RAT REVERSE GENETICS:
GENERATION AND CHARACTERIZATION
OF CHEMICALLY INDUCED RAT MUTANTS

Ruben van Boxtel

Een zucht is niet slechts een zucht.
We ademen de wereld in en ademen betekenis uit.
Zolang we kunnen. Zolang we kunnen.
- Salman Rushdie

Voor mijn ouders, Michelle, Godelieve en Farahilde

ISBN: 978-94-90371-25-8

Cover design: Judith van Boxtel
(judith.van.boxtel@live.nl)

Layout & printing: Off Page, Amsterdam

RAT REVERSE GENETICS:
GENERATION AND CHARACTERIZATION
OF CHEMICALLY INDUCED RAT MUTANTS

Gen-gedreven rattengenetica:
Het genereren en karakteriseren
van chemisch geïnduceerde rattenmutanten

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op maandag 23 augustus des middags te 2.30 uur

door

Ruben van Boxtel
geboren op 8 november 1979
te Kaatsheuvel

Promotor: Prof. dr. E. P. J. G. Cuppen

The research described in this thesis was performed at the Hubrecht Institute of the Royal Academy of Arts and Sciences (KNAW), within the framework of the Graduate School of Cancer Genomics and Developmental Biology (CDGB) in Utrecht.

Printing of this thesis was financially supported by the J. E. Jurriaanse Stichting.

CONTENTS

GENERAL INTRODUCTION: RAT GENE KNOCKOUT TECHNOLOGIES AND MODELS

Ruben van Boxtel[1] and Edwin Cuppen[1,2]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[2]Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, Utrecht, The Netherlands

ABSTRACT

Although the rat is one of the most used model organisms in biomedical research, its use as a mammalian genetic model system has long been hampered by the lack of pluripotent embryonic stem (ES) cells. Therefore, alternative techniques independent of ES cell manipulation were developed, like *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis, transposon-tagged mutagenesis and very recently zinc-finger nuclease (ZFN)-mediated gene targeting. These techniques are highly complementary and add to the increasing versatility of the rat as a genetic model system. In addition, successful isolation and propagation of genuine rat pluripotent ES cells and induced pluripotent stem cells (iPSCs) have recently been reported, opening the possibility for homologous recombination (HR)-based gene knockout and knock-in approaches. Here, we provide an overview of the current status and possibilities of these techniques and the application of genetically modified rat models in biomedical research.

UNDERSTANDING THE HUMAN GENOME

Almost a decade ago, the first drafts of the human genome were completed [1,2]. Knowing the sequence, however, does not mean that we understand the code. To crack the code, the use of genetic model organisms is crucial. Sequence conservation between different species is indicative of evolutionary selection and the functional importance of genomic elements. Genetic differences in highly conserved elements can be causative for naturally occurring phenotypic differences, so the availability of genome sequences of different organisms, both far diverged from humans as well as closely related, provides a useful tool to hypothesize about the potential function and importance of genomic elements in the human genome and changes in the DNA, which can be associated with disease. To test these hypotheses, genetic modification in model organisms has proven to be extremely powerful.

In general, technical tools to manipulate the genome can be split in two classes depending on the research question, namely classical phenotype-driven (forward genetic) or genotype-driven (reverse genetic) approaches (**Fig. 1**). To dissect developmental and biochemical pathways that underlie a given phenotype, like disease traits, forward genetic screens are excellent tools. Although naturally occurring genetic variations in humans or model organisms can be used to map phenotypic traits back to the genome, the number

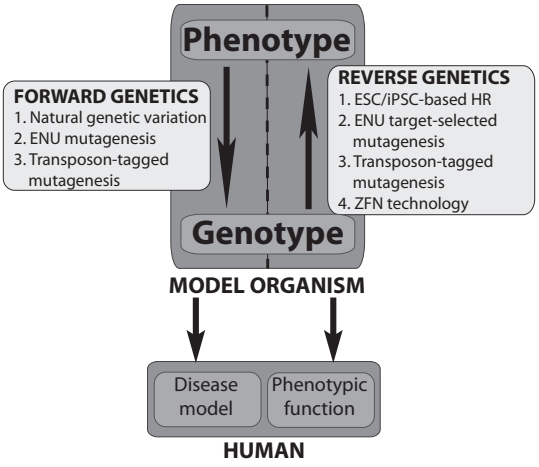

Figure 1: Genetic tools can be subdivided into two groups depending on the research question. 'Classical' or phenotype-driven approaches start off with a specified human disease phenotype. Animals displaying similar symptoms can be used to identify genetic elements underlying these disease traits by selective breeding and molecular biological techniques, like linkage analyses. Both naturally occurring genetic variation as well as artificially induced variation can be used to score disease phenotypes. Alternatively, genotype-driven approaches are based on systematically mutating known genes to determine their role in human physiology and pathology by analyzing the phenotypic effects. ENU, *N*-ethyl-*N*-nitrosourea; ESC, embryonic stem cell; iPSC, induced pluripotent stem cell; HR, homologous recombination; ZFN, zinc-finger nuclease.

of involved genetic elements can vary, making disease-gene discovery extremely complex. Therefore, forward genetic screens are often based on artificially introducing independent genetic variations in the germ line of model organisms using mutagenesis approaches. Hence, every mutant individual most probably carries a single causative genetic change, which can be traced back to the genome using molecular biological techniques. This way single genes involved in the phenotype of interest can be uncovered.

In contrast, genotype-driven approaches are based on manipulating coding genetic elements followed by phenotypic analysis. The availability of completely sequenced genomes of different model organisms has increased the popularity of this approach because prior sequence knowledge is required. Gene knockout technology using homologous recombination combined with pluripotent ES cells in the mouse has proven to be especially powerful for this [3]; however, for many other organisms this technology is not (yet) available. Still, many researchers are not willing to convert to the mouse because different model organisms of choice have different advantages. Therefore, in the last years alternative methods applicable to many other if not all organisms have been developed, which very efficiently allow for mutant generation.

Traditionally, the laboratory rat *Rattus norvegicus* has been the model organism of choice for physiologists, pharmacologists, nutritionists and other biomedical researchers owing to its large size, which facilitates experimental and surgical interventions [4]. For example, it was recently demonstrated that neurons beneath the surface of the brain could be imaged *in vivo* by fluorescent microscopy in a freely moving rat by mounting a miniature two-photon microscope to its head [5]. In addition, learning behavior is traditionally extensively studied in the rat. It was recently shown that neurogenesis and the maturation of these newborn neurons in the adult hippocampus of rats are enhanced when compared with the mouse brain [6]. Moreover, it was shown that these newborn neurons were more involved in response to behavioral activity in rats compared with mice [6]. These data suggest that the rat hippocampus may be a better model for that of the human. Importantly, some human diseases are best mimicked in the rat, in particular neurodegenerative diseases and disorders affecting higher brain function, schizophrenia, anxiety, depression and addiction. Selective breeding and characterization has led to hundreds of strains mimicking complex human disease [7]. To identify causal mutations underlying these disease phenotypes by quantitative trait loci (QTLs) mapping, detailed single nucleotide polymorphism (SNP) panels have been developed, enabling the genotyping of over 300 inbred strains and hybrid animals [8]. Furthermore, to locate genetic loci involved in complex phenotypes, several recombinant inbred panels as well as heterogeneous stocks and many congenic and consomic strains are available [9]. Nevertheless, identifying causative polymorphisms underlying disease phenotypes remains difficult. It is therefore not surprising that a lot of effort has been invested in generating and optimizing techniques to artificially manipulate the rat genome.

Recently, some great technological breakthroughs have been achieved in the field of rat genetics that will certainly have great impact on modeling and understanding

human physiology and disease and will finally make the rat a fully appreciated genetic model organism. Here, we review these technical improvements and newly developed methods. In addition, we will describe the implications of rat genetic mutant models for understanding human biology.

HOMOLOGOUS RECOMBINATION-MEDIATED GENE TARGETING

Maintaining pluripotency of rat ES cells

In the last two decades, 'classical' gene targeting based on homologous recombination (HR) in pluripotent ES cells has been one of the most powerful tools in the field of genetics [3]. For successful gene targeting, it is crucial to be able to maintain a cell type *in vitro* that is ultimately capable of germ line contribution when placed back in its original natural environment. A gene of choice is targeted by offering these cultured cells an artificially engineered piece of DNA, which contains parts of DNA homologous to the target sequence for recombination and a nonhomologous part, such as selection markers, reporter genes, sequence-specific recombinase genes, etc. (**Fig. 2A**). Successful gene targeting by HR is dependent on cell proliferation because colonies that derive from individual successfully recombined cells need to be selected for and expanded. Subsequently, these cells can be genotyped and re-implanted into their natural context. Currently, the only type of naturally occurring cells reaching these criteria is the pluripotent ES cell, which is a relatively rapidly dividing cell that can be placed back into blastocysts after gene targeting. In contrast, the multipotent spermatogonial stem cells (SSC), which can be isolated from rats, propagated in culture and can contribute to the germ line when placed back in recipient testes [10,11], are relatively slow-dividing cells and probably unsuitable for gene targeting by HR. Therefore, a prerequisite for this technique is the availability of pluripotent ES cells, but despite many efforts [12,13,14], these have been lacking for the rat for a long time.

Until very recently, the only targetable mammalian ES cells were derived from only a few mouse inbred strains, mainly 129 [15], and the isolation and culturing conditions were empirically based on these few cell lines. However, the same conditions did not yield ES cells from other mouse strains or species. In 2008, a groundbreaking study reported that external cues were dispensable for propagation of ES cells in culture. Instead, the elimination of internal differentiation-inducing signals was sufficient for self-renewal [16]. By adding three inhibitors (3i) that prevent differentiation cues through fibroblast growth factor (FGF)/ERK signaling or glycogen synthase kinase 3 (GSK3) activity, ES cells from other mouse strains [16,17] and also rats [18,19] maintained pluripotency when propagated *in vitro*. This ability to isolate and propagate pluripotent rat ES cells in a culture dish is the first and arguably most important breakthrough towards 'classical' gene targeting in this species. However, to date no transgenic animal using this technique has been reported.

There are several arguments possibly explaining the current lack of knockout rats by ES cell-based HR. First, genetic manipulation of rat ES cells in the 3i condition was reported to be technically challenging because of cell adhesion deficiency and high drug selection sensitivity [19]. Nevertheless, it was also postulated that rat ES cells that

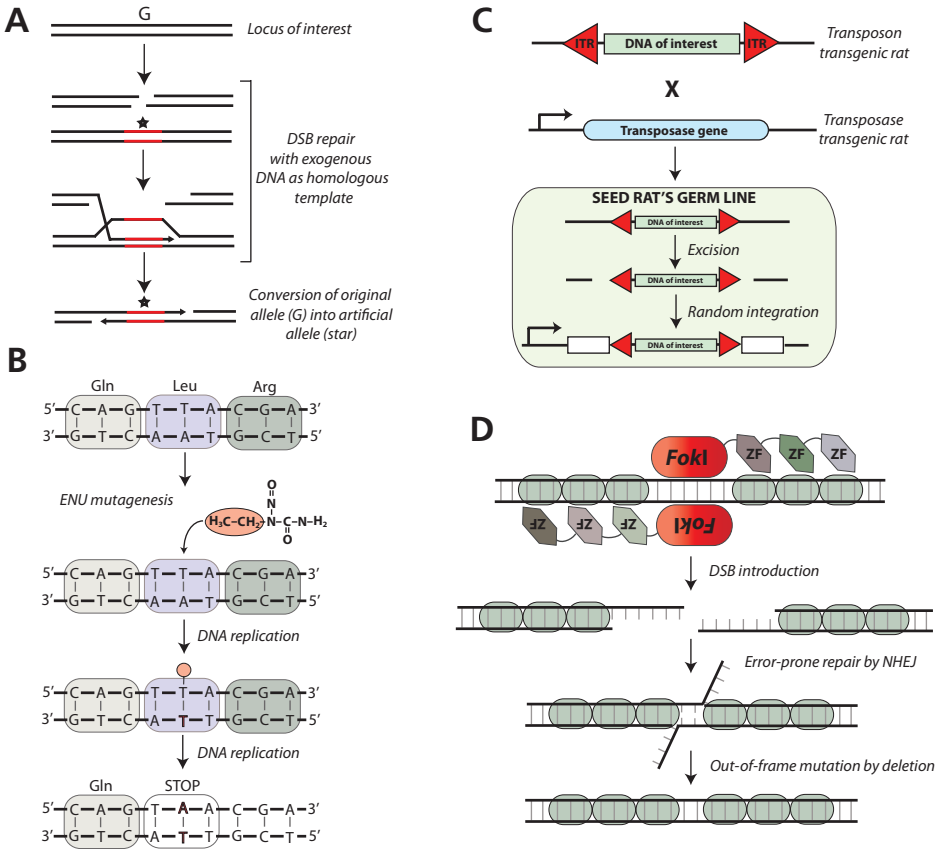

Figure 2: Different techniques available to manipulate the rat genome. (**A**) Schematic representation of gene targeting by homologous recombination (HR). A double strand break (DSB) near a gene of interest (G) is repaired with exogenous DNA as template. Black lines indicate DNA sequence homologous to the target, and the red lines indicate a nonhomologous part (*). (**B**) The mutagenicity of *N*-ethyl-*N*-nitrosourea (ENU) is the result of the capability of transferring the ethyl group to nucleotides in the DNA. During replication this can result in the mis-insertion of a nucleotide and after another round of replication in a single base pair substitution. (**C**) Schematic overview of germ line *Sleeping Beauty* (SB) transposition. A transgenic rat expressing the transposase gene is crossed with a transgenic rat that carries the transposon in its genome. This will produce double transgenic 'seed rats' with transposition events in their germ line, which can be fixed by outcrossing them with wild-type animals. (**D**) A DSB is introduced at a specific locus by fusing two zinc-finger (ZF) arrays to monomeric *Fok*I domains. When no homologous template is available for repair by HR, the DSB is repaired by the error-prone mechanism of nonhomologous end joining (NHEJ). This can result in insertions or deletions and subsequently out-of-frame mutations.

are cultured under 2i conditions, whereby the two inhibitors of fibroblast growth factor (FGF)/ERK signaling are replaced by one more potent MEK inhibitor [18,19], can overcome these challenges. However, it still has to be determined whether rat ES cells retain pluripotency after long-term culture under these conditions. Moreover, even if these problems are overcome, it still has to be determined whether the efficiency of HR as applied in mouse ES cells is sufficient enough for gene targeting. It is, for example, known that the application of this technique in human ES cells is highly inefficient [20]. Second, the incidence of germ line transmission is still low [18], which is also observed in mouse ES cells unless C57BL/6 strain blastocysts are used as hosts [21], underlining the necessity of systematically screening different donor and host strain combinations. Finally, although the karyotypes of the rat ES cells were found to be reasonably stable at earlier passages, chromosomal abnormalities increased at higher passages [18,19]. This finding can have big consequences for generating knockout animals because chromosomal abnormality is one of the major causes of loss of germ line competence of mouse ES cells [22]. Again, cells derived under 2i conditions did not display chromosomal abnormalities [18].

Generation of rat induced pluripotent stem cells (iPSCs)

Recently, a close relative of the ES cell, the iPSC, was generated for the rat [23,24]. This revolutionary technique is based on ectopic expression of 4 defined factors: *Oct-4*, *Sox2*, *c-myc* and *Klf4*, which initiates dedifferentiation of somatic cells, like fibroblast, into a pluripotent state [25]. If kept under the right culturing conditions, these cells retain their pluripotency. Importantly, it was shown that mouse iPSCs form viable chimaeras and can contribute to the germ line when injected into blastocysts [26,27]. It is conceivable that propagation of rat iPSCs under 3i or 2i conditions is essential to maintain pluripotency, like it is for rat ES cells. Indeed, one study reported that rat iPSCs, which were maintained under the standard conditions for mouse ES cells, did not yield chimaeras when injected into blastocysts [24]. In contrast, chimaeras were obtained when rat iPSCs were used that were maintained under slightly modified 3i conditions [23]. However, to date no germ line contribution has been reported yet, which is most probably because of the same problems that hinder successful HR in ES cells.

It is unclear when the first knockout rats generated using HR in stem cells can be expected. First, the conditions for HR in cultured stem cells have to be optimized, and most optimal strain combinations (donor cells and recipient strains) should be sorted out. Nevertheless, the isolation and generation of respectively pluripotent rat ES cells and iPSCs are major steps forward in the field of rat genetics.

ENU TARGET-SELECTED MUTAGENESIS

An alternative technique to manipulate the germ line in rats without the necessity to culture and genetically manipulate ES cells is based on random *N*-ethyl-*N*-nitrosourea

(ENU) mutagenesis of the male germ line, simply by treating the animal with an intraperitoneal injection of the mutagen ENU [28]. This is an approach that originally has been successfully used for phenotype-driven screens in a variety of model organisms, including mice [29,30] and rats [31]. Upon treatment, the ethyl group of ENU is transferred to oxygen or nitrogen radical present in the DNA. Subsequent replication cycles of this damaged DNA can result in mispairing followed by single base pair substitutions (**Fig. 2B**) [32]. Because the mutagenicity of ENU is dependent on multiple cycles of DNA replication, the alkylated DNA nucleotides of cells with a high turnover rate, like spermatogonial stem cells, are likely to result in point mutations. Because of the randomness of mutagenesis, every affected cell will contain a unique set of induced mutations, which will result in a genetically heterogeneous sperm cell population. Subsequently, the ENU-induced mutations are transferred through the germ line by crossing treated males with untreated females, resulting in an F_1 population in which every individual carries unique heterozygous mutations in their genome. Occasionally, knockout-like alleles are generated by introducing a nonsense mutation in the open reading frame (ORF) that results in a premature stop codon and an absent or truncated protein product or a mutation in a splice donor/acceptor site of a gene. Additionally, hypo- and hypermorphic alleles of the same gene can be generated by nonsynonymous mutations, which enable the study of gene-dosage effects and of amino acid residues important for protein-protein interaction or catalytic activity.

Because of the high mutator efficiency of alkylating agents like ENU, genotype-driven mutation discovery, also referred to as TILLING (Targeting Induced Local Lesions In Genomes) [33,34], has been widely implemented in a variety of species, such as the rat [35,36], zebrafish [37], medaka [38], *C. elegans* [39] and a wide range of plant species, including Arabidopsis [40] and maize [41]. In this setup, F_1 animals carrying interesting heterozygous mutations are identified using high-throughput mutation discovery methods, outcrossed and eventually bred to homozygosity. Notably, the first knockout rats reported were generated using ENU target-selected mutagenesis [35,36].

ENU mutagenesis as a gene-targeting tool

Because this technique is based on random introduction of germ line mutations, followed by targeted retrieval of mutations in selected genes of interest, it can be considered to be a 'semi'-targeted gene manipulation approach. The chance of identifying a knockout allele in any given gene depends on the size of that gene; to be specific, the chance increases with the size of the gene and the number of F_1 animals screened, namely the more base pairs covered, the more likely a knockout mutation occurs in a gene. Nevertheless, ENU mutagenesis is highly efficient, resulting in approximately 1–4 knockout alleles and many more missense mutations per F_1 animal, depending on the strain used for mutagenesis [28]. Consequently, the high occurrence of background mutations could form phenotypic complications and should be taken into account when characterizing a specific mutation. It should, however, be noted that in most approaches for the generation of mutant animals

background mutations can be problematic, including HR-based techniques because long-term culturing of ES cells may also result in the accumulation of genetic changes [42]. These issues should not be exaggerated and can relatively easily be overcome by outcrossing heterozygous carriers to the parental strain [43] and using wild-type and heterozygote littermates as controls in phenotypic characterization studies. A unique feature of the approach is that besides its high mutator efficiency, ENU mimics the most common form of human genetic variation, namely SNPs [44], making this technique highly complementary to other techniques, especially in large-scale mutagenesis projects [45]. Indeed, recently we reported highly efficient rat mutant generation for a large collection of genes that encode one-to-one orthologues of human G protein-coupled receptors (GPCRs) [46], demonstrating the effectiveness and utility of ENU target-selected mutagenesis in large-scale projects.

