculation

Rat Ensembl Build 29.3f was downloaded from ftp.ensembl.org. The total number of genes in the current annotation of the rat genome is 22 155 (23 751 Ensembl predicted genes – 1592 pseudogenes – 4 not analyzed transcripts). For every gene the lengthiest transcript was used to calculate the total number of coding nucleotides, which adds up to 28 402 044 bps. The most frequent ENU mutations (AT-TA 36% + AT-GC 37% + GC-AT 11% = 84%) were used to calculate the amount of bases that could be mutated into a stop codon, which was 1 696 035 bps (6% of the coding capacity) or 76.6 per average-sized (approx. 1300 bp) gene. The probability (P) to identify a knockout for any average-sized gene is calculated with the following formula: $P = 1 - (1 - f)^n$, where n is the number of animals and f is the knockout mutation frequency per gene (= 76.6/mutation rate).

Results and discussion

Optimization rat ENU-mutagenesis conditions

Efficiency of the target-selected mutagenesis approach strongly depends on the mutation rate. In mouse, the ENU-induced acute toxicity, long-term sterility, and mutation frequency was found to be strain and dose dependent. In addition, a regime of three weekly low-dose injections turned out to be more effective compared to a single high-dose injection [20]. Our previous experiments showed that doses of 3 × 60 mg/kg resulted in complete sterility for all strains tested (not shown) and a small-scale study indicated that the optimal dose for the Wistar strain would be between 3 × 30 and 40 mg/kg [19]. In line with these observations, we now mutagenized males from three commonly used inbred strains, Brown Norway (BN), Lewis (LEW), and Fisher (F344) with 3 × 20, 30 and 40 mg ENU/kg, as well as one outbred strain,

Table 1 Statistics of ENU mutagenesis and subsequent progeny production

Strain/dose[a]	BN 3 × 20	BN 3 × 30	BN 3 × 40	F344 3 × 20	F344 3 × 30	F344 3 × 40	F344 3 × 60[b]	LEW 3 × 20	WI 3 × 30	WI 3 × 35	WI 3 × 40	WI 3 × 60[b]	Total
# Males injected	24	24	26	20	18	18	5	18	10	10	10	10	193
# Fertile males	13	7	0	16	12	9	0	0	8	10	6	0	81
# Pups for screening[c]	187	119	0	n.d.	742	472	0	0	n.d.	757[d]	362[d]	0	2639
Average Pups/male	14 ± 7	17 ± 6			74 ± 18	43 ± 21				76 ± 23	60 ± 23		
# Bases screened	9.01×10^6	5.83×10^6			41.06×10^6	24.64×10^6				76.41×10^6	37.27×10^6		194.3×10^6
# Mutations-nonsense	1	0			0	0				5	0		6
missense	3	0			5	10				21	17		56
silent	2	1			3	1				11	7		24
non-coding	2	1			7	3				14	6		34
total	8	2			15	14				51	30		120
Mol. mut. freq.	8.81×10^{-7}	3.43×10^{-7}			3.65×10^{-7}	5.68×10^{-7}				6.67×10^{-7}	8.05×10^{-7}		
Mutation rates	1 in 1.13×10^6	1 in 2.91×10^6			1 in 2.74×10^6	1 in 1.76×10^6				1 in 1.50×10^6	1 in 1.24×10^6		

[a]BN, Brown Norway; F344, Fisher; WI, Wistar; Dose is indicated as mg of ENU per kg bodyweight (three weekly injections).
[b]F344 3 × 60 and WI 3 × 60 were adapted from our previous study [19].
[c]For BN, all fertile males were selected for F1 pup production for screening. For F344 and WI, only the males from the highest-dosed fertile groups were selected F1 pup production.
[d]The WI animals were used only 8 months for F1 pup production, the F344 and BN males were used for 1 year.