Improving ENU target-selected mutagenesis

The efficiency of the approach essentially depends on the ENU-induced germ line mutation frequency and the efficiency of the mutation identification method. By taking advantage of DNA mismatch repair (MMR) deficiency, a system shown to be involved in repairing ENU damage [47], in the MSH6 knockout rat [48], the efficiency of ENU mutagenesis was increased by 2.5-fold [49]. In addition, massively parallel sequencing technology with microarray-based capture of genomic regions was recently used for re-sequencing almost 1000 genes per individual of up to 20 F_1 animals simultaneously [50], greatly enhancing mutation discovery output. It is conceivable that these next-generation sequencing platforms in the near future will allow for whole exome and even whole genome sequencing of F_1 animals, enabling the identification of all coding mutations, including unwanted 'background' mutations. The next challenge will be to generate archives of large libraries of mutant F_1 individuals, which potentially can be screened indefinitely. Protocols for cryopreserving rat sperm have been optimized [51], and archives of cryopreserved sperm are currently being established for the rat, which can be used for re-derivation of the mutant strains by intracytoplasmic sperm injection (ICSI) [52]. Eventually, mutant models (knockout as well as allelic series of missense mutations) for genes of interest in a range of species may be ordered from pre-screened public collections, thereby creating an invaluable tool for rat genetics by facilitating functional analysis of allelic series.

TRANSPOSON-TAGGED MUTAGENESIS

Transposable elements are relatively large DNA fragments (2,5–10 kb) that can jump around in a genome. If a transposable element is incorporated into the coding region of a gene, it is likely to result in a knockout of that particular gene (although alternative splicing over the transposon-containing exon may occur). Two classes of transposable elements can be distinguished, namely class I elements or retrotransposons and class II

or DNA transposons [53]. Transposons have been successfully used for transgenesis and insertional mutagenesis in the nematode *Caenorhabditis elegans* [54] and *Drosophila melanogaster* [55], but it was the class II-based *Sleeping Beauty* (SB) transposon system, awakened from fish, which is sufficiently active for insertional mutagenesis in vertebrates [56]. After applying this transposon-mediated forward genetic approach in mice [57,58,59], it was adopted for germ line mutagenesis in rats [60,61].

Germ line transposition

A variety of transposon systems has been implemented in mutagenesis and transgenesis studies [53] and consists of two components: a transposable element and a transposase enzyme that catalyzes the transposition. The transposon is flanked by two transposase recognition sites and upon recognition is excised, followed by integration into a new sequence environment. Insertional mutagenesis is achieved by expressing the transposase in cells that carry the transposable element. For germ line transmission of the transposed element, two transgenic animals are crossed, one expressing the transposase in either all cells [60] or specifically in spermatocytes and spermatids [61] and one carrying a nonautonomous transposon. This crossing results in the generation of double transgenic rats or 'seed rats', which carry transposition events in their germ line (**Fig. 2C**). A unique feature of this technique is that the tranposon can be equipped with gene trap cassettes, which contain reporter genes like *LacZ* and *GFP* [60] or a tyrosinase mini-gene that rescues albinism and changes coat color when expressed [61]. The 'seed rats' are crossed with wild-type rats to generate an F_1 generation with fixed gene trap events in their genomes, which can be easily scored by expression of the reporter gene. The transposition insertion sites are subsequently located by standard PCR protocols from genomic DNA isolated from F_1 animals [53]. This manner of transposition is referred to as chromosomal transposition because the transposon is located in the genome of the 'seed rats'. Alternatively, transposon delivery can be achieved by expressing the transposase in pluripotent ES cells or multipotent SSCs, which carry the nonautonomous transposon. After transposition events, the cells can be placed back in the blastocyst for generating chimaeras or recipient testes for germ line transmission. There are two ways of allowing transposition in cultured cells, either by transfection-based 'plasmid-to-genome' delivery, which yields relatively unbiased genome-wide integration [53] or in the case of ES cells, by intragenomic, genome-to-genome mobilization by integrating the transposon into a specific locus in the genome by HR [57]. However, to date no *in vitro* transposon-mediated mutagenesis in rat ES cells or SSCs has been reported.

Randomness of transposon integration

The randomness of integration is strongly affected by integration site preference of the transposable element [53]. Firstly, the SB transposon, like other members of the *Tc1/mariner* family, always integrates into a TA target dinucleotide that is duplicated

upon insertion [56,62]. Importantly, although the SB transposon shows a small but significant bias insertion into genes and their upstream regulatory sequences, the gene hits are predominantly located in intronic regions [63]. Furthermore, it has been shown that chromosomal transposition is heavily biased toward local hopping into closely linked loci [57,58,64,65]. This latter bias can be overcome by using the 'plasmid-to-genome' delivery, which allows for more genome-wide integration coverage. Nevertheless, local hopping can also be beneficial when interested in a specific genomic locus, like a QTL. In this case a nonautonomous transposable element is integrated in a specific genomic locus, followed by transposase expression and local hopping. Notably, although it is also suggested that particular transposable elements are biased towards integration into expressed genes versus silent genes [53], it was observed not to be the case for the SB transposon [63]. Taken together, it can be concluded that transposons integrate nonrandomly throughout the genome. Although this could greatly influence its utility for knocking out genes in a genome-wide fashion, the ease of scoring gene trap events and locating integration sites makes transposon-tagged mutagenesis a powerful forward genetic technique in the rat genetic toolbox.

ZINC-FINGER NUCLEASES

An alternative relatively novel technique to knocking out genes in a targeted fashion without the need for pluripotent cells is based on the use of genetically engineered zinc-fingers nucleases (ZFNs). This approach is based on the observation that double-strand breaks (DSB), which are potentially lethal to the cell when they remain unrepaired, either increase HR and gene targeting or repair by error-prone nonhomologous end joining (NHEJ) [66]. By fusing sequence-specific zinc-fingers (ZFs), which are found in the DNA-binding domain of most transcription factors in most eukaryotic genomes [67], to the nonsequence-specific cleavage domain of the *Fok*I endonuclease, pre-determined genomic DSBs can be introduced. In the absence of a homologous template for error-free repair, DSBs will be repaired by NHEJ, which is often accompanied by deletions or insertions. If a DSB is introduced in the coding region of a gene or at intron-exon boundaries, repair by NHEJ can result in out-of-frame mutations or aberrant splicing and consequently in a knockout allele. This gene targeting approach has been successfully applied in a variety of model organisms, including *Drosophila melanogaster* [68], *Arabidopsis thaliana* [69], zebrafish [70,71] and the rat [72]. The main challenges for successful ZFN-mediated gene targeting are the design of the zinc-fingers arrays to achieve sufficient specificity for the targeted gene and correct expression of the ZFNs to ensure germ line transmission of the targeted gene.

Design and specificity of ZFNs

The unique features that makes zinc-fingers ideal for directing enzymatic domains, like *Fok*I, to pre-determined genetic loci are that each finger binds its 3-bp target site

independently [73] and that ZFs have been identified for almost all of the 64 DNA triplets [74]. By fusing independent fingers, target site-specificity is achieved and should increase with the number of fingers used. In addition, to cut DNA, the *Fok*I cleavage domain must dimerize [75,76], which is achieved by binding two sets of ZFs, each linked to a monomeric cleavage domain, with binding sites in an inverted orientation and thereby enhancing site specificity (**Fig. 2D**) [74].

There are different ways for generating ZFNs of which the most accessible method is modular assembly via standard recombinant DNA technology [74]. The key for success using this approach is finding a suitable target site in the gene of interest. In particular, ZFs that target 5'-GNN-3' (where N is any base) triplets in the target sequence have been tested extensively and give the most encouraging results [74]. However, high failure rates have been reported for modularly assembled ZF arrays, especially target sites composed of two, one or no 5'-GNN-3' triplets [77]. Although some successful targeting has been reported with modularly assembled ZFN in human cells [78,79] and *Drosophila melanogaster* [80], inconsistencies in the success rate [81] make this method until now inefficient for routine gene targeting in model organisms. Alternatively, zinc-finger arrays can successfully be constructed in an unbiased way by using a cell-based selection method, like the publicly available oligomerized pool engineering (OPEN) [82]. However, cell-based selection methods are labor intensive and time consuming, and currently ZFNs made using OPEN are thus far limited to targeting 5'-GNN-3' repeats, which occurs rarely in a given gene of interest [81]. Finally, a proprietary method to design ZFNs used by Sangamo Biosciences exists [71], which is licensed to Sigma-Aldrich. To date, this system has been the only used method that has successfully generated ZFN-modified knockout rats [72,83]; however it is an expensive approach. Custom-made ZFNs are sold for $35,000 to researchers capable of injecting them on their own (see below). Alternatively, a knockout breeding pair can be bought for $95,000, and the company keeps the intellectual property.

Expressing the ZFNs

To establish germ line transmission of an aberrantly repaired gene of interest, the ZFNs are injected into fertilized oocytes, which can give rise to chimaeric genetically modified offspring [72,83]. Subsequently, these ZFN-modified founders are identified and crossed with wild-type animals to generate an F_1 population carrying the modified allele in their genome. However, off-target effects, like the cleavage and mutagenesis of genomic loci other than the target, of the ZFNs should be taken into account because this increases toxicity and background mutations [84]. Nevertheless, short-term expression of the ZFN, by injecting mRNA instead of plasmid DNA, will most probably decrease these effects, without affecting the efficiency of the approach [72]. Furthermore, outcrossing to the parental strain, just as with ENU-induced mutant animals, should eliminate background mutations.

Besides the targeted characteristic, the main advantage of ZFN-mediated gene knockout technology is that it is fast after obtaining the ZFNs, which can be the time-limiting step especially when the finger arrays are modularly designed. After injecting the ZFNs into embryos, in just a few months ZFN-modified founders can be scored. Like ENU and transposon-mediated mutagenesis, there is no dependence on ES cell or iPSC culturing and manipulation. Furthermore, because ZFN-mediated DSBs in a gene of choice increases efficiency of HR *in vivo* [85], this technique could potentially allow for targeted knock-in animals, by simply co-injecting an artificially assembled construct together with the ZFNs. This would broaden the genetic toolbox in the rat by allowing techniques that otherwise depend on the necessity of culturing and manipulating ES cells or iPSCs, like generation of conditional knockout alleles and *in vivo* cell lineage tracing.

GENETICALLY MODIFIED RAT MODELS

Cancer-related models

The first genetic knockout rat models generated lacked well-known tumor suppressor proteins, like BRCA2 [36,86], APC [87] and MSH6 [48]. Although the mouse has been the mammalian model organism of choice for studying human tumorigenesis by overexpressing oncogenes or inactivating tumor suppressor genes [88], these rat models do add additional knowledge and are in fact highly complementary.

Human carriers of germ line mutations in *BRCA2* have increased risk of developing breast cancer, pancreas cancer, gastrointestinal cancer and prostate cancer [89], which is probably the result of the important role of BRCA2 in repairing DSBs through interaction with the RAD51 recombination enzyme [90]. Truncating mutations that occur 5' of exon 11 and thereby eliminating the RAD51-interacting BRC repeats of both BRCA1 or BRCA2 in mice results in embryonic lethality [91]. Using ENU mutagenesis combined with a yeast-based assay to specifically screen for mutations that truncate target proteins, nonsense mutations were identified in the ORF of both *Brca1* and *Brca2* in the rat. Whereas homozygosity of the *Brca1* mutation in rats also resulted in embryonic lethality (Michal Gould, personal communication), homozygous mutant *Brca2* rats were viable even though the mutation truncated the protein 5' of exon 11 [86]. These animals displayed sterility, lower body weight and a significant reduction in survival as a result of increased spontaneous tumorigenesis [36,86].

Using the same approach, a nonsense mutation was identified in the colon cancer-associated gene encoding adenomatous polyposis coli (APC) protein [87]. The majority of spontaneous and familial adenomatous polyposis (FAP) colonic tumors carry inactivating mutations in this gene, and therefore APC has been designated as a gatekeeper in colonic epithelial cells [92]. Mouse strains carrying heterozygous mutations in the *Apc* gene, like the Apc^{Min} strain, primarily develop small intestinal tumors and die within several months [93]. Interestingly, heterozygous mutant rats

primarily developed colorectal cancers, and the size of the rat allowed for colonoscopy to image colonic tumorigenesis *in vivo* [87].

The DNA MMR machinery safeguards genomic integrity by repairing mismatches, insertion or deletion loops and responding to genotoxic agents [94]. In humans, mutations in MMR genes have been linked to hereditary non-polyposis colorectal cancer (HNPCC), also referred to as Lynch syndrome, which is characterized by early-onset colon cancer [95]. In an ENU target-selected mutagenesis screen, a nonsense mutation was identified in the fourth exon of the MMR gene encoding for MSH6 [35], which together with MSH2 forms a heterodimer that is necessary for the recognition of single nucleotide mismatches and small insertion or deletion loops [94]. The nonsense mutation resulted in protein truncation and loss of DNA MMR [48]. As a consequence, homozygous mutant rats showed a decrease in life span as the result of increased spontaneous tumorigenesis [48]. However, the survival of *msh6*$^{-/-}$ rats is considerably longer than in mice lacking MSH6 [96,97], which display a much earlier onset of tumorigenesis. Furthermore, although the tumor spectra between MSH6 knockout mice and rats showed high similarities, a notable difference was the high incidence of uterus cancer in female *msh6*$^{-/-}$ rats [48]. In humans, mutations in *MSH6* have been associated with atypical HNPCC, which is characterized by a late onset and a high occurrence of extracolonic tumors, especially in the endometrium [98].

These rat models exemplify the utility of genetically modified rats in cancer research. Although mouse knockout models have been extremely powerful tools for identifying important oncogenes and tumor suppressor genes, there are discrepancies between the human disease phenotype and phenotypes observed in mouse models, even between mouse models that lacked the same gene but in different strain background. Indeed, the rat models described here also do not perfectly model the associated human tumorigenesis. However, the availability of comparable mammalian mutant models in different species does allow for *in vivo* phenotypic comparison and filtering out of species-specific effects. In addition, both species have advantages, like the broad genetic toolbox in the mouse. The rat seems to be more tolerant to tumorigenesis because they generally display a later onset of spontaneous tumorigenesis, increased survival and a capacity for large tumor sizes [48,87]. Recently, using ENU target-selected mutagenesis, a knockout allele of the gene encoding for the tumor suppressor p53 was identified (RB, EC unpublished results). This mutation has been bred to homozygosity and does develop tumors at high incidence, although more detailed phenotyping will be required to fully characterize this model.

Immune system-related models

Several rat knockout models have been generated for genes involved in immune response. Severe combined immunodeficiency (SCID) in mice has been extremely

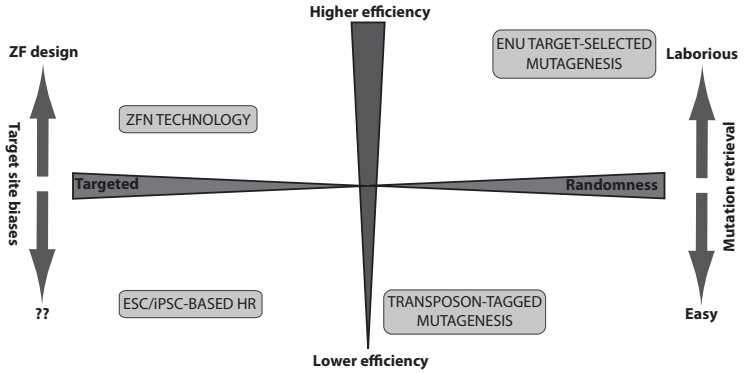

Figure 3: Schematic representation of complementarities between the discussed techniques. Thickness of the axes represents the weight of indicated factor. Both *N*-ethyl-*N*-nitrosourea (ENU) target-selected mutagenesis and transposon-tagged mutagenesis are dependent on the random introduction of mutagenic lesions. The latter technique is, however, more biased towards specific genetic loci, whereas the mutagenicity of ENU is more random from a genome-wide perspective. Also the number of germ line transposition events is much lower than the number of ENU-induced germ line mutations. Nevertheless, identifying gene trap events and locating the transposon insertion site is much easier than ENU-induced mutation identification. In contrast, mutagenesis by homologous recombination (HR) in embryonic stem (ES) cells and induced pluripotent stem cells (iPSCs) or zinc-finger nuclease (ZFN)-mediated nonhomologous end joining (NHEJ) is targeted to predetermined genetic loci. Gene targeting by HR is highly inefficient and occurs in only several cells per million, urging the need for a rapidly dividing cell system. ZFNs increase the efficiency by specifically introducing a double strand break at the desired locus thereby enhancing repair by either NHEJ or even HR. Nevertheless, the design of zinc finger (ZF) arrays remains difficult.

useful for generating humanized models by engrafting haematopoietic cells or tissues or by transgenically expressing human genes [99]. In humans, X-linked SCID (X-SCID) is the most common form of SCID and is caused by mutation in the gene encoding the interleukin 2 receptor gamma (*Il2rg*) [100,101]. Using commercially custom-designed ZFN, an IL2RG-deficient rat model was generated, which successfully served as a host for xenotransplantation of human ovarian cancer tumor cells [83]. Using the same technology, the gene that encodes for immunoglobulin M (*IgM*) was knocked out in the rat, which can be used to manufacture human monoclonal antibodies in the rat for therapeutic use [72].

By re-sequencing a large collection of GPCRs in an F_1 library generated by outcrossing mutagenized *msh6*$^{-/-}$ males, multiple rat models were generated with mutant alleles for receptors involved in immune function [46]. These included knockout alleles for the interleukin 8 receptor *Cxcr2*, the proton-sensing receptor *Gpr65*, the medium fatty-acid receptor *Gpr84* and the chemokine receptor *Ccr4*. Notably, rats homozygous for the nonsense mutation in *Cxcr2* were highly susceptible to external pathogens, and some individuals showed highly increased lymph nodes and spleen sizes resulting from a bacterial infection (RB, EC unpublished results).

Complex behavior-related models

Traditionally, the rat has been the preferred organism for modeling complex human behavior and neurobiology. The opportunity to generate genetically modified rats will most probably enhance the use of the rat for studying the genetic aspects underlying human behavioral processes and disease. Indeed, several genetic key players in emotion, motivation, cognition and food intake have been knocked out and studied in the rat, like the serotonin transporter (SERT) [102], the melanin-concentrating hormone precursor (PMCH) [103] and the melanocortin 4 receptor (MC4R) [46].

Serotonergic signaling activity heavily depends on the extracellular fluid concentration of the neurotransmitter serotonin (5-HT), which is regulated by SERT. Given the important role of the highly evolutionarily conserved serotonergic system in regulating emotion, motivation and cognition, it is not surprising that genetic variants in humans have been associated with many behavioral disorders, including depression, drug addiction, schizophrenia and eating disorders [104]. Importantly, pharmacologically blocking SERT activity by selective serotonin re-uptake inhibitors (SSRIs) is among the most used therapeutic intervention to treat several neuropsychiatric disorders, including depression [105]. An ENU-induced premature stop codon was identified in offspring of mutagenized animals in the third exon of the gene encoding SERT [35], which resulted in the complete absence of the protein in homozygous mutant animals [102]. Although SERT knockout mice and rats show high similarity in behavioral endophenotypes, significant differences were also found [106], underlining the importance of both genetic knockout mice and rats for phenotypic comparison to understand the physiological role of SERT in humans. Interestingly, heterozygous mutant animals display an intermediate phenotype with respect to expression and function indicative of gene-dosage effects, which may mimic decreased transcription efficiency associated with human anxiety-related traits as a result of polymorphisms in the promoter region of SERT [107].

Using the same approach, rat knockout models were generated for genes involved in food intake. Rats deficient in PMCH were lean and hypophagic as compared with wild-type controls, which was mainly because of PMCH deficiency during early development, arguing for a critical role of MCH in body set point determination [103]. In contrast, an ENU-induced nonsense mutation in the gene encoding MC4R, which truncated the C-terminus of the receptor, resulted in increased body weight and size [46], mimicking the many point mutations identified in the human *MC4R* gene that are associated with severe forms of obesity [108]. Interestingly, the nonsense mutation identified in rat MC4R, causing truncation of the last 18 amino acids, was still expressed *in vitro* but was unable to be located at the plasma membrane, the site of function for the receptor [46]. Many mutations identified in human *MC4R* associated with obesity were also shown *in vitro* to result in impaired plasma membrane expression [109].

These models exemplify the utilization of genetically modified rat models for understanding the genetic basis of human complex behavior and disease. Moreover,

because therapeutic intervention for many behavioral disorders is characterized by unwanted severe side effects and the laboratory rat has made an enormous contribution to pharmacological screening and testing, these genetic knockout models are likely to contribute to improving current drug development.

CONCLUSIONS

The many technical developments described in this review have enabled the generation of rat genetic mutants in a systematic and highly efficient manner. The different techniques are highly complementary, all having special features, advantages and disadvantages (**Fig. 3**). It is therefore not probable that one technique will prevail over another, but rather that all of them will be implemented depending on the research question. It is also likely that certain aspects of the different techniques as described here will be combined to strengthen the approach or facilitate specific output. For example, ES cells or iPSCs can be used to specifically target a specific locus, like QTLs, by incorporating a transposon by HR as has been done in mice [57], followed by local hopping to identify cis-acting modifiers in an objective manner. Furthermore, the stem cells could also be used for *in vitro* chemical mutagenesis to generate large archives of mutant alleles, which has also been done with mouse ES cells [110]. To knockout 95% of all the rat genes, a living library or sperm archive of 40,000 F_1 animals has to be generated using an MSH6-deficient background [49], which is currently not achievable. However, a high number of ES cells or iPSCs can easily be mutagenized in a petri dish, clonally expanded and split for DNA isolation and cryopreservation. In addition, different alkylating agents can be tested and even RNAi can be used to knockdown specific DNA repair pathways during mutagenesis. It was, for example, shown that lack of DNA MMR protected mouse ES cells from the toxic effects of alkylating agents, while increasing the chemically-induced mutation frequencies [47]. The rat genetic community is still eagerly waiting for the first reports describing gene targeting by HR in ES cells or iPSCs because this would allow for conditional knockout alleles and knock-ins. However, the emerging ZFN-mediated mutagenesis also could allow HR with exogenous DNA, without the need for ES cells manipulation and time-consuming selection procedures, by simply co-injecting the DNA construct for recombination together with the mRNA encoding the ZFNs [85], although a proof-of-principle for this remains to be demonstrated for the rat.

Although numerous rat knockout models have already been generated (see, for example, www.ratknockout.org), systematic characterization and application of these animals in modeling human disease are still in their infancy. It could be that, because of the emphasis of the rat genetics field on genomic manipulation and technological developments, systematic phenotypic screening protocols are lacking. On the other hand, it could be that researchers that traditionally work with rats find it hard to apply the genetic models in their analyses and prefer, for example, pharmacological

inhibition or stimulation. Extensive phenotype protocols as they exist for mice [111] would definitely accelerate the use and application of genetically modified rat models.

In conclusion, technical developments for manipulating the rat genome have enormously contributed to expanding the genetic toolbox in this model organism, and many more improvements can be expected in the near future. In addition, the use of rat knockout models will significantly contribute to biomedical research by enabling mammalian interspecies phenotypic comparisons and by taking advantage of species-specific characteristics for studying different aspects of human physiology and disease.

AIMS AND OUTLINE OF THIS THESIS

The work described in this thesis is the result of the efforts to generate genetically modified rat models using chemical mutagenesis and to use these models in biomedical research.

Chapter 2 describes detailed protocols for ENU mutagenesis approaches in the rat. The outline of the method is simple: male animals are treated with the mutagen and crossed with untreated females to generate an F_1 population, which can be screened for mutations that affect protein function. These mutations can be retrieved via genotype-driven of phenotype-driven approaches.

In **Chapter 3** the characterization of an ENU-induced knockout allele of the DNA MMR protein MSH6 is described. This rat model displays high levels of genomic instability and consequently spontaneous late-onset tumorigenesis, which partially mimics human hereditary non-polyposis colorectal cancer.

In **Chapter 4** we take advantage of MMR-deficiency in the MSH6 knockout rat in order to improve the efficiency of ENU mutagenesis. MMR-deficiency not only increases the ENU-induced germ line mutation rate, but also the mutation spectrum in such a way that the chance of introducing premature stop codon is significantly elevated.

A large collection of ENU-induced *in vivo* GPCR mutants has been generated in **Chapter 5**. In a predetermined panel of one-to-one orthologs of human GPCRs multiple novel knockout and missense alleles are discovered. Using comprehensive computational analyses, the impact of the mutations on normal receptor function is predicted. We experimentally show that a nonsense mutation in *Mc4r* and a missense mutation in the lysophosphatidic acid receptor *Lpar1* indeed result in loss-of-function phenotypes.

Chapter 6 provides a proof of principle for the use of next-generation sequencing technologies for target-selected reverse genetics procedures and demonstrates a very significant improvement in the efficiency of the approach.

In **Chapter 7** molecular and functional analyses are described, suggesting that the apparent loss-of-function phenotype in $LPAR1^{M318R}$ may be caused by constitutive arrestin-mediated desensitization.

Finally, in the **General discussion** the future of ENU target-selected mutagenesis as a tool for rat reverse genetics, compared to the novel techniques described in Chapter 1, is discussed. Furthermore, future technological possibilities of ENU mutagenesis are described.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304-1351.
3. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6: 507-512.
4. Jacob HJ (1999) Functional genomics and rat models. Genome Res 9: 1013-1016.
5. Sawinski J, Wallace DJ, Greenberg DS, Grossmann S, Denk W, et al. (2009) Visually evoked activity in cortical cells imaged in freely moving animals. Proc Natl Acad Sci U S A 106: 19557-19562.
6. Snyder JS, Choe JS, Clifford MA, Jeurling SI, Hurley P, et al. (2009) Adult-born hippocampal neurons are more numerous, faster maturing, and more involved in behavior in rats than in mice. J Neurosci 29: 14484-14495.
7. Lazar J, Moreno C, Jacob HJ, Kwitek AE (2005) Impact of genomics on research in the rat. Genome Res 15: 1717-1728.
8. Saar K, Beck A, Bihoreau MT, Birney E, Brocklebank D, et al. (2008) SNP and haplotype mapping for genetic analysis in the rat. Nat Genet 40: 560-566.
9. Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, et al. (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40: 516-522.
10. Hamra FK, Chapman KM, Nguyen DM, Williams-Stephens AA, Hammer RE, et al. (2005) Self renewal, expansion, and transfection of rat spermatogonial stem cells in culture. Proc Natl Acad Sci U S A 102: 17430-17435.
11. Ryu BY, Kubota H, Avarbock MR, Brinster RL (2005) Conservation of spermatogonial stem cell self-renewal signaling between mouse and rat. Proc Natl Acad Sci U S A 102: 14302-14307.
12. Buehr M, Nichols J, Stenhouse F, Mountford P, Greenhalgh CJ, et al. (2003) Rapid loss of Oct-4 and pluripotency in cultured rodent blastocysts and derivative cell lines. Biol Reprod 68: 222-229.
13. Fandrich F, Lin X, Chai GX, Schulze M, Ganten D, et al. (2002) Preimplantation-stage stem cells induce long-term allogeneic graft acceptance without supplementary host conditioning. Nat Med 8: 171-178.
14. Vassilieva S, Guan K, Pich U, Wobus AM (2000) Establishment of SSEA-1- and Oct-4-expressing rat embryonic stem-like cell lines and effects of cytokines of the IL-6 family on clonal growth. Exp Cell Res 258: 361-373.
15. Gardner RL, Brook FA (1997) Reflections on the biology of embryonic stem (ES) cells. Int J Dev Biol 41: 235-243.
16. Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, et al. (2008) The ground state of embryonic stem cell self-renewal. Nature 453: 519-523.
17. Nichols J, Jones K, Phillips JM, Newland SA, Roode M, et al. (2009) Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat Med 15: 814-818.
18. Buehr M, Meek S, Blair K, Yang J, Ure J, et al. (2008) Capture of authentic embryonic stem cells from rat blastocysts. Cell 135: 1287-1298.
19. Li P, Tong C, Mehrian-Shai R, Jia L, Wu N, et al. (2008) Germline competent embryonic stem cells derived from rat blastocysts. Cell 135: 1299-1310.
20. Zwaka TP, Thomson JA (2003) Homologous recombination in human embryonic stem cells. Nat Biotechnol 21: 319-321.
21. Seong E, Saunders TL, Stewart CL, Burmeister M (2004) To knockout in

129 or in C57BL/6: that is the question. Trends Genet 20: 59-62.
22. Liu X, Wu H, Loring J, Hormuzdi S, Disteche CM, et al. (1997) Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. Dev Dyn 209: 85-91.
23. Li W, Wei W, Zhu S, Zhu J, Shi Y, et al. (2009) Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. Cell Stem Cell 4: 16-19.
24. Liao J, Cui C, Chen S, Ren J, Chen J, et al. (2009) Generation of induced pluripotent stem cell lines from adult rat cells. Cell Stem Cell 4: 11-15.
25. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663-676.
26. Okita K, Ichisaka T, Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. Nature 448: 313-317.
27. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, et al. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448: 318-324.
28. van Boxtel R, Gould MN, Cuppen E, Smits BM ENU mutagenesis to generate genetically modified rat models. Methods Mol Biol 597: 151-167.
29. Hrabe de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, et al. (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25: 444-447.
30. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A (1999) Mouse ENU mutagenesis. Hum Mol Genet 8: 1955-1963.
31. Smits BM, Peters TA, Mul JD, Croes HJ, Fransen JA, et al. (2005) Identification of a rat model for usher syndrome type 1B by N-ethyl-N-nitrosourea mutagenesis-driven forward genetics. Genetics 170: 1887-1896.
32. Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. Mamm Genome 11: 478-483.
33. Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135: 630-636.
34. Stemple DL (2004) TILLING--a high-throughput harvest for functional genomics. Nat Rev Genet 5: 145-150.
35. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
36. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
37. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RH, et al. (2003) Efficient target-selected mutagenesis in zebrafish. Genome Res 13: 2700-2707.
38. Taniguchi Y, Takeda S, Furutani-Seiki M, Kamei Y, Todo T, et al. (2006) Generation of medaka gene knockout models by target-selected mutagenesis. Genome Biol 7: R116.
39. Cuppen E, Gort E, Hazendonk E, Mudde J, van de Belt J, et al. (2007) Efficient target-selected mutagenesis in Caenorhabditis elegans: toward a knockout for every gene. Genome Res 17: 649-658.
40. Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13: 524-530.
41. Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, et al. (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12.
42. Liang Q, Conte N, Skarnes WC, Bradley A (2008) Extensive genomic copy number variation in embryonic stem cells. Proc Natl Acad Sci U S A 105: 17453-17456.
43. Keays DA, Clark TG, Flint J (2006) Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. Mamm Genome 17: 230-238.
44. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-236.
45. Gondo Y (2008) Trends in large-scale mouse mutagenesis: from genetics to functional genomics. Nat Rev Genet 9: 803-810.

46. van Boxtel R, Vroling B, Toonen P, Nijman IJ, van Roekel H, et al. (2010) Systematic generation of in vivo G protein-coupled receptors mutants in the rat. Pharmacogenomics J in press.
47. Claij N, van der Wal A, Dekker M, Jansen L, te Riele H (2003) DNA mismatch repair deficiency stimulates N-ethyl-N-nitrosourea-induced mutagenesis and lymphomagenesis. Cancer Res 63: 2062-2066.
48. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
49. van Boxtel R, Toonen PW, Verheul M, van Roekel HS, Nijman IJ, et al. (2008) Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics 9: 460.
50. Nijman IJ, Mokry M, van Boxtel R, Toonen P, de Bruijn E, et al. (2010) Targeted genomic enrichment and next-generation sequencing for gene-driven reverse genetic approaches. Submitted.
51. Nakatsukasa E, Inomata T, Ikeda T, Shino M, Kashiwazaki N (2001) Generation of live rat offspring by intrauterine insemination with epididymal spermatozoa cryopreserved at -196 degrees C. Reproduction 122: 463-467.
52. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
53. Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, et al. (2009) Transposon-mediated genome manipulation in vertebrates. Nat Methods 6: 415-422.
54. Zwaal RR, Broeks A, van Meurs J, Groenen JT, Plasterk RH (1993) Target-selected gene inactivation in Caenorhabditis elegans by using a frozen transposon insertion mutant bank. Proc Natl Acad Sci U S A 90: 7431-7435.
55. Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, et al. (2004) A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nat Genet 36: 283-287.
56. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91: 501-510.
57. Luo G, Ivics Z, Izsvak Z, Bradley A (1998) Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc Natl Acad Sci U S A 95: 10769-10773.
58. Fischer SE, Wienholds E, Plasterk RH (2001) Regulated transposition of a fish transposon in the mouse germ line. Proc Natl Acad Sci U S A 98: 6759-6764.
59. Horie K, Kuroiwa A, Ikawa M, Okabe M, Kondoh G, et al. (2001) Efficient chromosomal transposition of a Tc1/mariner- like transposon Sleeping Beauty in mice. Proc Natl Acad Sci U S A 98: 9191-9196.
60. Kitada K, Ishishita S, Tosaka K, Takahashi R, Ueda M, et al. (2007) Transposon-tagged mutagenesis in the rat. Nat Methods 4: 131-133.
61. Lu B, Geurts AM, Poirier C, Petit DC, Harrison W, et al. (2007) Generation of rat mutants using a coat color-tagged Sleeping Beauty transposon system. Mamm Genome 18: 338-346.
62. Plasterk RH (1996) The Tc1/mariner transposon family. Curr Top Microbiol Immunol 204: 125-143.
63. Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, et al. (2005) High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 25: 2085-2094.
64. Carlson CM, Dupuy AJ, Fritz S, Roberg-Perez KJ, Fletcher CF, et al. (2003) Transposon mutagenesis of the mouse germline. Genetics 165: 243-256.
65. Horie K, Yusa K, Yae K, Odajima J, Fischer SE, et al. (2003) Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. Mol Cell Biol 23: 9189-9207.
66. Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14: 8096-8106.
67. Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23: 967-973.
68. Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using

zinc-finger nucleases. Genetics 161: 1169-1175.

69. Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102: 2232-2237.
70. Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26: 695-701.
71. Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, et al. (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26: 702-708.
72. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, et al. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325: 433.
73. Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252: 809-817.
74. Carroll D, Morton JJ, Beumer KJ, Segal DJ (2006) Design, construction and in vitro testing of zinc finger nucleases. Nat Protoc 1: 1329-1341.
75. Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, et al. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res 28: 3361-3369.
76. Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A 95: 10570-10575.
77. Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5: 374-375.
78. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res 19: 1279-1288.
79. Lee HJ, Kim E, Kim JS (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res 20: 81-89.
80. Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, et al. (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105: 19821-19826.
81. Kim JS, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc-finger nucleases. Nat Methods 7: 91; author reply 91-92.
82. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, et al. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294-301.
83. Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, et al. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5: e8870.
84. Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases. Genetics 172: 2391-2403.
85. Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300: 764.
86. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.
87. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
88. Resor L, Bowen TJ, Wynshaw-Boris A (2001) Unraveling human cancer in the mouse: recent refinements to modeling and analysis. Hum Mol Genet 10: 669-675.
89. (1999) Cancer risks in BRCA2 mutation carriers. The Breast Cancer Linkage Consortium. J Natl Cancer Inst 91: 1310-1316.
90. Shivji MK, Venkitaraman AR (2004) DNA recombination, chromosomal stability and carcinogenesis: insights into the role of BRCA2. DNA Repair (Amst) 3: 835-843.
91. Moynahan ME (2002) The cancer connection: BRCA1 and BRCA2 tumor suppression in mice and humans. Oncogene 21: 8994-9007.
92. Kinzler KW, Vogelstein B (1996) Lessons from hereditary colorectal cancer. Cell 87: 159-170.

93. Moser AR, Pitot HC, Dove WF (1990) A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science 247: 322-324.
94. Jiricny J (2006) The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7: 335-346.
95. Lynch HT, Smyrk T (1996) Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. Cancer 78: 1149-1167.
96. de Wind N, Dekker M, Claij N, Jansen L, van Klink Y, et al. (1999) HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nat Genet 23: 359-362.
97. Edelmann W, Yang K, Umar A, Heyer J, Lau K, et al. (1997) Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell 91: 467-477.
98. Wijnen J, de Leeuw W, Vasen H, van der Klift H, Moller P, et al. (1999) Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat Genet 23: 142-144.
99. Shultz LD, Ishikawa F, Greiner DL (2007) Humanized mice in translational biomedical research. Nat Rev Immunol 7: 118-130.
100. Noguchi M, Yi H, Rosenblatt HM, Filipovich AH, Adelstein S, et al. (1993) Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. Cell 73: 147-157.
101. Leonard WJ (2001) Cytokines and immunodeficiency diseases. Nat Rev Immunol 1: 200-208.
102. Homberg JR, Olivier JD, Smits BM, Mul JD, Mudde J, et al. (2007) Characterization of the serotonin transporter knockout rat: a selective change in the functioning of the serotonergic system. Neuroscience 146: 1662-1676.
103. Mul JD, Yi CX, van den Berg SA, Ruiter M, Toonen PW, et al. (2010) Pmch expression during early development is critical for normal energy homeostasis. Am J Physiol Endocrinol Metab 298: E477-488.
104. Murphy DL, Lerner A, Rudnick G, Lesch KP (2004) Serotonin transporter: gene, genetic disorders, and pharmacogenetics. Mol Interv 4: 109-123.
105. Wong ML, Licinio J (2004) From monoamines to genomic targets: a paradigm shift for drug discovery in depression. Nat Rev Drug Discov 3: 136-151.
106. Kalueff AV, Olivier JD, Nonkes LJ, Homberg JR (2010) Conserved role for the serotonin transporter gene in rat and mouse neurobehavioral endophenotypes. Neurosci Biobehav Rev 34: 373-386.
107. Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, et al. (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science 274: 1527-1531.
108. Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, et al. (2003) Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. N Engl J Med 348: 1085-1095.
109. Nijenhuis WA, Garner KM, van Rozen RJ, Adan RA (2003) Poor cell surface expression of human melanocortin-4 receptor mutations associated with obesity. J Biol Chem 278: 22939-22945.
110. Chen Y, Yee D, Dains K, Chatterjee A, Cavalcoli J, et al. (2000) Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. Nat Genet 24: 314-317.
111. Rogers DC, Fisher EM, Brown SD, Peters J, Hunter AJ, et al. (1997) Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mamm Genome 8: 711-713.

2

ENU MUTAGENESIS TO GENERATE GENETICALLY MODIFIED RAT MODELS

Ruben van Boxtel[1], Michael N. Gould[2], Bart M. G. Smits[2] and Edwin Cuppen[1]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2]McArdle Lab for Cancer Research University of Wisconsin-Madison Madison, WI 53706-1599 USA

Adapted from Methods in Molecular Biology (2010) 597: 151-67 and Methods in Molecular Biology (in press)

ABSTRACT

The rat is one of the most preferred model organisms in biomedical research and has been extremely useful for linking physiology and pathology to the genome. However, approaches to genetically modify specific genes in the rat germ line remain relatively scarce. To date, one of the most efficient approaches for generating genetically modified rats has been the target-selected *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis-based technology. Here, we describe the detailed protocols for ENU mutagenesis and mutant retrieval in the rat model organism.