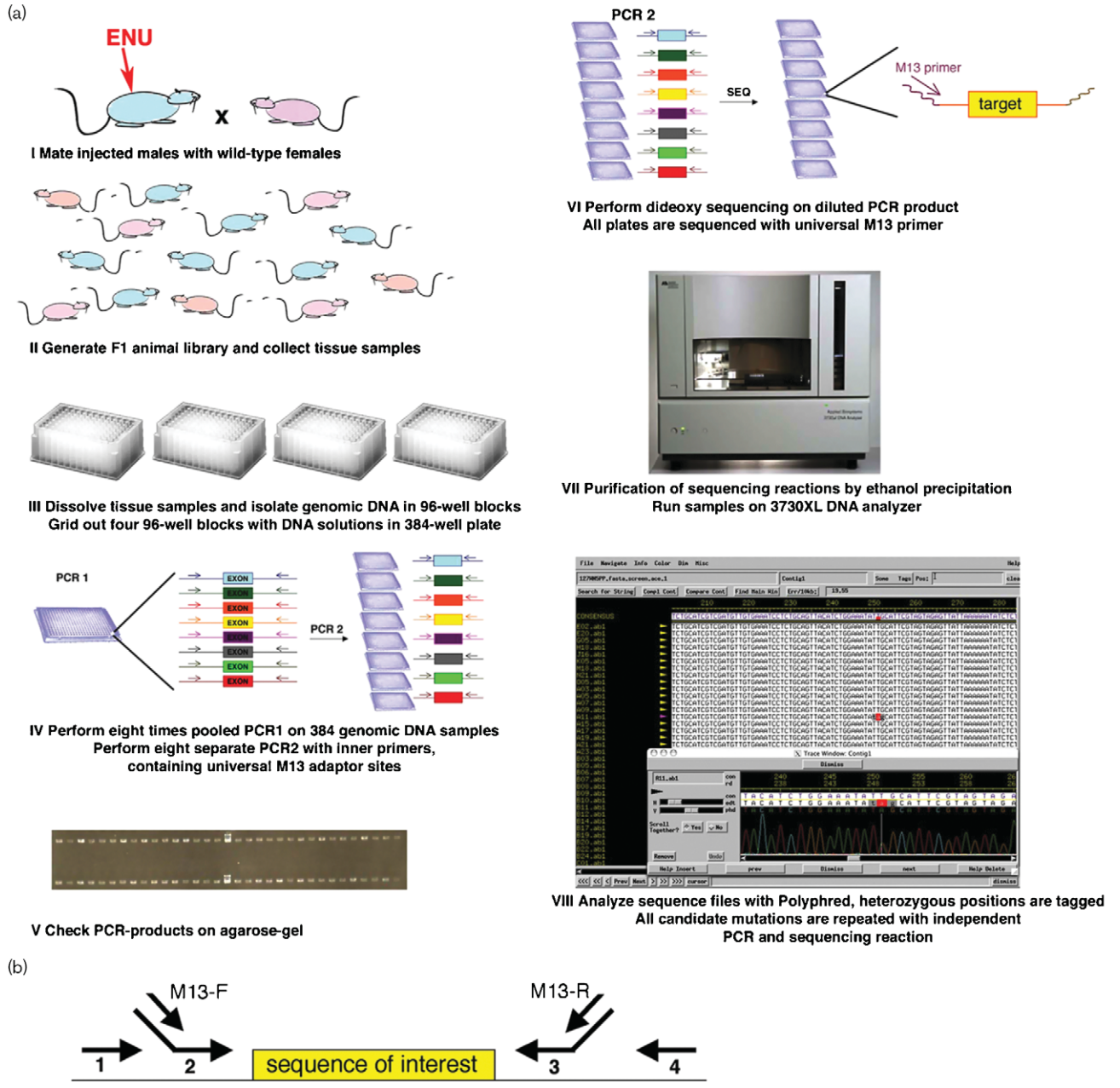
Wistar (WI) with 3 × 30, 35 and 40 mg ENU/kg. Remarkably, all three doses caused permanent sterility in LEW, and this strain was therefore excluded from further analysis. For the other strains, fertility was measured by up to four matings during 12 weeks (WI) or 13 weeks (BN, F344) following the last ENU injections. For BN, all animals in the 3 × 40 mg/kg group ($n = 26$) were sterile and 13 and 7 animals from the 3 × 20 and 3 × 30 mg/kg group ($n = 24$ for each group), respectively, were found to be fertile (Table 1 and Appendix Table 1, available online). For F344 and WI all experimental groups contained fertile animals: 16, 12 and 9 fertile males for F344 groups 3 × 20 ($n = 20$), 30 ($n = 18$), and 40 ($n = 18$), respectively, and 8, 10 and 6 for WI groups 3 × 30, 35 and 40 ($n = 10$ for each group), respectively (Table 1 and Appendix Table 1, available online). Males from the two highest dosed fertile groups for each strain were selected for further breeding. All early progeny was discarded, because these animals could be chimeric resulting from ENU-induced DNA adducts in spermatozoa that are repaired in (part of) the fertilized oocyte. Since we are only interested in the genetically fixed mutations resulting from mutagenized spermatogonial stem cells, we started collecting progeny for mutation analysis at least a full round of spermatogenesis (60 days) after the last ENU treatment. A continuous breeding program, where females were only removed when they were pregnant, was set up for up to one year after ENU mutagenesis for F344 and BN, and eight months for WI, to produce F1 offspring carrying random heterozygous mutations.