INTRODUCTION

The availability of technology that allows for introducing targeted genetic modifications in model organisms has greatly contributed to our understanding of specific gene function. In the mouse, homologous recombination in embryonic stem (ES) cells has proven to be a powerful tool for generating genetic knockouts [1]. Although the laboratory rat (*Rattus norvegicus)* is one of the preferred model organisms in

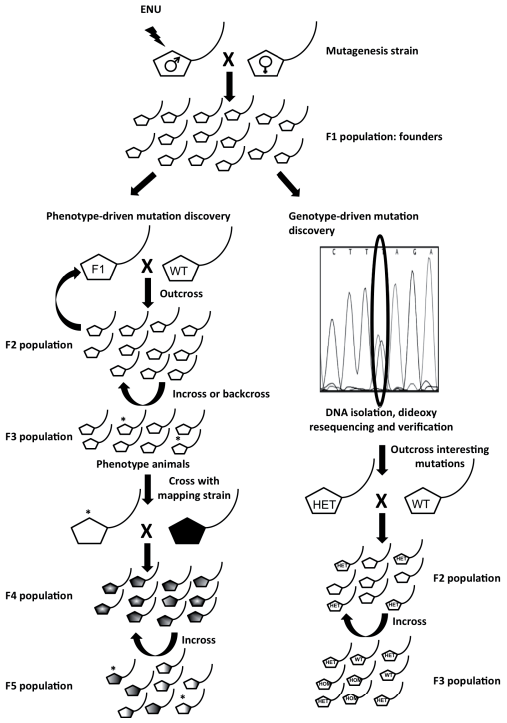

Figure 1: Schematic overview of the ENU mutagenesis approach. By crossing mutagenized males with untreated females a F_1 population, in which all individuals carry unique random ENU-induced mutations, is generated. Depending on the research interest, both genotype-driven as phenotype-driven approaches can be applied for mutation discovery. For a phenotype-driven approach the F_1 animals are crossed out with an untreated animals of the same strain (or alternatively of a mapping strain). The animals of the F_2 generation are subsequently backcrossed (for high retrieval of homozygous mutations) with the F_1 founder or crossed with brother-sister mating. The animals of the F_3 generation will carry random ENU-induced homozygous mutations and can be assessed for the phenotype of choice. For mapping, an animal that displays the phenotype of choice is crossed with an animal of a mapping strain. Subsequently, the progeny is crossed with brother-sister mating in order to generate unaffected (wild-type and heterozygous mutant) and affected (homozygous mutant) animals, which can be used for mapping using linkage analysis. For a genotype-driven approach DNA is taken from the F_1 animals, which is screened for heterozygous ENU-induced mutations in genes of interest. After verifying mutations of interest in an independent PCR and sequencing reactions, the F_1 animals carrying the mutations are crossed with untreated animals of the same strain. Homozygous mutant animals are obtained by crossing heterozygous mutant animals.

physiological and pharmacological research, gene knockout technology using homologous recombination approaches is still not available, because of the lack of pluripotent embryonic stem cells [2]. A successful alternative approach for the production of knockout rats is using an ENU mutagenesis-based technology. An advantage of this approach is that the created mutants are not 'transgenic' of nature, since no artificial DNA construct is integrated into the genome. And the technology does not require special (ES) cell lines and/or advanced oocyte or embryo manipulation, since it is based on the mutagenic property of the germ line mutagen ENU, which is applied *in vivo* [3]. The ENU mutagenesis-based knockout technology protocol follows a simple outline: ENU mutagenesis of male founder rats, F_1 library generation by mating with untreated females, mutation discovery using a platform of interest, and isolating an induced mutation into the desired genetic background (**Fig. 1**). The effectiveness of the technique depends on the efficiency of the mutagenesis and the mutation discovery methodology.

MATERIALS

Reagents and equipment

1. Safety wear for protection during ENU solution preparation and animals treatment: lab coat, gloves, mouth mask, and goggles.
2. One-way absorption paper and 0.1 M sodium hydroxide (NaOH) solution.
3. Syringes of 10 ml and 50 ml, needles of 21G, filters (∅ 0.2 µm) disposable cuvettes and parafilm.
4. Ethanol (100%).
5. Phosphate-citrate buffer: 0.1 M NaH_2PO_4, 0.05 M citric acid and set pH at 5.0 with phosphoric acid. The buffer is filter sterilized.
6. Weigh scale.
7. Spectrophotometer.

ENU stock solution preparation

1. Roughly 1 hour prior to the planned injections the ENU bottle (IsoPac) is unpacked and the metal lid is removed (*see* **Note 1**).
2. Using a syringe 5 ml of 100% ethanol is injected and the solution is shaken vigorously (*see* **Note 3**).
3. The ENU is dissolved in 95 ml of the phosphate citrate buffer by slowly adding the solution using a syringe and allowing for depressurization, followed by shaking vigorously for approximately 5 minutes.
4. The ENU solution is filter sterilized (∅ 0.2 µm).
5. The concentration of the ENU is determined by measuring the optical density (OD) using a spectrophotometer at 395 nm wavelength of a 1:10 dilution in the phosphate citrate buffer of the ENU stock solution. Subsequently, the concentration is calculated by assuming that 1 OD unit equals a concentration of approximately 1 mg/ml. (*see* **Note 4**).
6. ENU solutions should always be transported in a closed vial. A 0.1M NaOH solution should always be within reach to quickly neutralize any spillages (*see* **Note 2**).
7. Dissolved ENU should be used within one hour after preparation. Do not store dissolved ENU (*see* **Note 5**).

Waste removal

All disposals should be soaked in 0.1M NaOH for a few minutes before discarding. All waste should be designated as hazardous and removed by a central waste facility. The working areas

that were used for preparing the ENU stock solution and for injections should be cleaned with 0.1M NaOH.

DNA isolation and mutation discovery by dideoxy resequencing

1. Instruments and tubes/deep-well plates for tissue collection.
2. Tissue lysis buffer: 100 mM Tris-HCl (pH 8.5), 200 mM NaCl, 0.2% SDS (w/v), 5 mM EDTA and 100 µg/ml of freshly added Proteinase K.
3. Chemicals: Phenol:chloroform (1:1), isopropanol, 70% ethanol and sterile 10 mM Tris-HCl (pH 8.0).
4. Primers designed for a nested PCR amplifying exons of genes of interest.
5. 5x PCR buffer: 25 mM of tricine, 7.0% glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$ and 85 mM NH_4Ac, pH 8.7 with 25% ammonia (w/v).
6. Taq polymerase for PCR amplification.
7. dNTPs for PCR amplification.
8. PCR machine with 96- and/or 384-well blocks.
9. Standard gel electrophoresis unit, agarose and ethidium bromide.
10. Sequencing chemicals: BigDye (v3.1; Applied Biosystems) and Sanger BigDye Dilution Buffer version 2 (SBDDv2; Applied Biosystems). BigDye mix should be kept at –20°C and in the dark until use.
11. Ice-cold 80% ethanol and precipitation mix: 80% ethanol (w/v) supplemented with 40 mM NaAc.

METHODS

ENU mutagenesis

The mutagenicity of ENU results from the ability to transfer its ethyl group to oxygen or nitrogen radicals present in the DNA [4]. Subsequent replication cycles of this damaged DNA can result in mispairing followed by single base pair substitutions. Treatment of male rats with ENU will cause DNA alkylation and subsequently point mutations in the DNA of cells with a high turnover rate, like the spermatogonial stem cells. Because of the randomness of the mutagenesis every affected cell will contain a unique set of induced mutations, which will result in a genetically heterogeneous sperm cell population. To transfer the induced mutations through the germ line, the treated males are crossed with untreated females.

Theoretically, treatment with high doses of ENU will yield a high molecular mutation frequency and therefore a high chance of generating modified alleles of genes of interest. However, higher doses of ENU will also negatively affect fertility and even viability of the treated animals. Hence, the optimal dose is largely determined by the tolerance to the toxic effects of the mutagen. In general, the optimal dose will have to result in more than 25% fertile males to be able to efficiently generate a large F_1 population. It has been demonstrated in the mouse that two or three weekly administrations of low doses of ENU will yield higher mutagenic efficiency compare to a single high dose [5]. The optimal dose is strongly strain-dependent (**Table 1**), which could be a major consideration when selecting a strain to initiate an ENU mutagenesis experiment. Notably, outbred strains seem to have higher tolerance to the toxic effects of ENU when compared with inbred strains and yield higher mutation frequencies. The highest mutation frequency has been obtained with a Wistar-derived line, *Msh6*^{1Hubr} (**Table 1**).

Table 1: Differences in strain-dependent ENU tolerance

Strain / Line	Optimal dose (mg ENU/kg bodyweight)[a]	Sterility dose (mg ENU/kg bodyweight)[a]	Mutation Rate	Reference
BN	3 x 20	3 x 40	1 in 2.91 x 10^6 bp	[14]
F344/Crl	3 x 40	3 x 60	1 in 1.76 x 10^6 bp	[14]
F344/NHsd	2 x 60	2 x 75	1 in 29 pups[b]	[12]
LEW	Unknown	3 x 20	Unknown	[14]
WF	Unknown	2 x 50	Unknown	[12]
WKy	Unknown	2 x 50	Unknown	[12]
Hsd:SD	2 x 60	2 x 100	1 in 64 pups[b]	[12]
Wistar	3 x 40	3 x 60	1 in 1.24 x 10^6 bp	[14]
Msh6^{1Hubr} [d]	3 x 30	3 x 35	1 in 5.85 x 10^5 bp[c]	[8]

[a]Doses are listed as number of injections in weekly intervals. [b]For these strains the mutation rate is shown as the rate of appearance of phenotypically aberrant pups. [c]Mutation frequency was shown to decline in F_1 progeny that was generated more than 14 weeks after the last ENU. Bp, base pairs. [d]This mutant rat line is Wistar-derived.

This rat line lacks the MSH6 component of the mismatch repair (MMR) machinery and specifically fails to recognize single nucleotide mismatches [6]. In rodents, it has been shown that the MMR system accounts at least partially for the repair of ENU-induced mutations [7,8]. Interestingly, when males of this rat line were mutagenized, not only the mutation frequency was increased more than two-fold, but the mutation spectrum was also changed resulting in an increased chance of introducing a premature stop codon [8].

1. The volume of the ENU solution to be injected should be calculated for each male rat and follows the equation (optimal strain-dependent split dose of ENU * kg bodyweight) / (10 * measured OD). If the calculated volume of the ENU solution is too low for accurate injection, dilute the solution with phosphate-citrate buffer.
2. Prepare a room for the ENU treatment by covering the floor with the one-way absorption paper to absorb any spilled ENU.
3. Male rats of 11 – 12 weeks of age are mutagenized by intraperitonially (IP) injecting a pre-calculated volume of the ENU solution (*see* **Note 6**).
4. The ENU treatment is repeated weekly until the optimal dose is reached.
5. It is recommended to monitor the animals' health at least once a week after the last ENU injection by determining their bodyweight and general status. Temporal stabilization of body weight is normal, but a decrease in body weight after 1 week after injection could be a measure for taking that animal out of the study.
6. The fertility rate is determined in the first 10 weeks after the last injection as a surrogate for the efficiency of mutagenesis. Three weeks after the last injection

mutagenized males are crossed with untreated females and the progeny are counted. A reduction and regain of the fertility rate, as well as smaller litter sizes, in the first 10 weeks are indicative of an effective mutagenesis [9], as ENU is capable of killing the mature sperm population. F_1 pups sired in the first 10 weeks after mutagenesis are not suitable for screening (*see* **Note 7**).

F_1 library generation

After a full cycle of spermatogenesis (~60 to 70 days after first ENU injection) the ENU-treated males are crossed with untreated females to produce the F_1 progeny, which can be screened for mutations using phenotype-driven (forward genetics) or genotype-driven (reverse genetics) approaches. There is a high chance of retrieving animals that are genetically chimaeric, when animals are used that were born before 10 weeks after the last injection. This is probably because of ethyl-adducts that originates from mutagenized sperm in the fertilized oocyte, which can result in heterogeneous mutation fixation in different lineages.

In case of a genotype-driven approach a large living repository can be generated, which can subsequently be screened for mutations in genes of interest. However, if space is limited, breeding using a rolling-cycle model can be applied. In this model the F_1 animals are screened before weaning for a panel of genes of interest and only animals that carry interesting mutations are retained. Notably, male F_1 animals that will be discarded can be archived by freezing sperm and tissue samples. These permanent frozen libraries can be screened for mutations in genes of interest indefinitely and have been created for both mice [10] and rats [11].

Genotype-driven mutation discovery

After a F_1 population has been established, the mutations that affect protein function have to be retrieved. Depending on the research question, two different approaches can be used. Gene-driven screens are extremely useful for determining the function of specific genomic elements. Especially the availability of fully sequenced genomes has increased the popularity of the latter approach since prior knowledge about the gene sequences is required in order to amplify the target regions. DNA is taken from all individuals of the F_1 population derived from chemically mutagenized founders, and is screened for induced heterozygous mutations in pre-selected genes of interest. Occasionally, knockout-like alleles are generated by introducing a mutation in the open reading frame (ORF) or a splice donor/acceptor site of a gene of interest that will result in a premature stop codon and an absent or truncated protein product. Additionally, hypo- and hypermorphic alleles of the same gene could be generated by nonsynonymous mutations, which enable the study of gene-dosage effects and of amino acid residues important for protein-protein interaction or catalytic activity. Animals carrying interesting heterozygous mutations are identified using high-throughput mutation discovery methods, outcrossed and eventually bred to homozygosity.

The first knockout rats were identified using a yeast-based screening assay that specifically identifies mutations interfering with translation of the protein [12]. Although it has been shown that this assay is highly effective for identifying knockout-like alleles, it neglects mutations that potentially results in interesting amino acids substitutions and the method is relatively laborious and difficult to scale. Two other methods have been successfully applied to discover induced mutations in rat ENU mutagenesis experiments, namely CELI cleavage-mediated [13] and Mu transposase-based heteroduplex identification [11]. These methods rely on the discovery of mismatches that arise after denaturing and reannealing a DNA fragment containing a heterozygous mutation. Both methods have the potency to identify all mutation types, however, the observed moderate mutation rates suggest false negatives in the discovery. These problems are overcome by resequencing the genes of interest of animals of the F_1 population [14]. Resequencing is considered to be the golden standard for mutation discovery. This procedure is cost-effective and can be applied for small-scale experiments as well as easily be automated using standard liquid handling robotic systems for systematic and large-scale screenings.

DNA isolation

1. A tissue sample, like ear, toe or tail material, is collected from each F_1 animal for DNA extraction.
2. The sample is incubated overnight at 55°C in 400 µl of tissue lysis buffer, preferably under shaking or rotating conditions.
3. If tissue debris is present (which is likely from a tail clip), the samples should be centrifuged at maximal speed for 1 min and the supernatant transferred to a fresh tube. The pellet can be discarded.
4. 400 µl of phenol/chloroform (1:1, v/v) is added, and the mixture is vortexed vigorously for 2 min. followed by centrifugation for 3 min. at maximal speed. Subsequently, the aqueous layer is transferred to a new tube and this step is repeated. After the second centrifugation 300 µl of the aqueous layer is transferred to a new tube.
5. The gDNA is precipitated by adding 300 µl of isopropanol and by inverting the tube 10 times. The sample is centrifuged at 14,000 rpm, at 4°C for 20 min. The supernatant is removed and the pellet is washed with 100 µl of 70% ethanol. The sample is again centrifuged at 14,000 rpm, at 4°C for 5 min. and the supernatant is discarded. The sample is spun an additional minute and all remaining ethanol is carefully removed using a pipette.
6. The DNA is dissolved by adding 500 µl 10 mM Tris-HCl (pH 8.0) and incubating at 55°C for 10 min. with occasional vortexing and can be stored at -20°C.

High-throughput resequencing

This protocol assumes the use of standard 384 well plates and a GeneAmp® PCR system 9700 (Applied Biosystems), although other brands are expected to perform equally.

1. The first PCR reaction contains 5 – 10 ng of template DNA, 2 µl of 5x PCR buffer, 0.1 mM of each dNTP, 0.2 µM of each primer (*see* **Note 10**) and 0.2 units of Taq polymerase in a total volume of 10 µl and is carried out using a touchdown thermocycling program (94°C for 60 sec.; 15 cycles of 92°C for 30 sec., 65°C for 30 sec. with a decrement of 0.2°C per cycle and 72°C for 60 sec.; followed by 30 cycles 92°C for 30 sec., 58°C for 30 sec. and 72°C for 60 sec.; 72°C for 180 sec.).
2. After thermocycling the first PCR reaction is diluted with 20 µl of Milli-Q water and 1 µl is hatched into the second PCR mix (by pipetting or using 96 or 384 needle pin-replicators), which contains 1 µl of 5x PCR buffer, 0.1 mM of each dNTP, 0.2 µM of each primer and 0.1 units of Taq polymerase in a total volume of 4 µl. The second PCR is carried out using a standard thermocycling program (94°C for 60 sec.; 35 cycles of 92°C for 20 sec., 58°C for 30 sec. and 72°C for 60 sec.; 72°C for 180 sec.).
3. It is highly recommended to test several samples of the second PCR on a 1% agarose gel containing ethidium bromide for the presence of the correct amplification product before sequencing.
4. If the correct amplification product is present, the second PCR mix is diluted with 20 µl of Milli-Q and 1 µl is hatched into the sequencing mix, which contains 1.9 µl of sequencing buffer, 0.1 µl of BigDye v3.1 and 0.4 µM of sequencing primer in a total volume of 4 µl. The sequencing reaction is carried out using a specially designed thermocycling program (40 cycles of 92°C for 10 sec., 50°C for 5 sec. and 60°C for 120 sec.).
5. Before proceeding to capillary sequencing, the sequence fragments have to be purified using ethanol precipitation. To each well 30 µl of precipitation mix is added and the mixtures are vortexed vigorously and spun at maximal speed, at 4°C for 40 min. in a cooled plate centrifuge. The supernatant is discarded and the pellet is washed with 30 µl of ice-cold 80% ethanol. The samples are spun an additional 10 min., the supernatant is discarded and the samples are air-dried for approximately 15 min. preferably protected from light. Finally, the precipitate is dissolved in 10 µl of Milli-Q.
6. The plates are analyzed on the 96-capillary 3730XL DNA analyzer (Applied Biosystems) using the standard RapidSeq protocol on a 36 cm array.
7. The sequencing reads that belong to one amplicon are aligned, which facilitates the discovery of heterozygous point mutations (*see* **Note 11**).
8. All candidate mutations have to be verified in an independent PCR and sequencing reaction in order to exclude false positives.

Phenotype-driven mutation discovery

Phenotype-driven screens are excellent tools for gene discovery and to dissect developmental and biochemical pathways that underline a given phenotype. The challenge in these forward screens lies in assessing the phenotype of interest in a consistent and systematic manner, i.e. the phenotype should be clearly distinguishable

in affected compared with unaffected animals. Multiple phenotypes can be scored in rapid but broad primary screens [15], followed by more detailed, phenotype-specific follow-up screens. Mutations underlying the phenotype of interest are mapped by crossing the mutagenized genetic background into a mapping strain. Subsequently, known single nucleotide polymorphisms between the two strains are genotyped in order to locate the genomic locus containing the mutation.

The F_1 population carrying heterozygous mutations can directly be tested to screen for dominant phenotypes or the animals can be outcrossed followed by either a backcross to the F_1 founder or brother-sister incrosses (**Fig. 1**), resulting in F_3 animals with random homozygous ENU-induced mutations, which can be screened for recessive phenotypes [16]. In addition, mutations underlying a phenotype of choice can be screened for in a genome-wide fashion [17] or by applying a region-specific approach, like using a balancer chromosome in mice [18]. Here we will limit ourselves to explaining phenotype-driven screens for recessive ENU-induced mutations.

Breeding scheme

Recessive mutation discovery depends on crossing animals with heterozygous ENU-induced mutations and scoring the phenotype of interest in their progeny, which will have random homozygous mutations. After assessing these animals for the phenotype of choice the causative mutation has to be mapped by crossing affected animals to a mapping strain, followed by brother-sister incrosses of their progeny. These animals are then assessed for the phenotype and the causative mutation can be mapped using linkage analysis (**Fig. 1**). Alternatively, F_1 animals can be directly outcrossed with the mapping strain. In this way, the F_3 progeny that is assessed for the phenotype of choice can be directly used for mapping. The mapping strain is usually another, relatively far diverged strain and sufficient genomic information should be available of both strains in order to perform linkage analysis. The genomic loci containing the mutations that are underlying the phenotypes of interest can be mapped using linkage analysis, which is currently most efficiently performed using panels of single nucleotide polymorphism (SNPs) for which the different strains have been genotyped, as previously described [19,20].