For all fertile males the average pup production was calculated and the average pup production per injected group was determined (Table 1). The male founders from the BN groups gave the lowest average pup production, namely only 14 (± 8) and 17 (± 6) pups per fertile founder for groups 3 × 20 and 30, respectively. In contrast, fertile F344 males and WI males, gave much higher average pup productions of 74 (± 18), 43 (± 21), 76 (± 23), and 60 (± 23) for groups F344 3 × 30, F344 3 × 40, WI 3 × 35, and WI 3 × 40 respectively. Regardless of the molecular mutation frequency, the Brown Norway strain may not be the strain of choice, because of its poor breeding capacity.

Mutation discovery by high-throughput resequencing

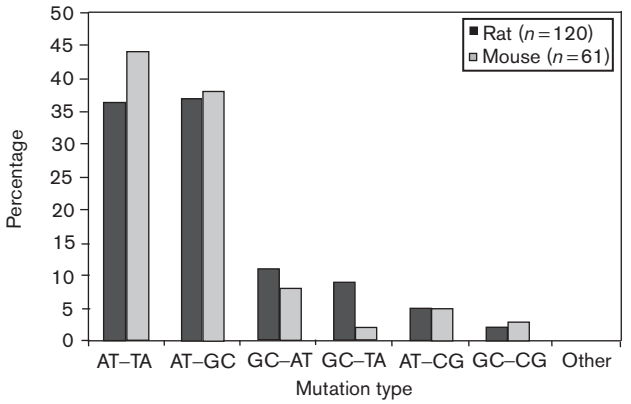
Dideoxy sequencing is generally accepted to be the most reliable and robust technology for mutation discovery [26], although traditionally one of the most costly methods as well. However, we developed a resequencing protocol that is well suited for automation and high-throughput processing and very cost-effective, competing in costs per sample with alternative technologies such as DHPLC, TGCE, SSCP, MALDI-TOF, etc. The setup described here can be operated by only two technicians and has a capacity of up to 70 000 samples

Fig. 1



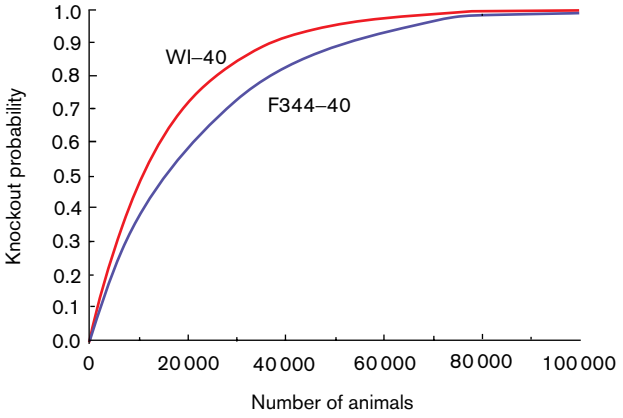
Rat target-selected mutagenesis. (a) Overview of the procedure. Male rats are injected with stem cell mutagen ENU (I), mated with untreated females to generate an F1 population that harbors random heterozygous point mutations in their genomes. From the F1 animals tissue samples are collected (II), genomic DNA is extracted in 96-well blocks (III), gridded out in 384-well plates and PCR is used to specifically amplify regions of interest. The first PCR is done in multiplex, generating eight amplicons in a single reaction (IV). In the second PCR, the multiplex PCR1 products are used as template for eight separated nested PCRs (IV). PCR samples are checked on an agarose gel for the presence of the correct product (V). Cost-effective dideoxy sequencing is performed using universal M13 primer (VI) and subsequent purification of products by precipitation. Samples are analyzed on a 96-capillary 3730XL DNA analyzer (VII) producing sequence files that are inspected using the PolyPhred software package for the presence of heterozygous positions (VIII). Finally, candidate mutations, which are tagged by PolyPhred, are reconfirmed in an independent PCR and sequencing reaction to verify the mutation. (b) Graphical representation of the PCR and sequencing assay design. Gene-specific oligonucleotides 1 and 4 are used in the first multiplexed PCR. The PCR product is diluted and used as template in the second PCR using oligonucleotides 2 and 3 that contain universal M13 tails. The second PCR contains limited amounts of dNTPs and oligonucleotides, eliminating the need for purification of the PCR product before sequencing. Sequencing is performed with diluted PCR product as template in a universal reaction mixture using universal M13 oligonucleotide and minimized amounts of sequencing chemicals.

Fig. 2



ENU-induced mutation spectrum of the rat (this study) and mouse [27]. Only point mutations have been identified in these studies.

Fig. 3



Probability for the identification of a rat knockout using optimized target-selected mutagenesis conditions. The chance to retrieve a knockout for any given average-sized gene (approx. 1300 bp) is plotted as a function of the number of mutant F1 animals for Wistar (WI-40, red line; dose of 3×40 mg/kg, mutation rate of 1 per 1.24×10^6 bp) and Fisher (F344-40, blue line, dose of 3×40 mg/kg, mutation rate of 1 per 1.76×10^6 bp) rats. The data for the BN strain is not shown, as the number of mutations and their distribution over the founder animals is not suited for generalization.

per week. Costs per sample, including all reagents and disposables needed for the PCR and sequencing reactions and sequencing analysis on a capillary sequencer, but excluding costs for personnel and equipment, is currently less than 20 eurocents per sample (sequencing read of 500 bp).

The procedure starts with rapid isolation of genomic DNA from tail clips, followed by a nested PCR for amplification of the region of interest (Fig. 1a). The first PCR is performed in multiplex, followed by dilution and non-multiplexed second PCR reactions. We found that most amplicons (~90%) work in arbitrary combinations of up to eight amplicons. The second PCR reaction is modified in two ways. First, gene-specific primers with universal tail sequences (M13) are used to allow a subsequent standardized sequencing reaction with a universal primer (Fig. 1b). Second, both primer and dNTP concentrations in this reaction are decreased such that only small amounts remain after thermocycling. As a result, reproducible amounts of clean PCR products with only very little sample to sample variation are obtained, allowing immediate cycle sequencing after dilution of the template without the need for additional purification. Furthermore, we found that the combination of the high quality PCR template and the universal sequencing reaction using the M13 adapter allowed us to titrate down the costly sequencing chemicals significantly. The sequencing reactions, which are performed in a volume of only 5 µl, are purified by ethanol precipitation and analyzed on ABI 3730XL DNA analyzers, resulting in 96 sequencing reads (500–600 bp) every 30 min. We used PolyPhred [23] to automatically score potential heterozygous peaks in the sequencing reads, followed by manual inspection of these positions. Every candidate mutation is repeated in an independent PCR and bidirectional sequencing reaction for verification. Initial candidate mutation scoring was deliberately very tolerant to prevent missing rare mutations, resulting in relatively low reconfirmation rates (~7.5%, data not shown).

Selected animals carrying interesting mutations were outcrossed with a wild-type animal to set up a mutant line. Progeny resulting from these crosses was genotyped and in all cases the mutation was found to be genetically inherited (data not shown).

ENU-induced mutation frequency

We have resequenced on average 340 amplicons in 2639 F1 progeny from the three different experimental groups, covering 194×10^6 bases, and identified 120 induced mutations (Table 2). The mutation spectrum nicely overlaps with the ENU-induced mutation spectrum obtained in forward genetic screens in mice [27] (Fig. 2).

The F344 strain displayed the lowest mutation rate: 1 mutation per 2.7 and 1.8×10^6 bp for the 3×30 and 40 mg/kg groups, respectively (Table 1). The mutation rate in the BN strain is much higher for the 3×20 mg/kg group (1 mutation per 1.1×10^6 bp), but unexpectedly lower for the 3×30 mg/kg group (1 per 2.9×10^6 bp). However, the number of animals and bases screened for these groups are relatively small due to the poor breeding characteristics of the mutagenized BN animals. In addition, analysis of the mutation rate per individual founder (Supplementary Table 2) revealed that the