1. Cross the F_1 animals with untreated animals of the same or, alternatively, a mapping strain of choice (*see* **Note 8**) to generate the F_2 generation.
2. An F_3 generation, carrying random homozygous ENU-induced mutations in their genomes, is subsequently generated by incrossing the F_2 animals or by backcrossing to the F_1 founder in order to increase the amount of random homozygous mutations.
3. The F_3 animals are screened for the phenotype of choice.
4. Affected F_3 animals are crossed with the mapping strain to produce an F_4 population in which all individuals are heterozygous for the causative mutation (*see* **Note 9**).
5. DNA is isolated of both affected and unaffected animals using the protocol described above.

Mapping

The genomic loci containing the mutations that are underlying the phenotypes of interest can be mapped using linkage analysis, which is currently most efficiently performed using panels of single nucleotide polymorphism (SNPs) for which the different strains have been genotyped [21,22], as previously described [19,20].

Outcrossing

After identifying interesting mutations or an interesting phenotype, it is important to maintain the line by outcrossing the heterozygous carriers with untreated animals of the same strain. In addition, these crosses are important to eliminate unwanted ENU-induced background mutations. Generally, six to ten backcrosses to the parental strain are considered to eliminate most of the background mutations. However, background mutations that are close to the mutation of interest are hard to evade and could be a potential problem. Nevertheless, the chance that a linked mutation landed in a genetic feature that attenuates the phenotype under study is small. Indeed, estimates concerning the impact of the background mutations, indicate that the potential problem should not be exaggerated [23]. When phenotyping animals there should always be a 1:1 relationship between phenotype and genotype, and littermates should always be included as controls. The ultimate control would be the generation of a second allele of the same gene, which would not carry the confounding background mutation and should verify the phenotype.

1. Animals of the F_1 population that harbor mutations of interest are crossed with untreated animals of the parental strain.
2. If the causative mutation is known the progeny are genotyped to determine if they inherited the induced mutation and the heterozygous animals are retained. If the causative mutation is not known the carriers in the progeny are discriminated by brother-sister incrosses and assessing the phenotype of interest in their offspring.
3. Further outcrossing of the mutation to the parental strain genetic background is performed by backcrossing heterozygous animals of each subsequent generation to the parental strain. In every outcrossing stage, homozygous animals can be generated by intercrossing two heterozygotes. This procedure also yields non-homozygous mutant littermates that could serve as controls in experiments.

CONCLUSIONS

Several rat knockout models for genes involved in human inherited diseases have already been made using ENU mutagenesis. In a forward genetics screen a nonsense mutation was identified in *Myo7a*, which resulted in a rat model for Usher syndrome type 1B, a severe autosomal inherited recessive disease that involves deafness and vestibular dysfunction [20]. In addition, characterization of the rat knockout models for the tumor suppressor genes *BRCA1, BRCA2, APC* and *MSH6*, which were

produced using reverse genetics approaches, demonstrated that the rat complements the equivalent mouse models for studying specific aspects of tumorigenesis, especially owing to its larger size, prolonged viability and ability to bear larger tumors [6,12,24,25]. The serotonin transporter (*SERT*) knockout rat model has been shown to exhibit a disturbed serotonin homeostasis [26]. Serotonin is evolutionarily the most ancient neurotransmitter. It is involved in a wide array of biological functions, such as emotion, motivation, and cognition, for which the rat is the preferred model organism to study. These examples illustrate that rat knockout models are a valuable addition to the toolkit to study the genetic basis of specific aspects of human health and disease.

NOTES

1. People who work with ENU should be aware of the genotoxic effects and should take some precautions beforehand. When preparing the ENU stock solution wear gloves, lab coat, goggles and a mask. ENU solution should be prepared in a high-flow chemical hood. Prepare a 0.1M NaOH solution to neutralize spillage (*see* **Note 2**).
2. Since the half-life of ENU decreases by increasing pH and is less than a minute at high pH alkaline solutions (0.1 – 1M NaOH) should always be in close proximity when handling ENU. Any ENU spillage should be poured with 0.1M NaOH and left stand for few minutes before cleaning.
3. It is recommended to thaw the ENU isopac before dissolving it, because it greatly facilitates dissolving the ENU.
4. The final concentration of the ENU stock differs between batch numbers and should always be determined by measuring the OD_{395nm}. Typically this concentration varies between 6 and 8.5 mg/ml. In ENU mutagenesis experiments in rodents, the concentration of ENU is generally measured at a wavelength of 395 nm, however, in other ENU experiments a wavelength of 238 nm might be used (e.g. zebrafish ENU mutagenesis). For consistency purposes it is recommended to use OD_{395nm}.
5. The half-life of ENU in the phosphate-citrate buffer, which is at pH 5.0 is approximately 80 hours. At pH 7.0 this decreases to 34 minutes, which would approximate the half-life of the chemical in the animal body. Therefore, we advise using the ENU solution for injection within 1 hour after preparation, because of its high instability when dissolved.
6. For safety reasons two persons should perform the ENU injections. The first person will restrain the animal and the other person will perform the injection of the ENU solution.
7. There is a high chance of retrieving chimaeric animals when animals are used that were born before 10 weeks after the last injection. This is probably because of ethyl-adducts that originates from mutagenized sperm in the fertilized oocyte, which can result in heterogeneous mutation fixation in different lineages.
8. Depending on the phenotype of interest and mutation frequency, another round of outcrossing with an untreated animal of the same genetic background of the mutagenized strain can be assessed in order to dilute the ENU-induced mutations.
9. If affected F_3 animals are unable to breed, a F_2 animal that was used to generate affected animals, can be used to cross with the mapping strain. However, since this F_2 animal is a heterozygous carrier of the causative mutation, by chance only 50% of the progeny will be heterozygous carrier of the mutation. Brother-sister incrosses have to be used to determine which individuals are the carriers.
10. The nested PCR setup described here assumes the use of primers that were designed with an optimal annealing temperature of 58°C. We make use of publicly available web-based information system called LIMSTILL, LIMS for Induced Mutations by Sequencing and TILLing (http://limstill.niob.knaw.nl), which allows for automated primer design.
11. For aligning and analyzing the sequence reads PolyPhred **[27]** software can be used, which automatically detects the presence of heterozygous single nucleotide substitutions. The

LIMSTILL software (*see* **Note 7**) also allows for automated annotation in order to determine the effect of the retrieved mutation.

REFERENCES

1. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6: 507-512.
2. Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM, et al. (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40: 516-522.
3. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, et al. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A 76: 5818-5819.
4. Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. Mamm Genome 11: 478-483.
5. Justice MJ, Carpenter DA, Favor J, Neuhauser-Klaus A, Hrabe de Angelis M, et al. (2000) Effects of ENU dosage on mouse strains. Mamm Genome 11: 484-488.
6. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
7. Claij N, van der Wal A, Dekker M, Jansen L, te Riele H (2003) DNA mismatch repair deficiency stimulates N-ethyl-N-nitrosourea-induced mutagenesis and lymphomagenesis. Cancer Res 63: 2062-2066.
8. van Boxtel R, Toonen PW, Verheul M, van Roekel HS, Nijman IJ, et al. (2008) Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics 9: 460.
9. Smits BM, Haag, J. D., Cuppen, E., and Gould, M. N. (2007) Rat Knockout and Mutant Models. In: Conn PM, editor. Sourcebook of Models for Biomedical Research. Totowa NJ: Humana Press Inc. pp. 171-178.
10. Coghill EL, Hugill A, Parkinson N, Davison C, Glenister P, et al. (2002) A gene-driven approach to the identification of ENU mutants in the mouse. Nat Genet 30: 255-256.
11. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
12. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
13. Smits BM, Mudde J, Plasterk RH, Cuppen E (2004) Target-selected mutagenesis of the rat. Genomics 83: 332-334.
14. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
15. Rogers DC, Fisher EM, Brown SD, Peters J, Hunter AJ, et al. (1997) Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mamm Genome 8: 711-713.
16. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A (1999) Mouse ENU mutagenesis. Hum Mol Genet 8: 1955-1963.
17. Hrabe de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, et al. (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25: 444-447.
18. Kile BT, Hentges KE, Clark AT, Nakamura H, Salinger AP, et al. (2003) Functional genetic analysis of mouse chromosome 11. Nature 425: 81-86.
19. Nelms KA, Goodnow CC (2001) Genome-wide ENU mutagenesis to reveal immune regulators. Immunity 15: 409-418.
20. Smits BM, Peters TA, Mul JD, Croes HJ, Fransen JA, et al. (2005) Identification of a rat model for usher syndrome type 1B by N-ethyl-N-nitrosourea mutagenesis-driven forward genetics. Genetics 170: 1887-1896.

21. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, et al. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 24: 381-386.
22. Saar K, Beck A, Bihoreau MT, Birney E, Brocklebank D, et al. (2008) SNP and haplotype mapping for genetic analysis in the rat. Nat Genet 40: 560-566.
23. Keays DA, Clark TG, Flint J (2006) Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. Mamm Genome 17: 230-238.
24. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
25. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.
26. Homberg JR, Olivier JD, Smits BM, Mul JD, Mudde J, et al. (2007) Characterization of the serotonin transporter knockout rat: a selective change in the functioning of the serotonergic system. Neuroscience 146: 1662-1676.
27. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25: 2745-2751.

3

LACK OF DNA MISMATCH REPAIR PROTEIN MSH6 IN THE RAT RESULTS IN HEREDITARY NON-POLYPOSIS COLORECTAL CANCER-LIKE TUMORIGENESIS

Ruben van Boxtel[1], Pim W. Toonen[1], Henk S. van Roekel[1], Mark Verheul[1], Bart M. G. Smits[1], Jeroen Korving[1], Alain de Bruin[2] and Edwin Cuppen[1]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2]Utrecht University, Faculty of Veterinary Medicine, Department of Pathobiology, Yalelaan, 3584 CL Utrecht, The Netherlands.

Adapted from Carcinogenesis (2008) 29(6):1290-7

ABSTRACT

To understand genetic instability in relation to tumorigenesis experimental animal models have proven very useful. The DNA mismatch repair (MMR) machinery safeguards genomic integrity by repairing mismatches, insertion/deletion loops and responding to genotoxic agents. Here, we describe the functional characterization of a novel rat mutant model in which the mismatch repair gene *Msh6* has been genetically inactivated by N-ethyl-N-nitrosourea (ENU)-driven target-selected mutagenesis. This model shows a robust mutator phenotype that is reflected by microsatellite instability and an increased germ line point mutation frequency. Consequently these rats develop a spectrum of tumors with a high similarity to atypical hereditary non-polyposis colorectal cancer in humans. The MSH6 knockout rat complements existing models for studying genetic instable tumorigenesis as it provides experimental opportunities that are not available or suboptimal in current models.

INTRODUCTION

The mismatch repair system (MMR) recognizes mismatches that are introduced during DNA replication and surpass the proofreading activity of DNA polymerase. The MMR machinery has been highly conserved in evolution, and the prototypic system was first elucidated in Escherichia coli [1]. In this model a mismatch is recognized by the MutS homodimer followed by binding of the MutL homodimer and activation of the repair pathway. This induces the excision and re-synthesis of the error containing DNA strand. In eukaryotes MutS function is separated in the recognition of single-base mismatches and small insertions or deletions loops (IDL's) of one or two extra helical nucleotides by MutSα and the recognition of larger IDL's by MutSβ. Both are heterodimers consisting of MSH2 - MSH6 and MSH2 - MSH3, respectively [2].

Besides maintaining the genomic integrity during replication, the MMR system has been shown to mediate DNA damage-induced apoptosis. Cell lines deficient for MMR are more resistant to cell death induced by alkylating agents [3,4], antimetabolites and intrastrand crosslinking agents [5]. Two hypotheses have been proposed to explain MMR-mediated apoptosis. The repeated activation of MMR because of wrong nucleotide insertion by polymerase at the site of DNA damage may cause double stranded breaks, which ultimately results in lethality [6]. Alternatively, the MMR system may function as a molecular sensor that can directly activate the apoptotic machinery in case of high levels of DNA damage [7].

In humans, mutations in MMR genes have been linked to hereditary non-polyposis colorectal cancer (HNPCC), also referred to as Lynch syndrome, which is characterized by early-onset colon cancer [8]. Mutations in the *MSH6* gene have been associated with an atypical HNPCC phenotype characterized by a late onset and a high occurrence of extracolonic tumors, especially in the endometrium [9]. A hallmark of MMR deficiency in HNPCC tumors is a significantly higher mutation rate, like the instability of simple repeat lengths during DNA replication [10]. This phenomenon is referred to as microsatellite instability (MSI) and has proven to be a powerful tool to diagnose HNPCC tumors [11].

Mouse knockout models have provided insight about in vivo deficiency of MMR function. Complete loss of MMR recognition in the *msh2*$^{-/-}$ mouse results in a strong reduction in the survival of these mice with a median survival time of 5 – 6 months [12]. At an early stage in life these mice develop lymphomas followed by a later onset of adenocarcinomas in the gastrointestinal tract [12,13]. Inactivation of the *Msh6* gene in mice also results in a decreased life span, however, this survival seems to be prolonged compared with *msh2*$^{-/-}$ mice. *Msh6*$^{-/-}$ mice show a median survival time of 6 – 10 months and predominantly develop lymphomas and rarely intestinal tumors [14,15]. Deletion of the *Msh3* gene in mice did not induce a cancer phenotype, however, deletion of this gene in a MSH6 deficient background led to a cancer phenotype indistinguishable from that of the *msh2*$^{-/-}$ mouse, suggesting redundancy between MSH6 and MSH3 [14]. Although mouse knockout models have provided very valuable insight in mammalian

MMR, modeling human HNPCC has only been successful in a limited way. First, heterozygous MMR mouse knockouts do not develop tumors and secondly, the spectra of tumors that develop in homozygous knockouts differ from the human situation. The human HNPCC phenotype primarily shows colorectal tumors, whereas the mouse models develop primarily lymphomas and tumors in the small intestine [12,15]. Furthermore, the high occurrence of endometrial cancers found in human atypical HNPCC, especially in the case of *MSH6* germ line mutations [9], is not reflected by the mouse model. Hence, novel models have the potential to complement exiting model systems.

Over the last decades, the rat has become an important genetic model system [16,17]. The complete genome sequence of the rat is now available [18] and recently developed techniques make it possible to generate knockout rats using a target-selected N-ethyl-N-nitrosourea (ENU)-driven mutagenesis approach [19,20]. There are indications that some human cancer syndromes, which have been difficult to model in genetically engineered mice, are more closely reflected in rats. One example is the recent report of the Pirc rat that carries a knockout allele of the *Apc* gene. The mouse knockout model, Apc^{min}, which has been studied for decades, develops primarily tumors in the small intestine [21], whereas the rat knockout primarily develops colonic tumors [22], similar to human patients [23]. Another example is the inability to study the loss-of-function of BRCA1 or BRCA2, which are often mutated in human breast tumors [24], because knocking out these genes in mice negatively affected viability. It turned out to be possible to generate viable *Brca2* knockout rats that develop tumors, although no increased incidence for breast tumor development was observed [20,25]. Nevertheless, this rat provides a unique model to study the in vivo role of this gene in tumorigenesis.

Here, we describe the genetic inactivation of the *Msh6* gene in the rat and the phenotypic consequences. A premature stop codon in the *Msh6* gene, introduced using an ENU-driven target-selected mutagenesis approach [19], was shown to result in a full functional knockout of the gene. Primary characterization of this $msh6^{-/-}$ rat revealed a strong mutator phenotype resulting in a reduced life span because of the development of different types of tumors, including lymphomas and endometrial cancers.

RESULTS

Generation of the MSH6 knockout rats

In a large ENU-driven target-selected mutagenesis screen we identified a rat carrying a heterozygous mutation in the *Msh6* gene [19]. This T to A transition in exon 4 resulted in a premature stop codon at position 306 (**Fig. 1A**). After outcrossing the mutant rat, heterozygous offspring was used to obtain homozygous mutant rats, which occurred in a Mendelian fashion (Data not shown). The homozygous mutants showed normal growth and fertility. Confirmation of the *Msh6* gene knockout phenotype at the molecular level was established by Western blotting, showing that the full-length

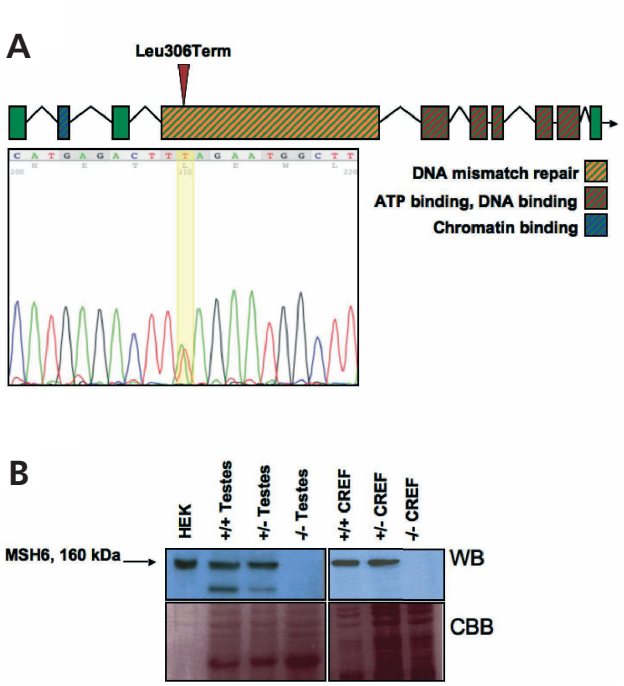

Figure 1: Molecular characterization of the *msh6*$^{-/-}$ rat. (A) Genomic organization of the *Msh6* gene in the rat genome. The arrow marks the position of the ENU-induced mutation that results in a premature stop codon. The sequence trace indicates the T to A transition in a heterozygous rat. (B) Western blot (WB) of testes and cultured rat embryonic fibroblast (CREF) lysates stained with antibodies against human MSH6 (160 kDa). Lysate of human embryonic kidney (HEK) cells was loaded as a positive control. MSH6 protein is completely absent in *msh6*$^{-/-}$ testes and CREF. Coomassie brilliant blue staining (CBB) was used as a control for protein loading.

MSH6 protein of approximately 160 kDa in the testes of *msh6*$^{+/+}$ and *msh6*$^{+/-}$ males was completely lacking in the testes of *msh6*$^{-/-}$ males (**Fig. 1B**). We could not detect any truncated protein, suggesting that either the mRNA with the premature stop codon or the truncated protein is unstable, because of nonsense-mediated decay or protein folding defects, respectively. Cell lysates of cultured rat embryonic fibroblasts also showed complete absence of MSH6 in homozygous mutant cultures, further confirming these results.

MSH6 knockout rats show a germ line mutator phenotype

To confirm loss of MSH6 function, we analyzed the presence of MSI in the germ line of *msh6*$^{-/-}$ males. Two mononucleotide repeats, $(G)_{20}$ and $(A)_{30}$, and two dinucleotide repeats, $(CA)_{36}$ and $(CA)_{40}$, were analyzed in the outcrossed offspring of 9 *msh6*$^{-/-}$ males and 19 *msh6*$^{+/+}$ males. In both mononucleotide repeats (**Fig. 2A**) and dinucleotide repeats (**Fig. 2B**) MSI was observed as a mono-allelic change of the length of the tested repeat. Although for some of the repeats the paternal (-/-) and maternal (+/+) contribution could not be distinguished (**Fig. 2A**), in other cases the size of the repeat was polymorphic within the strain and showed that the MSI phenotype is exclusively contributed by the paternal allele on an *msh6*$^{-/-}$ background (**Fig. 2B**). In total, 168 progeny from 9 homozygous mutant males were tested and revealed that germ line MSI occurred in all males and in all the repeats tested (**Table 1**). None of the 100 offspring samples from 19 control wild-type males and females showed MSI.