Table 2 Overview of mutations in target genes

Gene	Description	ENSEMBL_ID	Strain	Base change	AA change	Pred. cat.[a]	Alt. Splic.[b]	Domain information[c]
nonsense mutations: 6 (~7% of coding mutations)								
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T10185A	C413X		YES	
MCH	Melanin Conc. Hormone	ENSRNOG00000004632	WI	A1205T	K50X		NO	
MSH6	Mismatch Repair Protein 6	ENSRNOG00000016134	WI	T12645A	L373X		NO	
NNOS	Neuronal Nitric-Oxide Synthase	ENSRNOG00000001130	WI	A28708T	K407X		NO	
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	BN	C3657G	Y62X		YES	
SERT	Serotonin Transporter	ENSRNOG00000003476	WI	C3924A	C209X		NO	
missense mutations: 56 (~64% of coding mutations)								
ADRA2B	Alpha2B Adrenergic Receptor	ENSRNOG00000013887	WI	T1116C	I39T	1	NO	In 1st tmd
AVP1A	Vasopressin V1A Receptor	ENSRNOG00000004400	WI	T1674G	V143G	1	NO	In 3rd tmd
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T8553C	I75T	1	NO	In 1st tmd
CHRND	Acetylcholine Receptor Delta	ENSRNOG00000019527	WI	T6147A	V308E[d]	1	NO	1st residue of 2nd tmd
DBH	Dopamine beta-Hydroxylase	ENSRNOG00000006641	WI	A15984G	Y518C	1	NO	Not annotated
DRD1	Dopamine Receptor 1	ENSRNOG00000023688	F344	T1215A	V72E[d]	1	NO	In 2nd tmd
FSHR	Follicle Stimulating Horm. Rec.	ENSRNOG00000016783	F344	A235725G	D566G	1	NO	Between 3rd and 4th tmd
GPR54	G Protein-coupled Receptor 54	ENSRNOG00000011954	F344	T1801G	V96G	1	NO	In 2nd tmd
IEDA	Orphan Receptor IEDA	ENSRNOG00000014793	BN	A81288T	D679V	1	NO	C-terminal part
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	T90087C	L42P	1	NO	Not annotated
LHR	Luteinizing Hormone Receptor	ENSRNOG00000016712	BN	T37228C	L269P[d]	1	NO	In 1st tmd
MSH6	Mismatch Repair Protein 6	ENSRNOG00000016134	WI	A15933G	N1113D	1	NO	MutS-C
MYOC	Myocilin	ENSRNOG00000003221	WI	T6634A	L124Q	1	NO	Coiled coil
NLG3	Neuroligin 3 Precursor	ENSRNOG00000003812	WI	T8023A	V230D	1	NO	Not annotated
NNOS	Neuronal Nitric-Oxide Synthase	ENSRNOG00000001130	WI	A38127T	I616F	1	NO	NO-synthase
NPY2R	Neuropeptide Y Receptor 2	ENSRNOG00000022004	F344	T1830A	V277E[d]	1	NO	In 6th tmd
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	F344	A3856T	I129F	1	YES	In 3rd tmd
RELN	Reelin Precursor	ENSRNOG00000006665	WI	A20691G	D2814G	1	YES	Not annotated
SLC6A8	Sodium-dep. choline transporter	ENSRNOG00000018857	WI	T6922C	F407L	1	NO	In 8th tmd
SYPH	Synaptophysin	ENSRNOG00000010223	WI	T7519A	C187S	1	NO	Between 3rd and 4th tmd
BDNF	Brain-Derived Neurotrophic Factor	ENSRNOG00000005051	F344	A46980T	H131L	2	NO	Not annotated
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T8592C	V88A	2	NO	Between 1st and 2nd tmd
CHRNA6	Acetylcholine Receptor 6 alpha	ENSRNOG00000012283	WI	T6119G	I432S	2	NO	Between 3rd and 4th tmd
CYSLT2	Cysteinyl Leukotriene Receptor	ENSRNOG00000015042	F344	A740T	Y247F	2	NO	Between 3rd and 4th tmd
GLUR5	Glutamate Receptor 5	ENSRNOG00000001575	WI	A108310G	Y451C	2	NO	N-terminal part
GPR115	G Protein-coupled Receptor	ENSRNOG00000012535	WI	C4780G	T84R	2	NO	N-terminal part
GRPR	Gastrin Releasing Peptide Receptor	ENSRNOG00000004124	F344	T40617C	S289P	2	NO	3 residues behind 6th tmd
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	A299209C	Q511P	2	NO	Not annotated
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	T299400C	Y575H	2	NO	Not annotated
NMN	Neurotensin/Neuromedin N Precur.	ENSRNOG00000004179	WI	T6729A	M77K	2	NO	Not annotated
NMU2R	Neuromedin U Receptor 2	ENSRNOG00000014081	F344	T1512C	V129A	2	NO	In 3rd tmd
NNOS	Neuronal Nitric-Oxyde Synthase	ENSRNOG00000001130	WI	A38049T	S590C	2	NO	NO-synthase
NR0B2	Nuclear Receptor Subfamily 0 B2	ENSRNOG00000007229	F344	T3843A	L202M	2	NO	Hormone_rec_lig
NR1D1	Orphan Nuclear Receptor	ENSRNOG00000009329	F344	T4477A	V178D	2	NO	Znf_C4steroid
NTRK2	TRKB Tyrosine Kinase	ENSRNOG00000018839	WI	C57323A	T340K	2	NO	Between 2 tmd
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	WI	T1095C	S23P	2	YES	N-terminal part

Table 2 (continued)

Gene	Description	ENSEMBL_ID	Strain	Base change	AA change	Pred. cat.[a]	Alt. Splic.[b]	Domain information[c]
PDYN	Prodynorphin Precursor	ENSRNOG00000016036	WI	A2671T	Q188L	2	NO	Not annotated
PINK	Serine/Threonine Kinase PINK1	ENSRNOG00000015385	WI	G12449A	V549M	2	NO	Not annotated
POMT1	Protein O-Mannosyl-Transferase 1	ENSRNOG00000010477	WI	T1825A	L47Qd	2	NO	In 1st tmd
ABCG5	ATP-Binding Cassette, G5	ENSRNOG00000005250	WI	C25706A	T319K	3	YES	N-terminal part
ATRX	Transcriptional Regulator ATRX	ENSRNOG00000002550	WI	C23842T	R271C	3	NO	Not annotated
CHRM5	Muscarinic Acetylcholine Rec. 5	ENSRNOG00000006397	F344	G2300A	V343I	3	NO	Between 5th and 6th tmd
DOPTA	Dopamine Transporter	ENSRNOG00000017302	F344	G1070T	V24L	3	NO	N-terminal part
ESDN	Endoth-/SMC-der. neuropilin-like	ENSRNOG00000001651	WI	A119852G	M458V	3	NO	1 Residue before tmd
GPR115	G Protein-coupled Receptor 115	ENSRNOG00000012535	BN	A8746T	L284F	3	NO	N-terminal part
HSP90	Heat-Shock Protein 90	ENSRNOG00000019834	WI	G2412T	V253L	3	NO	Low complexity
HD	Huntington Disease Protein	ENSRNOG00000011073	F344	T116025A	S1264T	3	NO	Not annotated
MSH2	Mismatch Repair Protein 2	ENSRNOG00000015796	F344	A5135G	N105S	3	NO	MutS_N
MYOC	Myocilin	ENSRNOG00000003221	WI	T15966A	Y458F	3	NO	Olfac_like
NLG3	Neuroligin 3 Precursor	ENSRNOG00000003812	WI	A1073T	I25F	3	NO	Carboxylesterase
NPY2R	Neuropeptide Y Receptor 2	ENSRNOG00000022004	WI	T1161C	V54A	3	NO	In 1st tmd
NPY4R	Neuropeptide Y Receptor 4	ENSRNOG00000020282	WI	A1587C	E196A	3	NO	Between 3rd and 4th tmd
OPRM1	MU-Type Opioid Receptor 1	ENSRNOG00000018191	WI	A34267G	I256V	3	NO	Last residue of 5th tmd
P53	p53	ENSRNOG00000010756	WI	C1039G	I10M	3	NO	P53
TH	Tyrosine 2-Monooxygenase	ENSRNOG00000020410	WI	C5502A	D275E	3	NO	Aaa-hydroxylase
GRM8	Metabotrophic Glutamate Rec. 8	NM_022202	WI	C71547T	T341I	n.d.	NO	Between 2nd and 3rd tmd

Silent mutations: 24 (~28% of coding mutations)
Non-coding mutations: 34

Total mutations: 120 (of which 86 in coding regions)

[a]SIFT/Polyphen analysis: cat. 1 is predicted to have functional consequences by both programs; cat. 2 is predicted to have functional consequences for one of the programs; cat. 3 is predicted to have no consequences by either program.
[b]Indicates if the mutation resides in an exon that may be alternatively spliced out.
[c]Indicates the position of the mutation in a domain of the protein; tmd=transmembrane domain.
[d]Indicates five potential severe mutations in transmembrane domains (see text for details).

Table 3 Estimated number of PCR and sequencing reactions needed to obtain a mutation or stopcodon in an exon or gene of interest with 70% or 95% probability

	Exon of interest[a]		Gene of interest[b]	
	PCRs[c]	SEQs[c]	PCRs[c]	SEQs[c]
1 mutation 70%[d]	6514	5790	9135	8120
1 mutation 95%[d]	16 200	14 400	22 603	20 090
1 stopcodon 70%[e]	149 963	133 300	181 916	161 700
1 stopcodon 95%[e]	324 900	288 800	452 816	402 500

[a]Average size of an exon of 250 bp is used for this calculation.
[b]Average size of a gene of 1.25 kbp is used for this calculation. We assume that on average seven amplicons have to be screened to cover an average gene's coding region.
[c]In order to perform one sequencing reaction (SEQ), a nested PCR is required. The first PCR is performed in a pooled fashion with eight amplicons. The amount of PCRs and SEQs are thus related 9:8.
[d]Numbers are based on Wistar mutation frequency (1 in 1.2 million bases).
[e]Based on this study, 1 in 20 mutations in an amplicon turns a codon into a stopcodon (6 of 120 mutations).

apparent high rate for the 3×20 group of BN animals results mainly from the contribution of a single mutagenized male. Finally, the WI strain gives the best overall mutation rates of 1 per 1.5 and 1.2×10^6 bp for groups 3×35 and 40, respectively (Table 1).

Mutation spectrum and characteristics

The amplicons used for determining the molecular mutation frequency were designed to contain mostly exonic sequence. Eighty-six out of the 120 mutations are located in the protein coding sequences of target genes; six result in the introduction of a premature stop codon in the open reading frame (nonsense, ~7%) and are most likely to result in a functional knockout of the gene, 56 change an amino acid (missense, ~65%), and 24 have no effect on the protein coding capacity (silent, ~28%) (Table 2). According to the rat genome annotation (Ensembl v29.3f) two of the nonsense mutations reside in an exon that may be spliced out, resulting in a splice form-specific knockout (Table 2).