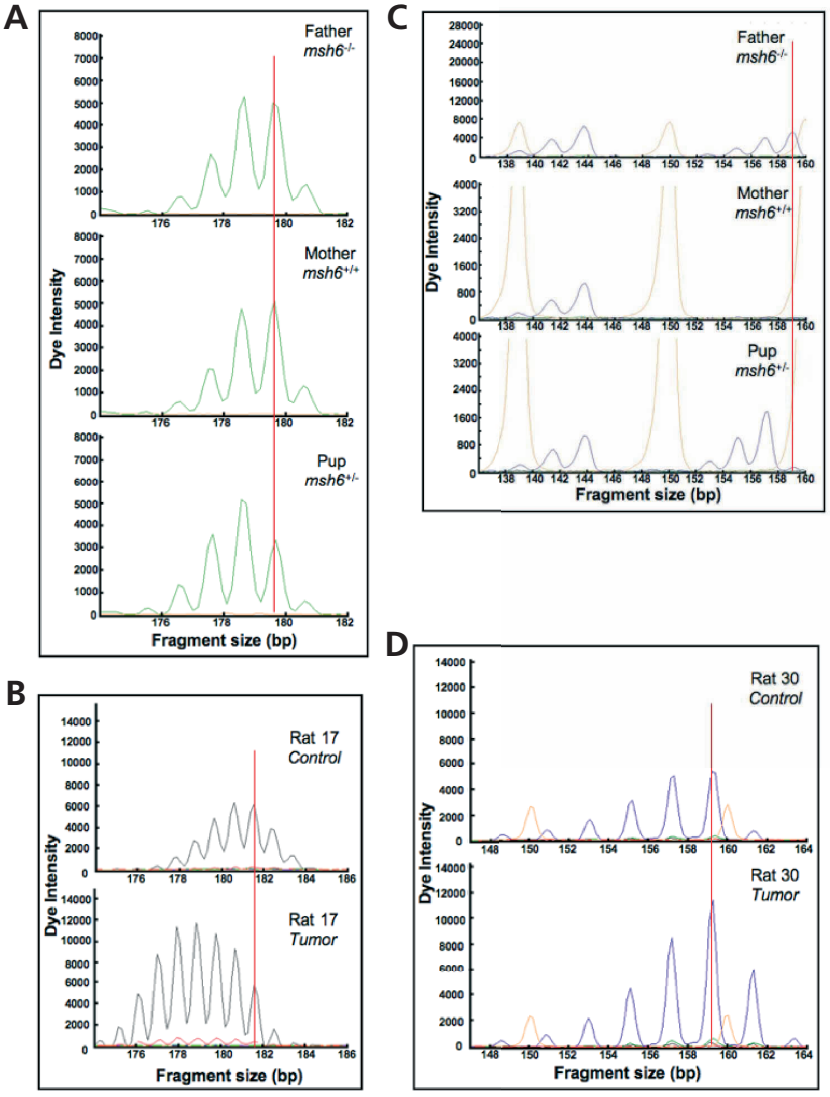

Figure 2: Microsatellite instability in the germ line and tumors. (A) A mononucleotide repeat of 20 repetitive subunits $(G)_{20}$ was analyzed in a cross between a *msh6*$^{-/-}$ father and a wild type mother and their *msh6*$^{+/-}$ offspring. The DNA of the parents and pups was isolated from ear and tail cuts, respectively. The red line indicates the size of the most prominent amplification product in both parents. In the DNA sample of the offspring one of the alleles has lost one repetitive subunit. (B) Fragment analysis of a dinucleotide repeat $(CA)_{36}$ that is polymorphic between the founder animals. As a result, progeny is heterozygous for this marker and the depicted F_1 animal shows a deletion of one repetitive subunit in the larger allele, which was inherited from the *msh6*$^{-/-}$ father. (C) Mononucleotide repeat instability in tumors. Three repetitive subunits in a mononucleotide repeat of 30 repetitive subunits $(A)_{30}$ were deleted in the DNA of tumor tissue (mediastinum) as compared with control DNA (ear cut). (D) Dinucleotide repeat instability in the DNA of tumor tissue (mediastinum) compared with control DNA (liver). The orange peaks represent the size marker that was loaded to determine the size of the PCR product.

Table 1: MSI in the germ line of MSH6 deficient rats

Father (genotype)[a]	$(CA)_{36}$	$(CA)_{40}$	$(G)_{20}$	$(A)_{30}$	Pups tested[b]
F (-/-)	1	3	2	0	17
G (-/-)	1	2	3	1	24
HA (-/-)	0	1	1	0	18
I (-/-)	0	1	2	0	9
J (-/-)	2	0	1	3	22
K (-/-)	0	2	1	0	14
L (-/-)	1	0	0	2	3
M (-/-)	0	2	1	0	14
N (-/-)	2	3	1	3	47
Average MSI % of total pups (± SEM)	6% (± 4%)	9% (± 2%)	8% (± 2%)	10% (± 7%)	Total pups tested = 168
Control (+/+) n = 19	0	0	0	0	Total pups tested = 100

[a]DNA was isolated from ear cuts. [b]DNA was isolated from tail cuts. MSI, microsatellite instability.

Because the MSH6 protein is also involved in the recognition of single-base mismatches introduced during DNA replication, we anticipated finding an elevated spontaneous germ line mutation frequency. Indeed, whereas the mammalian germ line mutation frequency is estimated around 1 x 10^{-8} bp^{-1} per generation [26], high-throughput resequencing of preselected amplicons, followed by heterozygous mutation discovery in DNA samples from the offspring of 2 *msh6*$^{-/-}$ males and wild-type females, revealed 3 mutations in 8.3 x 10^6 resequenced base pairs (twice a C>T and once a G>A transition; **Table S1**). This indicates a ~30-fold increased spontaneous single base pair mutation rate of about 3.6 x 10^{-7} bp^{-1} in the male germ line of the *msh6*$^{-/-}$ rat. As a control, the same amplicons were sequenced in DNA samples from the offspring of 4 wild-type males and no mutation was found in the 6.6 x 10^6 base pairs, indicating that the wild type mutation rate in the genetic background used is lower than 1.5 x 10^{-7} bp^{-1}.

MSH6 knockout rats show a reduced life span and develop tumors

Studies in the mouse have shown that MSH6 deficiency results in a reduced life span because of the development of different types of tumors [14,15]. Homozygous mutant rats showed a median survival time of 14 months, whereas 95% of wild type rats normally survive at this age (**Fig. S1**). After 18 months all the *msh6*$^{-/-}$ rats had become moribund, showing a tumor incidence of ~88%. These results are consistent with the observation that MSH6-deficient mice show a reduced life span although they exhibited a median survival time of only 6 – 10 months [14,15].

Each rat was subjected to a complete necropsy and histopathological analysis. The first tumors were detected at 9 months of age in different locations and the first affected rats were males. Representative histological specimens of the tumor spectra that were observed are shown in Figure 3 and the complete data are summarized in Table 2. Macroscopical examination revealed a significantly enlarged spleen, which was the result of infiltration of lymphomas. In total 8 out of 17 rats developed highly invasive lymphomas in the spleen, liver, kidney, lung and mediastinum (**Fig. 3**). Histological analysis of the lymphomas revealed that neoplastic lymphocytes were organized in sheets separated by fine fibrovascular stroma. The round cells were uniform and of medium size with scant cytoplasm and round to ovoid nuclei containing fine chromatin and occasionally a prominent nucleolus located centrally (lymphoblastic lymphoma). Sheets of neoplastic cells infiltrate into the surrounding tissue and occasionally into vascular structures. Multiple areas within the neoplasm show apoptosis characterized by pyknotic and fragmented nuclei and bright eosinophilic cytoplasm (starry sky pattern). To ascertain the cellular origin of the lymphomas, sections of tumors were

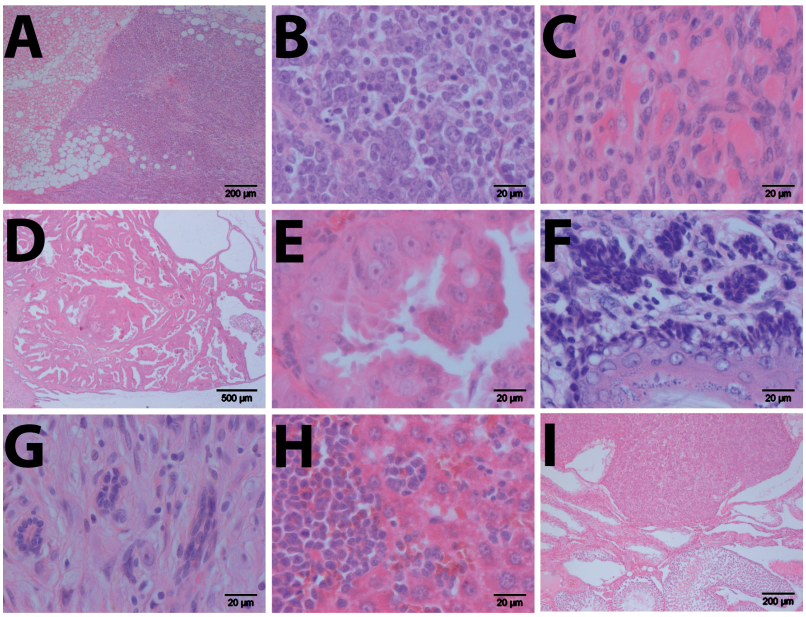

Figure 3: Tumors observed in the *msh6*$^{-/-}$ rat. (A) Section of a mediastinal lymphoblastic lymphoma of an *msh6*$^{-/-}$ male rat stained with HE. (B) A higher magnification of the lymphoblastic lymphoma. (C) Section of a leiomyosarcoma of the endometrium stained with HE. (D) Section of an endometrial carcinoma stained with HE. (E) A higher magnification of the endometrial carcinoma. (F) A section of a squamous cell carcinoma in the stomach of an *msh6*$^{-/-}$ female rat stained with HE. (G) Section of a fibroadenoma of a *msh6*$^{-/-}$ male rat stained with HE. (H) Section of a liver with lymphoblastic lymphoma leukemia stained with HE. (I) Section of a testicle containing a Leydig cell tumor stained with HE.

Table 2: Tumor spectrum of the MSH6 knockout rat

Rat ID (Genotype)	Gender	Age (months)	Tumor type	Tumor size[a]	Involvement	MSI	
						Mononucleotide repeat[b]	Dinucleotide repeat[c]
17 (-/-)	M	18	B-cell lymphoblastic lymphoma	~3 cm	Mediastinum	Yes	Yes
25 (-/-)	M	11	B-cell lymphoblastic lymphoma	N/D	Spleen, Liver, Kidney, Lung	No	No
30 (-/-)	F	12	B-cell lymphoblastic lymphoma	~3 cm	Mediastinum	Yes	Yes
42 (-/-)	M	17	B-cell lymphoblastic lymphoma	N/D	Bone marrow, mediastinum, liver	Yes	Yes
47 (-/-)	F	15	Squamous cell carcinoma	~2 cm	Stomach	No	No
49 (-/-)	F	12	Leiomyo-carcinoma	~5 cm	Uterus	No	No
52 (-/-)	M	18	Leydig cell tumor	N/D	Testis	No	No
62 (-/-)	M	9	B-cell lymphoblastic lymphoma	N/D	Spleen, Liver	Yes	No
68 (-/-)	F	12	Endometrial carcinoma	~1.5 cm	Uterus	Yes	No
70 (-/-)	F	12	Endometrial carcinoma	~1 cm	Uterus	Yes	Yes
72 (-/-)	F	18	B-cell lymphoblastic lymphoma	~1 cm	Mediastinum	Yes	Yes
74 (-/-)	M	14	Fibroadenoma and adeno-carcinoma	~2 cm	Mammary gland	Yes	No
81 (-/-)	F	18	Endometrial carcinoma	~0.5 cm	Uterus	No	No
85 (-/-)	M	17	T-cell lymphoblastic lymphoma	~2 cm	Thymus	Yes	No
89 (-/-)	M	10	B-cell lymphoblastic lymphoma	N/D	Spleen, Liver	Yes	No

[a]Approximate tumor diameter estimated from slides. [b]Two mononucleotide repeats were tested: $(A)_{30}$ and $(G)_{20}$. [c]Two dinucleotide repeats were tested: $(CA)_{36}$ and $(CA)_{40}$. MSI, microsatellite instability; N/D, not done.

stained with CD79, a B-cell specific antibody and a CD3 antibody that is T-cell specific. Of these, 7 were determined to be B-cell lymphoblastic lymphoma and one a T-cell lymphoblastic lymphoma.

Other tumors found in male *msh6*$^{-/-}$ rats included a testicular Leydig cell tumor, and a mammary fibroadenoma and adenocarcinoma. Four females suffered from vaginal bleeding and histological evaluation revealed the presence of endometrial carcinomas (**Fig. 3D**) in three animals and uterine leiomyosarcoma in one female (**Fig. 3C**). One female developed a gastric squamous cell carcinoma (**Fig. 3F**). Notably, some of the tumors reached large sizes, like some of the lymphomas that had an estimated diameter in excess of 3 cm. Remarkably, the leiomyosarcoma in the uterus reached a diameter of ~5 cm and also the gastric tumor reached a diameter of ~2 cm (**Table 2**).

Tumors of *msh6*$^{-/-}$ rats show MSI

To confirm the presence of genetic instable tumorigenesis in the *msh6*$^{-/-}$ rat, we analyzed the tumors for MSI. The four simple repeats described above were PCR amplified using DNA that was isolated from tumor tissues or control tissues (ear cuts or liver samples) and the product lengths were compared. 10 out of 15 tumors showed length polymorphisms in at least one of the four examined simple repeats (**Table 2**). All 10 tumors displayed mononucleotide repeat instability (**Fig. 2C**) and 3 showed also dinucleotide repeat instability (**Fig. 2D**). Some microsatellites were found to be highly unstable in the tumors, illustrated by deletions of up to 7 repetitive subunits and the presence of bi-allelic changes (**Table S2**). As expected, the longest microsatellite that was studied, $(A)_{30}$, was found to be the most instable repeat.

DISCUSSION

Carcinogenesis is believed to result from the accumulation of genomic mutations that change cells characteristics, which eventually result in the ability to proliferate and form tumors. DNA can be damaged by cellular metabolites and environmental agents, like alkylating agents which are widespread environmental mutagens that alter the chemical structure of DNA and cause G – T mismatches after subsequent replication [27]. Such damage can be recognized by MMR and indeed, two out of the three spontaneous mutations found in the rat MSH6-deficient germ line are likely to be unrecognized C to T transitions. Furthermore, mutations are generated because DNA polymerases make mistakes, although the fidelity of DNA polymerases is extremely accurate [28]. The initiation of tumorigenesis resulting from genomic mutations can be explained by the mutator phenotype hypothesis - mutations in genes that function in the maintenance of genomic stability, like MMR-genes, increase the mutation rate leading to a cascade of gene inactivation [29]. Alternatively, carcinogenesis has also been explained by increased selective advantages of cells as a result of mutagenesis [30]. MutSα has been found to play a role in DNA damage-induced apoptosis [4]

and inactivation would result in a selective advantage. In this model, the associated genetic instability would only be a secondary effect. Both hypotheses are not mutually exclusive and could explain the onset of tumor development in the MSH6 knockout rat. Deletion of the MMR system will result in both the accumulation of DNA damage like frame-shifts and point mutations, causing a mutator phenotype, and a selective advantage. Indeed, large-scale mutation discovery in a large set of tumors revealed extremely elevated somatic mutation prevalence in MMR-deficient cancer types [31].

Using ENU-driven target-selected mutagenesis we have generated a rat knockout model of the MutSα component MSH6 [19], which shows all the features of a mutator phenotype. Firstly, instability of both mono- and dinucleotide repeats was observed in both the mutant germ line as in tumors, but not in the wild-type germ line, confirming the role of MSH6 in the recognition of small insertion/deletions loops. MSI can result in hypermutability of expressed genes because of out-of-frame mutations [32] and because repetitive sequences have also been found in the coding regions of the *MSH6* gene itself, it is thought to be responsible for loss of heterozygosity (LOH) in humans by mutating the wild-type allele [33]. In mice lacking MSH6, replication error phenotype was observed at mononucleotide repeats in small intestinal epithelial cells that carried a lacI mutational reporter [34], and instability of dinucleotide repeats was observed in tumors [15]. The frequency of MSI in the tumors of the *msh6*$^{-/-}$ rat that was found in this study is much higher as compared with that found in the mouse [35], which could be because of differences in the length of the simple repeats that were tested. Most tumors displayed MSI at $(A)_{30}$ (10 out of 15 tumors), but for a much smaller repeat $(G)_{20}$ MSI frequency is much lower (2 out 15 tumors). Testing longer simple repeats increases the sensitivity of the analysis without generating false-positives, which is illustrated by the absence of length polymorphisms in the wild-type germ line.

Secondly, accumulation of point mutations as a result of MMR-deficiency can also attribute to the mutator phenotype. Indeed, lack of MSH6 results in a higher spontaneous germ line mutation rate compared with the wild-type germ line. Although the test groups were relatively small, the chance of discovering three mutations in 8.3 x 10^6 bp, given a 'normal' mammalian germ line mutation frequency of 1 x 10^{-8} bp^{-1} per generation [26], is extremely small ($p = 0.0017$; Fisher's exact test). Because ENU was used to generate the *msh6*$^{-/-}$ rat it could be argued that these mutations are because of the effect of this mutagen in the founder animal. However, analysis of the DNA of the parents showed that all mutations are unique to the progeny and can thus be classified as de novo mutations. Furthermore, the lack of any observed spontaneous mutations in 6.6 x 10^6 bp in wild type littermates suggests the lack of any experimental biases. In mice, small intestinal epithelial cells from MSH6-deficient animals were shown to exhibit a 41-fold increase in lacI mutation frequency compared with wild-type cells [34], which is in line with the approximate 30-fold increase that we found in the rat germ line.

The observed genomic instability illustrates that the *msh6*$^{-/-}$ rat is a functional gene knockout at the molecular level. As expected, the accumulation of both point and out-of-frame mutations in somatic tissue eventually results in a reduced life span of the MSH6-deficient rat. Although this decrease in life span is also observed in mice lacking MSH6, the survival is considerably longer in the *msh6*$^{-/-}$ rat. Two different MSH6-deficient mice strains have been generated, both of which have targeted exon 4 of the gene (comparable to the premature stop codon in this report), but on different genetic backgrounds (Msh6^{tm1Rak} was generated in hybrids of C57Bl/6 and WW6 [15]; Msh6^{tm1Htr} was generated in hybrids of 129/OLA and FVB [14]). These models were found to display clear phenotypic differences. For example, the median survival of Msh6^{tm1Rak} mice was found to be 10 months [15], whereas that of Msh6^{tm1Htr} mice was only 6 months [14]. The reason why the *msh6*$^{-/-}$ rat shows a delayed onset of tumorigenesis (median survival of 14 months) when compared with the *msh6*$^{-/-}$ mouse models remains unclear. The median life span of wild type mice and rats do not differ significantly, suggesting that other species-specific characteristics could underlie the observed differences. As rats and mice have diverged for about 40 million years this is not unlikely. Alternatively, because of its size, the rat could be able to sustain the growth of malignancies longer, allowing tumors to reach relatively larger sizes before the animal becomes moribund. Both the prolonged viability as well as the capacity of large tumor sizes could allow for longitudinal surgical, chemo-preventive, and/or therapeutic studies of genetic instable tumorigenesis. Increased life span when compared with similar mouse models was also observed in the *Apc* mutant Pirc rat, where tumors could reach a diameter larger than 1 cm [22].

The predominant development of lymphomas in the MSH6-deficient rats is also observed in mouse strains lacking MSH6 or other MMR components [13,14,15,36]. The occurrence of human patients with biallelic MMR mutations is rare [37,38], but the cases that have been reported also frequently develop lymphomas, with a relatively early age of diagnosis. A reason for the specificity of lymphoma development can be a high proliferation and turnover rate of lymphocytes during early development. MMR genes *MHL1*, *MSH2* and *MSH6* play a role in class switch recombination [39,40,41,42] and together with a possible increased accumulation of point and out-of-frame mutations, this could rapidly lead to genetic instability and the development of malignant cells. The extensive involvement of the small intestine in the neoplastic phenotype of the Msh6^{tm1Rak} mice [15] was not observed in the rat and could potentially explain the shorter life span of MSH6-deficient mice. However, the Msh6^{tm1Htr} strain rarely develops intestinal tumors, but does show the highest reduction in life span [14]. Interestingly, the *msh6*$^{-/-}$ rat shows a high incidence of endometrial cancers (3 of 7 *msh6*$^{-/-}$ females), resembling the atypical HNPCC spectrum of tumors. Although cancer in the uterus in mice has been reported in the Msh6^{tm1Htr} strain (3 in 22 *msh6*$^{-/-}$ mice), the other MSH6-deficient strain completely lacked involvement of the uterus in their cancer phenotype [14,15], indicating a considerably higher occurrence in the rat. It has

been reported that in human families bearing an *MSH6* germ line mutation, the two primary cancers found are colorectal and endometrial cancers [9,43,44]. Furthermore, defects in MSH6 have been found to be relatively common in an unselected series of endometrial cancers [45].

There are two major differences between the rat and human cancer phenotypes. First, human HNPCC patient are typically heterozygous for MMR germ line mutations and loose the wild-type allele in somatic cells only. Heterozygous MMR mutant mice do not develop early-onset tumors, which may be explained by their shorter lifespan and smaller size [35,46]. As for mice, heterozygous MSH6 knockout rats do not develop tumors either (not shown). However, it should be mentioned that tumors in both human patients as well as the rodent models are MMR-deficient. Second, the *msh6*$^{-/-}$ rats do not develop colorectal tumors, but they do show a high incidence of endometrial tumors. It has been suggested that endometrial cancer represents the most common manifestation of HNPCC among female *MSH6* mutation carriers and that colorectal cancer can not be considered an obligatory prerequisite to define the syndrome [9].

Taken together, the MSH6 knockout rat described here complements existing models for studying DNA mismatch repair and its relation to human genetic instable tumorigenesis. Not only the extended range of phenotypic characteristics, but also the size of the animal, the prolonged viability and the ability to bear large sized tumors make this rat an attractable model for specific experimental manipulations, such as endoscopy and local irradiation experiments for studying and treating residual tumor cells.

MATERIALS AND METHODS

Animals. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Science according to the Dutch legal ethical guidelines. Experiments were designed to minimize the number of required animals and their suffering. The MSH6 knockout rat (Msh6^{1Hubr}) was generated by target-selected ENU-driven mutagenesis (for detailed description, see [19]). Briefly, high-throughput resequencing of genomic target sequences in progeny from mutagenized rats revealed an ENU-induced premature stop codon in exon 4 of the *Msh6* gene in a rat (Wistar/Crl background). The heterozygous mutant animal was outcrossed at least three times to eliminate confounding effects from background mutations induced by ENU. To obtain homozygous animals the heterozygous offspring were crossed in. At three weeks of age ear cuts were taken and used for genotyping. Genotypes were reconfirmed after experimental procedures were completed. Animals were housed under standard conditions in groups of two to three per cage per gender under controlled experimental conditions (12-h light/dark cycle, 21±1°C, 60% relative humidity, food and water ad libitum).

Genotyping. Genotyping was performed using the KASPar SNP Genotyping System (KBiosciences, Hoddesdon, UK) and gene specific primers (forward common, CAG TGG ACC CAC TAT CTG GTA; reverse a1, GAA GGT GAC CAA GTT CAT GCT CTT CTC TGG CTT AAG CCA TTC TA; reverse a2, GAA GGT CGG AGT CAA CGG ATT CTC TTC TCT GGC TTA AGC CAT TCT T). Briefly, a PCR was carried out using the optimal thermocycling conditions for KTaq (94°C for 15 min; 20 cycles of 94°C for 10 sec, 57°C for 5 sec, 72°C for 10 sec; followed by 18 cycles of 94°C for 10 sec, 57°C for 20 sec and 72°C for 40 sec; GeneAmp9700, Applied

Biosystems, Foster City CA, USA). The PCR reaction contained 2 µl DNA solution, 1 µl 4X Reaction Mix, 165 nM reverse Primer a1 and a2 and 412.5 nM of the common forward primer, 0.025 µl KTaq polymerase and 0.4 mM $MgCl_2$ in a total volume of 4 µl. Samples were analyzed in a PHERAstar plate reader (BMG Labtech) and data were analyzed using Klustercaller software (KBiosciences). All genotypes were confirmed in an independent reaction.

Western Blot analysis. Proteins were extracted by adding lysis buffer (1% SDS, 1.0mM sodium ortho-vanadate, 10mM Tris pH 7.4) to approximately 0.25 g of tissue or cultured cells. The protein was separated on a SDS gel (6% acrylamide gradient, Bio-Rad) and transferred to a nitrocellulose membrane. The membrane was incubated overnight at 4°C with a 1:100 dilution of a monoclonal mouse anti-human MSH6 antibody (BD Biosciences Pharmingen) in blocking buffer followed by an incubation for 2 hours with peroxidase-conjugated, anti-mouse IgG diluted 1:2500 in blocking buffer at room temperature. Protein bands were detected by using the enhanced chemiluminescence detection method (ECL, Amersham Biosciences). Cell lysate from human HEK cells was used as a control and total protein was measured by Coomassie Brilliant Blue staining.

MSI analysis. A simple repeat containing a repetitive stretch of $(CA)_{40}$ (chr14:85120966-85121141, RGSC 3.4) was PCR amplified using the following primers: 5'-6FAM- TTC AAC CAC AAT CTC GAC AG -3' (forward) and 5'- AGG CAT GAG TTC TGA GGT TC -3' (reverse); the simple repeat (contig) with the stretch of $(CA)_{36}$ (chr13:105939209- 105939433) was PCR amplified using the following primers: 5'-6FAM- TGG CAC AGG TGT TTA GTG TC -3' (forward) and 5'- TGC AGA AGA AAT GAG AGG TG -3' (reverse); for PCR amplifying the simple repeat containing a $(G)_{20}$ (chr3:105633672-105633851) 5'-VIC- CAT TCT GGA AGT GAC TGC TG -3' (forward) and 5'- TCC ACG ATA CTG CAA TTC TC-3' (reverse) were used and the simple repeat containing the stretch of $(A)_{30}$ (chr8:118654926-118655555) was PCR amplified using the following primers 5'-NED- GCC CTC TTC TGG TGT ATC TG -3' (forward) and 5'- AGC TTC ATC CGT TAG TGT GG -3' (reverse). The appropriate volume of PCR product was mixed with 0.5 µl of GeneScan™-500 LIZ™ size standard (Applied Biosystems) in 5 µl of Milli-Q, denaturated for 5 minutes at 95°C and subsequently run on a AB3730WL capillary DNA analyzer (Applied Biosystems). Product lengths were analyzed using Genemapper software (Applied Biosystems).

Point mutation analysis. 768 pre-selected amplicons were amplified using a nested PCR setup followed by dideoxy sequencing. 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 the presence of heterozygous mutations using PolyPhred [47] and in-house developed software. All candidate mutations were verified in independent PCR and sequencing reactions.

Analysis of tumors. Animals were scarified by CO_2/O_2 suffocation. Tumors, when found, and organs including the gastrointestinal tract, lungs, liver, kidneys, spleen and thymus were removed and fixed in phosphate buffered 4% formaldehyde. Representative tissues from the tumors and organs were processed and embedded in paraffin. All tissues were prepared for Hematoxylin and Eosin stain. Lymphomas were studied for immunotyping. After deparaffination the sections were heated with citrate buffer pH 6.0 for antigen retrieval. The endogen peroxidase activity was blocked with 1% H_2O_2 in methanol for 30 min. The slides were incubated with 10% normal serum for 10 min followed by an incubation with Mouse anti CD 79 (1:80,M7051,DAKO,,Denmark) or Rabbit anti CD3 (1:1500,CMC 365,Cell Marque Corp,USA) overnight at 4°C.After 30 min incubation with Horse anti Mouse/biotin (1:125,BA-2000,Vector,USA) or Goat anti Rabbit/biotin (1:250,E0432 DAKO) the slides were incubated with Avidin-Biotin peroxidase complex (PK-4000,Vector) for 30 min. Visualization was performed with 3, 3 diaminobenzidine solution (DAB,Sigma,USA) for 10 min. Nuclei were stained with Mayer's Haematoxylin. After dehydrazition, the slides were mounted with Eukitt. (Kindller GmBH&Co, Germany) Washing steps were preformed with PBS/Tween, antibody dilutions and ABC/PO complex were made with PBS.

For MSI analysis slides were deparaffinated and lysed overnight followed by phenol/chloroform (1:1, vol/vol) extraction. DNA was precipitated by adding 300 µl isopropanol, mixing and centrifuging for 20 min, at 21,000 g at 4°C. The supernatant was discarded and pellets were washed with 70% ethanol and dissolved in 50 µl 10 mM Tris-Cl (pH 8.0).

ACKNOWLEDGEMENTS

The authors thank I.J. Nijman for bioinformatic help, J. van der Belt for technical assistance and H. Feitsma for critically reading the manuscript.

REFERENCES

1. Lahue RS, Au KG, Modrich P (1989) DNA mismatch correction in a defined system. Science 245: 160-164.
2. Jiricny J (2006) The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7: 335-346.
3. Claij N, Te Riele H (2002) Methylation tolerance in mismatch repair proficient cells with low MSH2 protein level. Oncogene 21: 2873-2879.
4. Claij N, van der Wal A, Dekker M, Jansen L, te Riele H (2003) DNA mismatch repair deficiency stimulates N-ethyl-N-nitrosourea-induced mutagenesis and lymphomagenesis. Cancer Res 63: 2062-2066.
5. Yang G, Scherer SJ, Shell SS, Yang K, Kim M, et al. (2004) Dominant effects of an Msh6 missense mutation on DNA repair and cancer susceptibility. Cancer Cell 6: 139-150.
6. Karran P, Bignami M (1992) Self-destruction and tolerance in resistance of mammalian cells to alkylation damage. Nucleic Acids Res 20: 2933-2940.
7. Fishel R (1999) Signaling mismatch repair in cancer. Nat Med 5: 1239-1241.
8. Lynch HT, Smyrk T (1996) Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. Cancer 78: 1149-1167.
9. Wijnen J, de Leeuw W, Vasen H, van der Klift H, Moller P, et al. (1999) Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat Genet 23: 142-144.
10. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363: 558-561.
11. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, et al. (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58: 5248-5257.
12. Reitmair AH, Redston M, Cai JC, Chuang TC, Bjerknes M, et al. (1996) Spontaneous intestinal carcinomas and skin neoplasms in Msh2-deficient mice. Cancer Res 56: 3842-3849.
13. de Wind N, Dekker M, Berns A, Radman M, te Riele H (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell 82: 321-330.
14. de Wind N, Dekker M, Claij N, Jansen L, van Klink Y, et al. (1999) HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nat Genet 23: 359-362.
15. Edelmann W, Yang K, Umar A, Heyer J, Lau K, et al. (1997) Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell 91: 467-477.
16. Lazar J, Moreno C, Jacob HJ, Kwitek AE (2005) Impact of genomics on research in the rat. Genome Res 15: 1717-1728.
17. Smits BM, Cuppen E (2006) Rat genetics: the next episode. Trends Genet 22: 232-240.
18. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. (2004) Genome sequence of the Brown Norway

rat yields insights into mammalian evolution. Nature 428: 493-521.

19. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
20. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
21. Moser AR, Pitot HC, Dove WF (1990) A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science 247: 322-324.
22. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
23. Rowley PT (2005) Inherited susceptibility to colorectal cancer. Annu Rev Med 56: 539-554.
24. Hutchinson JN, Muller WJ (2000) Transgenic mouse models of human breast cancer. Oncogene 19: 6130-6137.
25. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.
26. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148: 1667-1686.
27. Warren JJ, Forsberg LJ, Beese LS (2006) The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. Proc Natl Acad Sci U S A 103: 19701-19706.
28. Kunkel TA, Bebenek K (2000) DNA replication fidelity. Annu Rev Biochem 69: 497-529.
29. Loeb LA, Loeb KR, Anderson JP (2003) Multiple mutations and cancer. Proc Natl Acad Sci U S A 100: 776-781.
30. Tomlinson I, Bodmer W (1999) Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. Nat Med 5: 11-12.
31. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, et al. (2007) Patterns of somatic mutation in human cancer genomes. Nature 446: 153-158.
32. Eshleman JR, Lang EZ, Bowerfind GK, Parsons R, Vogelstein B, et al. (1995) Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. Oncogene 10: 33-37.
33. Yamamoto H, Sawai H, Weber TK, Rodriguez-Bigas MA, Perucho M (1998) Somatic frameshift mutations in DNA mismatch repair and proapoptosis genes in hereditary nonpolyposis colorectal cancer. Cancer Res 58: 997-1003.
34. Mark SC, Sandercock LE, Luchman HA, Baross A, Edelmann W, et al. (2002) Elevated mutant frequencies and predominance of G:C to A:T transition mutations in Msh6(-/-) small intestinal epithelium. Oncogene 21: 7126-7130.
35. Wei K, Kucherlapati R, Edelmann W (2002) Mouse models for human DNA mismatch-repair gene defects. Trends Mol Med 8: 346-353.
36. Reitmair AH, Schmits R, Ewel A, Bapat B, Redston M, et al. (1995) MSH2 deficient mice are viable and susceptible to lymphoid tumours. Nat Genet 11: 64-70.
37. Bandipalliam P (2005) Syndrome of early onset colon cancers, hematologic malignancies & features of neurofibromatosis in HNPCC families with homozygous mismatch repair gene mutations. Fam Cancer 4: 323-333.
38. Whiteside D, McLeod R, Graham G, Steckley JL, Booth K, et al. (2002) A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots. Cancer Res 62: 359-362.
39. Ehrenstein MR, Neuberger MS (1999) Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. Embo J 18: 3484-3490.
40. Martomo SA, Yang WW, Gearhart PJ (2004) A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med 200: 61-68.
41. Schrader CE, Edelmann W, Kucherlapati R, Stavnezer J (1999) Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med 190: 323-330.
42. Schrader CE, Vardo J, Stavnezer J (2002) Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class

switching shown by sequence analysis of recombination junctions. J Exp Med 195: 367-373.
43. Charames GS, Millar AL, Pal T, Narod S, Bapat B (2000) Do MSH6 mutations contribute to double primary cancers of the colorectum and endometrium? Hum Genet 107: 623-629.
44. Wagner A, Hendriks Y, Meijers-Heijboer EJ, de Leeuw WJ, Morreau H, et al. (2001) Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. J Med Genet 38: 318-322.
45. Goodfellow PJ, Buttin BM, Herzog TJ, Rader JS, Gibb RK, et al. (2003) Prevalence of defective DNA mismatch repair and MSH6 mutation in an unselected series of endometrial cancers. Proc Natl Acad Sci U S A 100: 5908-5913.
46. Edelmann L, Edelmann W (2004) Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. Am J Med Genet C Semin Med Genet 129: 91-99.
47. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25: 2745-2751.

SUPPLEMENARY DATA

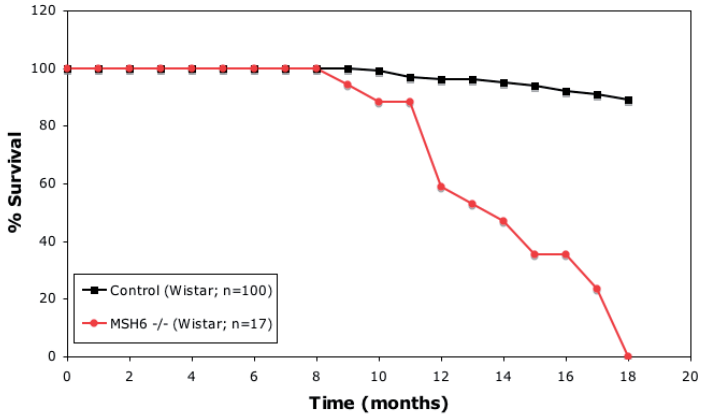

Figure S1: Survival of the *msh6*$^{-/-}$ rats. The time the rats became moribund is recorded. Red line, *msh6*$^{-/-}$ rats (n=17); black line, control group (n=100, Source Harlan).

Table S1: Spontaneous germ line mutation frequency

Genotype father[a]	*Msh6*$^{-/-}$ (n=2)	*Msh6*$^{+/+}$ (n=4)
Total base pairs resequenced	8,327,563	6,565,893
Number of mutations	3	0
Mutation frequency (bp^{-1})	3.6 x 10^{-7}	<< 1.5 x 10^{-7}

[a]DNA was isolated from ear or tail cuts

Table S2: Tumors of *msh6*$^{-/-}$ rats show MSI

	Microsatellite[b]			
Tumor sample[a]	**$(G)_{20}$**	**$(A)_{30}$**	**$(CA)_{37}$**	**$(CA)_{40}$**
17	-1	-2	+1	0
25	0	0	0	0
30	0	-1	+1	+1
42	0	-1	-7 / -1	+1
47	0	0	0	0
49	0	0	0	0
52	0	0	0	0
62	0	-3 / -1	0	0
68	0	-1	0	0
70	0	-1	+1	0
72	0	-3 / -1	+1	0
74	0	-1	0	0
81	0	0	0	0
85	-1	-5	0	+1
89	0	-1	0	0
Total[c]	2/15	10/15	5/15	3/15

[a]The rat ID is indicated. [b]The numbers indicate inserted/deleted repetitive subunits. [c]This number shows the number of instable tumors for the indicated microsatellite.

4

IMPROVED GENERATION OF RAT GENE KNOCKOUTS BY TARGET-SELECTED MUTAGENESIS IN MISMATCH REPAIR-DEFICIENT ANIMALS

Ruben van Boxtel, Pim W. Toonen, Mark Verheul, Henk S. van Roekel, Isaac J Nijman, Victor Guryev and Edwin Cuppen

Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

Adapted from BMC Genomics (2008) 9:460

ABSTRACT

Background. The laboratory rat *(Rattus norvegicus)* is one of the preferred model organisms in physiological and pharmacological research, although the availability of specific genetic models, especially gene knockouts, is limited. *N*-ethyl-*N*-nitrosourea (ENU)-driven target-selected mutagenesis is currently the most successful method in rats, although it is still very laborious and expensive.

Results. As ENU-induced DNA damage is normally recognized by the mismatch repair (MMR) system, we hypothesized that the effectiveness of the target-selected mutagenesis approach could be improved by using a MMR-deficient genetic background. Indeed, knockout rats were found to be more sensitive to ENU treatment and the germ line mutation rate was boosted more than two-fold to 1 mutation per 585 kb. In addition, the molecular mutation spectrum was found to be changed in favor of generating knockout-type alleles by ~20%, resulting in an overall increase in efficiency of ~2.5 fold. The improved effectiveness was demonstrated by high throughput mutation discovery in 70 Mb of sequence in a set of only 310 mutant F_1 rats. This resulted in the identification of 89 mutations of which four introduced a premature stop codon and 64 resulted in amino acid changes.

Conclusions. Taken together, we show that the use of a MMR-deficient background considerably improves ENU-driven target-selected mutagenesis in the rat, thereby reducing animal use as well as screening costs. The use of a mismatch repair-deficient genetic background for improving mutagenesis and target-selected knockout efficiency is in principle applicable to any organism of interest.