We used two programs, SIFT [24] and PolyPhen [25] that use phylogenetic conservation and structural information to predict the potential effect of a mutation on protein function, to analyze the 56 missense mutations. Based on the output of these programs, we defined three categories (Table 2). Sixteen mutations were predicted not to affect protein function by both programs (category 3), 19 are predicted to affect protein function by one of the two programs (category 2), and 20 mutations are predicted to affect protein function by both programs (category 1). These predictions may help selecting primary candidates for further analysis, as only experimental work will be able to give an answer to the question which missense mutations have a biological effect

Interestingly, 38 out of the 56 missense mutations occur in transmembrane domain-containing proteins and 15 mutations do alter an amino acid in a transmembrane domain (Table 2). Residue changes from a hydrophobic to a charged amino acid and the introduction of helix-distorting prolines may affect proper insertion of the protein in the membrane, resulting in inefficient or deficient trafficking to its proper location, or may affect channel properties of specific proteins. Five mutations in our set meet these criteria and four of them were indeed predicted by both SIFT and PolyPhen to affect protein function (Table 2). Together with the six knockouts, these mutations are most likely to have functional consequences for the protein. Currently, we are crossing out these animals and are breeding the mutations to homozygosity. Three of the potentially functional amino acid changes have been crossed to the next generation, including acetylcholine receptor delta subunit (CHRND), protein *O*-mannosyl transferase (POMT1), and neuropeptide Y receptor (NPY2R). Finally, four knockouts have currently been passed on to the next generation, namely the nociceptin receptor OPRL1 (opioid-like receptor-1 or orphanin FQ receptor), the DNA mismatch repair protein MSH6 that is involved in hereditary non-polyposis colorectal cancer (HNPCC), the melanin concentrating hormone MCH, and the serotonin transporter (SERT/SLC6A4). The latter is homozygous viable and has been confirmed to be a functional knockout by biochemical and pharmacological analysis (unpublished results).

Conclusions

We have established robust and scaleable technology to generate knockout and mutant rats using target-selected mutagenesis and have illustrated the effectiveness of the methodology by the identification of 120 induced gene mutations, of which six result in exclusive rat knockout models, due to induced nonsense mutations.

Optimal ENU mutagenesis conditions using four rat strains were determined, resulting in similar mutation spectra [27] and rates (about 1 mutation every 1–1.5×10^6 bp) as obtained for specific mouse strains [21,22]. Furthermore, we have set up an efficient resequencing approach for the identification of rare induced point mutations. The major advantages of this mutation

discovery technology are the robustness and low costs per reaction and the highly informative output. The resulting raw data is very well-suited for automatic bioinformatic analysis, providing detailed sequence information for every mutation per individual animal. The platform is very flexible and scaleable, as it can be implemented for both large-scale projects using automation by robotics, as well as for moderate-scaled experiments using multi-channel pipets. In contrast to previously used technology for the generation of rat knockouts, employing a yeast-based truncation assay [18], resequencing not only detects nonsense mutations, but also potentially interesting missense mutations. In addition, resequencing provides direct information on the actual ENU-induced molecular mutation frequency and spectrum. From the four strains tested, the WI strain was found to be the most robust strain, based on its molecular mutation frequency as well as its good breeding properties and it may therefore be the strain of choice for routine production of knockout rats.

We calculated the probability to identify a specific knockout based on the obtained molecular mutation frequencies (Fig. 3). The total number of coding bases in the current rat genome release is 28.4×10^6 bp, of which 1.7×10^6 bp (6%) could be mutated into a stop codon by the most frequent ENU mutations. Assuming random distribution of these potential ENU-inducible nonsense positions over the 22 155 annotated genes and a mutation rate of 1 per 1.2×10^6 bp for the Wistar strain, the chance to find a knockout for any average-sized (approximately 1250 bp) gene in a library of 5000, 10 000, 20 000 and 50 000 mutant F1 rats would be 27, 46, 71 and 96%, respectively. It should be mentioned that with the current ENU mutation frequencies obtained in model organisms, chances are still extremely low to identify a particular point mutation, e.g. exact homologs of human disease-causing mutations, as any given position in the genome has a likelihood to be hit at a rate equaling the mutation frequency, which would be only 1 in 1.2 million for the rat.

In Table 3, we illustrated how many PCRs and sequencing reactions would be required to obtain a mutation or a knockout in an exon or gene of interest with 70% and 95% probability. Based on the Wistar mutation frequency about 6500 or 16 000 PCRs and 6000 or 14 000 sequencing reactions are needed to find a mutation in a single amplicon (250 bp coding) with 70% or 95% probability, respectively. For finding a premature stopcodon in an average sized gene of 1.25 kbp (assuming seven protein-coding exons) with 95% probability, up to 450 000 PCRs and 400 000 sequencing reactions are needed.

Ideally, a permanent frozen library of mutant F1 animals would be generated, which could be screened indefinitely for mutations in any gene of interest. However, rat sperm cryopreservation is still in its infancy, with currently only a single example of successful rederivation from frozen sperm [28]. Alternatively, batches of living animals could be generated that can be screened in a rolling circle model. To benefit optimally from such a relatively short-lived resource, the mutation screening capacity will be the limiting factor. Although the technology described here is highly scaleable, emerging technologies such as microarray-based directed or random resequencing techniques could potentially boost throughput significantly [29,30]. However, costs associated with these technologies are currently still too high for routine large-scale multi-individual resequencing projects.

Taken together, the approach described here allows routine generation of rat knockouts, opening a whole new field in rat genetic and functional research that is likely to produce many valuable models for studying human health and disease.

Acknowledgements

We thank F. Vervoordeldonk for help with Polyphred analysis. We thank R. Malik for SIFT/PolyPhen analysis. This work was supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics and NV Organon.

References

1 Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004; **428**:493–521.
2 Jacob HJ, Kwitek AE. Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet* 2002; **3**:33–42.
3 Kolb B. Neurological models. In: Whishaw IQ, Kolb B, editors. *The Behavior of the Laboratory Rat.* Oxford: Oxford University Press; 2005. pp. 449–461.
4 Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci USA* 1979; **76**:5818–5819.
5 Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* 2000; **25**:440–443.
6 Hrabe de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, Marschall S, *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 2000; **25**:444–447.
7 Cox RD, Brown SD. Rodent models of genetic disease. *Curr Opin Genet Dev* 2003; **13**:278–283.
8 Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, *et al.* High-throughput screening for induced point mutations. *Plant Physiol* 2001; **126**:480–484.
9 McCallum CM, Comai L, Greene EA, Henikoff S. Targeted screening for induced mutations. *Nat Biotechnol* 2000; **18**:455–457.
10 Coghill EL, Hugill A, Parkinson N, Davison C, Glenister P, Clements S, *et al.* A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet* 2002; **30**:255–256.
11 Wienholds E, Schulte-Merker S, Walderich B, Plasterk RH. Target-selected inactivation of the zebrafish rag1 gene. *Science* 2002; **297**:99–102.
12 Beier DR. Sequence-based analysis of mutagenized mice. *Mamm Genome* 2000; **11**:594–597.
13 Jansen G, Hazendonk E, Thijssen KL, Plasterk RH. Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat Genet* 1997; **17**:119–121.

14 Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, *et al.* Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 2003; **13**:524–530.
15 Wienholds E, vanEeden F, Kosters M, Mudde J, Plasterk RH, Cuppen E. Efficient target-selected mutagenesis in zebrafish. *Genome Res* 2003; **13**:2700–2707.
16 Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, *et al.* Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol* 2004; **4**:12.
17 Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, *et al.* A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. *Plant Physiol* 2003; **131**: 866–871.
18 Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, Wang YR, *et al.* Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol* 2003; **21**:645–651.
19 Smits BM, Mudde J, Plasterk RH, Cuppen E. Target-selected mutagenesis of the rat. *Genomics* 2004; **83**:332–334.
20 Justice MJ, Carpenter DA, Favor J, Neuhauser-Klaus A, Hrabe de Angelis M, Soewarto D, *et al.* Effects of ENU dosage on mouse strains. *Mamm Genome* 2000; **11**:484–488.
21 Quwailid MM, Hugill A, Dear N, Vizor L, Wells S, Horner E, *et al.* A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm Genome* 2004; **15**:585–591.
22 Concepcion D, Seburn KL, Wen G, Frankel WN, Hamilton BA. Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. *Genetics* 2004; **168**:953–959.
23 Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997; **25**:2745–2751.
24 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**:3812–3814.
25 Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30**:3894–3900.
26 Larsen LA, Christiansen M, Vuust J, Andersen PS. Recent developments in high-throughput mutation screening. *Pharmacogenomics* 2001; **2**: 387–399.
27 Noveroske JK, Weber JS, Justice MJ. The mutagenic action of *N*-ethyl-*N*-nitrosourea in the mouse. *Mamm Genome* 2000; **11**:478–483.
28 Nakatsukasa E, Inomata T, Ikeda T, Shino M, Kashiwazaki N. Generation of live rat offspring by intrauterine insemination with epididymal spermatozoa cryopreserved at −196 degrees C. *Reproduction* 2001; **122**:463–467.
29 Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; **21**:20–24.
30 Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433–438.