BACKGROUND

The rat is one of the most widely used model organism in biomedical research and has proven to be a powerful tool for linking physiology and pathology to the genome [1,2]. Selective breeding and characterization of strains, mimicking complex human diseases have led to hundreds of useful models. However, identification of the underlying causative polymorphisms and genes has shown to be difficult. An alternative approach in elucidating specific gene functions is by introducing targeted genetic modifications. Gene knockout technology using homologous recombination in embryonic stem cells has proven to be extremely powerful for this [3]. However, because of the lack of pluripotent embryonic stem cells for the rat, ENU-driven target-selected mutagenesis, also known as TILLING [4] has been one of the most successful methods for generating knockouts in the rat [5,6]. This approach does not require any special cell lines and/or advanced oocyte or embryo manipulation as male animals are injected with the alkylating agent ENU, which very efficiently introduces random point mutations in the DNA of spermatogonial stem cells. Upon crosses with untreated female animals, an F_1 generation is established in which each individual carries different random heterozygous point mutations in its genome. The DNA of these animals is subsequently screened in pre-selected genes of interest with the goal to identify mutations that affect protein function, e.g. by introducing a premature stop codon, by changing a consensus splice site residue, or by mutating critical (conserved) amino acids.

The overall efficiency of this method depends essentially on the mutagenicity of ENU, which was found to be both strain and dose dependent [5,6,7]. The maximum ENU-induced mutation rate in rats is approximately one point mutation every 1.25 – 1.5 Mb for Wistar rats treated with 35-40 mg ENU per kg bodyweight [5], which is similar to the highest mutation frequency that can be obtained in mice [8,9,10,11,12]. Higher doses of ENU result in sterility of the animals and at much higher doses (>120 mg/kg for rats) in lethality. However, in Arabidopsis and *C. elegans* chemically induced (EMS) point mutation frequencies are as high as 1 in 100 kb [13,14] and in zebrafish ENU-induced frequencies of 1 in 150-250 kb can be obtained routinely [15] (E.C., unpublished results), suggesting that the maximum mutation load in a vertebrate genome that is compatible with viability, is much higher than what is currently reached in both rat and mouse. Although about 10 rat knockouts were successfully generated by ENU-driven target-selected mutagenesis [5,6,16] the relatively low mutation frequency makes the target-selected mutagenesis procedure laborious and costly.

We hypothesized that the efficiency of the ENU-driven target-selected mutagenesis approach could potentially be increased in a MMR-deficient genetic background, as this system is involved in the response to and repair of ENU-induced genetic damage [17]. Single nucleotide mismatches, as well as small insertion and deletion loops (IDLs), are recognized by MutSα heterodimer, which consists of MSH2 and MSH6 [18]. Because MSH2 also functions in another heterodimer with MSH3, which

recognizes larger IDLs, MSH6-deficiency is in principle best suited to test the idea if the ENU mutagenesis efficiency can be increased in a MMR-deficient background. The MutSα subunit has already been implicated in the recognition of different forms of DNA damage caused by genotoxic agents [19]. Indeed, cell lines deficient in MutSα function that were treated with different alkylating agents, including ENU, showed a higher survival and increased mutation rate [17].

A rat carrying a premature stop codon in the *Msh6* gene has been identified in a previous knockout screen in our lab [5]. This mutation was shown to result in a complete loss of function [20] and, as a result, these animals show genetic instability as indicated by microsatellite instability (MSI), an elevated spontaneous germ line point mutation rate and tumorigenesis [20].

Here, we show that MSH6-deficiency improves the efficiency of the ENU-driven target-selected mutagenesis procedure in the rat by ~2.5-fold. This does not only result in a vast decrease of the number of animals needed, but equally reduces screening costs.

RESULTS

Effect of ENU in MSH6 knockout rat

The efficiency of ENU is both strain and dosage-dependent, which is reflected by very large differences in mutation frequency in different mouse [7] and rat [5,6] inbred strains and by a higher mutagenicity after three weekly administrations of low doses of the mutagen as compared with a single high-dose injection [7]. Higher ENU concentrations not only increase the mutation frequency, but also affect fertility [5,21]. To determine the optimal concentration in the *msh6*$^{-/-}$ rat (which is in Wistar background), we performed a dose-response experiment and determined the impact of the mutagen on fertility. At 12 weeks of age *msh6*$^{-/-}$ males were treated with 3 weekly doses of ENU ranging between 1 and 35 mg/kg bodyweight and fertility was monitored after a full cycle of spermatogenesis (~ 60-70 days). At concentrations of 35 and 30 mg/kg, respectively 1 out of 8 and 5 out of 8 males were fertile (**Table 1**). Previous studies showed that wild-type Wistar rats are less sensitive to ENU as 6 out of 10 and 10 out of 10 rats were fertile at 40 and 35 mg/kg ENU, respectively [5].

We also found that ENU affects the survival of *msh6*$^{-/-}$ rats. Untreated *msh6*$^{-/-}$ rats show a median survival of 60 weeks [20], while ENU treatment decreased survival of mutant males to an average mean of 37 ± 3 weeks (**Fig. 1**). This decrease in lifespan was found to be because of tumor development, mainly lymphomas (data not shown). It is crucial for performing ENU target-selected mutagenesis screens that animals are viable and healthy for at least 26 weeks to allow for mutagenesis and subsequent breeding. In previous studies, no reduction in lifespan was observed in both 40 mg/kg (10 out of 10 survived) and 35 mg/kg (9 out of 10 survived) ENU-treated wild-type males until 1.5 years of age (E.C., unpublished data).

Table 1: ENU-induced mutation frequencies

Genotype/Dose[a]	*Msh6*$^{-/-}$ 3 x 25	*Msh6*$^{-/-}$ 3 x 30	*Msh6*$^{-/-}$ 3 x 35	Total	WT Control 3 x 40[b]
# Males injected	6	8	8	24	10
# Fertile males[c]	N/D	5	1		6
# Pups for screening	16	291	3	310	362
# Bases screened	3.6 x 10^6	69.7 x 10^6	0.7 x 10^6	74.0 x 10^6	37.3 x 10^6
# Mutations					
- nonsense	0	4	0	4	0
- missense	4	59	1	64	17
- silent	0	16	0	16	7
- non-coding	0	5	0	5	6
- total	4	84	1	89	30
Mutation Frequencies (bp^{-1})	1.12 x 10^{-6}	1.21 x 10^{-6}	1.38 x 10^{-6}		8.05 x 10^{-7}
Mutation Rates (bp)	1 in 8.93 x 10^5	1 in 8.29 x 10^5	1 in 7.22 x 10^5		1 in 12.4 x 10^5

[a]Dose is indicated as mg of ENU per kg bodyweight (three weekly injections). [b]Wild-type (WT) control 3 x 40 mg/kg measurements were adapted from [5] and represent the highest ENU-induced mutation frequency in rats. [c]Fertility is measured by the presence of at least one litter within 10 weeks after the last injection. Sterility was confirmed by histological examination. N/D, not determined.

Mutation screening procedure

We have developed an automated, high-throughput and cost-effective dideoxy-based DNA resequencing protocol (~0.15 € for chemicals and disposables per sample) for the discovery of ENU-induced point mutations [5]. To reduce the costs of the knockout procedure further, we now established an efficient breeding and screening schedule in which the offspring of mutagenized animals are screened for the presence of interesting mutations before they are weaned (3 weeks of age). This rolling-circle model significantly reduces the amount of animal space needed.

At two weeks of age tissue samples from uniquely tagged F_1 offspring were taken and DNA was isolated. Subsequently, DNA samples were screened for 768 pre-selected amplicons by nested PCR amplification and sequencing using two sets of 384 wells plates, containing pre-gridded primers. Resequencing data for the same amplicons from different pups were automatically processed by PolyPhred [22] as well as semi-manually inspected using in-house developed software. All candidate mutations were verified in independent PCR and sequencing reactions and within one week interesting mutants could be selected and weaned.

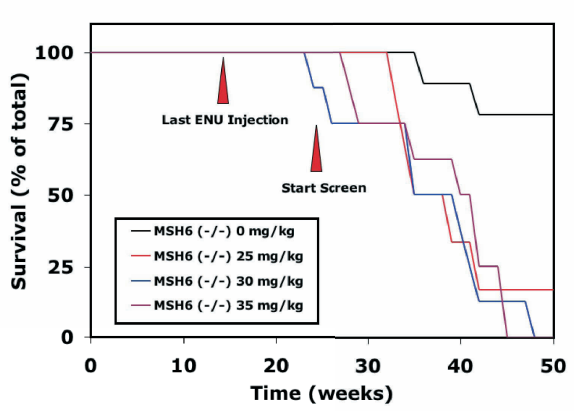

Figure 1: Effect of ENU on the survival of *msh6*$^{-/-}$ males. ***Msh6***$^{-/-}$ male rats show an increased mortality after treatment with different concentrations of ENU compared with untreated ***msh6***$^{-/-}$ male rats. Survival of untreated and ENU-treated wild-type rats is 100% in the indicated time-period (data not shown). Red arrowheads indicate the time points of the last ENU injection and the start of mating for F_1 progeny that can be screened for mutations without risk for chimaeric progeny.

Increased ENU mutagenicity in MSH6-deficient background

The highest ENU dose resulting in more than 25% fertile males after a full cycle of spermatogenesis was found to be 3 x 30 mg/kg for the *msh6*$^{-/-}$ background. In wild-type Wistar rats this dose was 3 x 40 mg/kg [5,21] and in SD and F344 2 x 60 mg/kg [6]. We screened 768 amplicons in 291 F_1 progeny of the 3 x 30 mg/kg group and covered ~70 Mb of sequence. A total of 84 ENU-induced mutations were identified (**Table 1**). As a result, the overall mutation rate is 1 mutation every 8.29 x 10^5 bp, a 1.5-fold increase compared with the highest mutation rate in MSH6-proficient animals.

Only one nest of three pups could be recovered from the single fertile male that was treated with 3 x 35 mg/kg (this animal was sterile in subsequent mating) and one mutation was discovered in the 0.72 Mb that was sequenced (**Table 1**). 16 F_1 progeny from a single male that was treated with 25 mg/kg were screened and in this group 4 mutations were discovered in 3.6 Mb, resulting in a similar mutation rate as observed in the 3 x 30 mg/kg treated group (**Table 1**).

Although several candidate mutations were found more than once, though at low frequency, analysis of family relationship revealed that these are not because of clonal effects, as in all cases, F_1 progeny bearing the same mutations descended from different founders. These polymorphism where thus classified as low frequency SNPs (note that Wistar is an outbred strain) and not included in the results. Unfortunately, mutation density in our study is not high enough to draw any conclusions on the existence of hypersensitive or ENU-resitant regions in the rat genome.

Mutation frequency changes over time

We analyzed the mutation frequency of the F_1 offspring in time bins after mutagenesis of the founders. Offspring from mutant males treated with 3 x 30 mg/kg demonstrate a reduction of mutation frequency in time (**Fig. 2A**). In the 79 animals that were born 14 – 17 weeks after the last ENU injection ~ 18.7 Mb was analyzed and 32 mutations

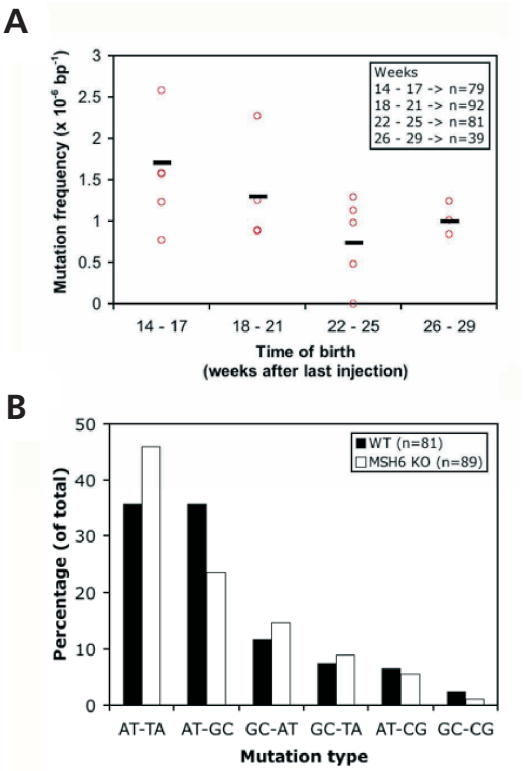

Figure 2: ENU-induced mutation frequency and spectrum. (A) The ENU-induced mutation frequency differs in time. Red circles indicate the mutation frequency of the F_1 progeny of the 5 different *msh6*$^{-/-}$ founders treated with 30 mg ENU/kg. The black stripes indicate the mutation frequency of all the F_1 progeny of these founders together for the different time bins. A decrease of mutation frequency is observed in time, which reaches a steady level at 1.0 x 10^{-6} per bp (1 mutation per 1 Mb). The letter n indicates the number of total F_1 screened per time bin. (**B**) ENU-induced mutation spectrum in MSH6-deficient background (white bars) compared with MMR-proficient animals (black bars).

were discovered, reflecting a mutation frequency of 1.71 x 10^{-6} per bp and a rate of 1 mutation every 585 kb. Animals that were born 18 – 21 weeks and 22 – 25 weeks after the last ENU injection show a decreased mutation frequency of 1.30 x 10^{-6} and 0.74 x 10^{-6} per base pair, respectively. We observed some notable variation in germ line mutation frequencies between the 5 founders. However, this was mainly the result of the differences in litter sizes at the analyzed time points and the variation decreased almost completely when the overall germ line mutation frequency of the individual founders was compared (1.20 x $10^{-6} \pm 0.25 \times 10^{-6}$), suggesting only limited inter-animal variability. F_1 progeny born later than 25 weeks after the last injection show a slight increase in frequency (~1.0 x 10^{-6} per bp). However, this observation is based on a relatively low numbers of F_1 progeny, derived from a limited number of founders, as the mortality started to increase at this time. The mean survival of ENU-treated *msh6*$^{-/-}$ rats (3 x 30 mg/kg) was 37 ± 3 weeks, which is 23 weeks after the last injection (**Fig. 1**). Notably, all fertile males lived at 37 weeks of age, while the infertile males had mostly died, suggesting a direct relationship between fertility and viability, and probably also the amount of ENU-induced genetic damage.

ENU-induced mutation spectrum differs in MSH6-deficient background

Interestingly, it was found that ENU induces a different mutation spectrum in an MSH6-deficient background as compared with an MMR-proficient background (**Fig. 2B**). The most common mutations in both wild-type as well as MSH6-deficient background were alternations of A-T base pairs (~75%). This is consistent with previous reports and is thought to result from O^2-ethylthymine (O^2-etT) and O^4-ethylthymine (O^4-etT) lesions [17,23]. However, for the A-T pairs a significant increase in A-T to T-A transversions ($p = 0.045$) and decrease of A-T to G-C transitions ($p = 0.019$) was observed in the MSH6-deficient background. It has been shown that bypass of O^2-etT lesions induces A-T to T-A transversions, whereas O^4-etT lesions results in A-T to G-C transitions [24,25]. Our data suggests that MutSα preferentially recognizes O^2-etT lesions in the rat.

Because of non-random codon usage and the highly biased composition of translation termination codons, the chance for introducing a premature stop codon strongly depends on the mutation spectrum. For example, A-T to T-A transversions can introduce a premature stop codon in 7 out of the total 183 possible changes in the genetic code (excluding the combinations that already code for a stop signal), whereas A-T to G-C transitions will never result in a premature stop codon. Based on the mutation spectra obtained in wild-type and MSH6-deficient backgrounds we calculated the chance of generating a knockout-type allele by introducing a premature stop codon or mutating splicing donor/acceptor site for the 768 amplicons that were used in our screen. The chance of generating a knockout-type allele was found to be increased by ~21% in the MMR-deficient background (from 4.26% for wild-type to 5.16% for the *msh6*$^{-/-}$ mutation spectrum).

ENU-induced mutations

The 768 amplicons that were used in the mutation screening were designed to cover exons of genes of interest. 84 out of the 89 mutations that were identified reside in protein-coding regions (**Table 1**). Four mutations (~ 5%) introduce a premature stop codon in an open reading frame and are thus most likely to result in a functional knockout of the gene, which corresponds nicely with the expected chance of introducing a premature stop codon, as discussed above. 64 mutations cause an amino acid change (missense, ~ 76%) and 16 mutations do not affect protein coding (silent, ~ 19%), which also correspond with calculated predictions (74% and 22%, respectively). None of the non-coding mutations resided in splicing donor/acceptor sites.

DISCUSSION

The mutagen ENU can transfer its ethyl group to oxygen or nitrogen radicals present in DNA, which results in lesions that can cause mispairing during replication and eventually give rise to a single base pair substitution [23]. We hypothesized that a

deficiency in the repair system that detects or corrects such single base pair damages or mismatches would result in an increased ENU-induced mutation frequency and associated improvement of the target-selected mutagenesis-based knockout procedure. Indeed, deficiency of MSH6 was found to improve the ENU-driven knockout procedure in two ways; 1) it increases the ENU-induced germ line mutation frequency up to 1 mutation every 585 kb and 2) the mutation spectrum is changed and enhances the chance of generating a knockout-type allele by 21%. Cumulatively, this results in a ~2.5-fold increase in knockout efficiency.

The increased mutation frequency was reached at an ENU dose that was 25% lower than in wild-types and together with the change in mutation spectrum this argues for the specificity of MMR-deficiency underlying the observed effects. In line with this, an increase in mutagenicity of ENU in MMR-deficient background was also shown in mouse ES cells lacking Msh2 [17]. However, it has to be noted that in zebrafish, which lack MSH6, no difference in ENU-induced mutation frequency has been observed compared with MMR-proficient fish [26]. This apparent contradiction with the results presented here could be explained by the large difference in mutation frequency. In wild-type zebrafish, the ENU-induced mutation frequency is about 1 mutation every 100,000 bp, whereas, this frequency in wild-type rats is more than 10-fold lower [5,8,9,10,11,12]. Proposedly, the zebrafish mutation load is the maximum that is compatible with viability, a suggestion that is corroborated by comparable maximal mutation frequencies observed in *C. elegans* [13] and Arabidopsis [14]. Our results suggest that the lower ENU-induced mutation frequencies in rodents can at least partially be explained by more efficient mismatch repair in the testis that counteracts mutagenicity, and is less likely because of increased sensitivity to genotoxic damage in general. The decline of the mutation frequency in time that is observed in this study, however, could indicate the presence of genotoxic effect. Initial depletion of spermatogonial stem cells could be because of apoptosis induced by ENU damage – a mechanism that is presumed to be underlying the sterility effect at higher doses. Selective repopulation of the testis by the most viable stem cells with presumably the lowest amount of genotoxic damage, decreases the apparent mutation frequency (and potentially increases the change for clonal progeny). Our results indicate that the target-selected mutagenesis works most efficient for F_1 progeny resulting from matings at about 10 – 14 weeks after the last ENU injection. However, it should be kept in mind that only F_1 animals, generated after one full cycle of spermatogenesis (in rat and mouse about 60 to 70 days) after mutagenesis, should be screened in order to prevent retrieval of chimaeras. Such chimaeras can be induced by ethyl-DNA adducts in the fertilized oocyte that originate from mutagenized post-meiotic sperm cells and which could result in fixation of mutation in different lineages during embryonic development.

Besides the ENU mutagenesis approach, other gene targeting technologies are being developed for the rat, which include nuclear transfer, although to date no genetic modification has been reported, and knockdown by RNA interference (reviewed in

[27]). Recently, transposon-tagged mutagenesis [28] was successfully applied to the rat [29] as well and although this technique is currently less amendable for scaling, it should be considered as highly complementary to the existing ENU-based efforts. The ENU-driven target-selected mutagenesis approach has already been used successfully for generating a variety of novel rat knockout models [5,6,16] and with the improvements described here, this approach does provide realistic technological requirements for screens on a genome-wide scale. In a wild-type background, ~110,000 F_1 rats would be needed to knockout any given gene with 95% probability (**Fig. 3**). This number is reduced to 40,000 in the MMR-deficient background. When reasoning the other way around, 40,000 F_1 rats will contain knockouts of 95% of all rat genes and when considering 'only' 5,000 F_1 rats, knockout alleles for ~50% of all the rat genes will be present. It should be said, however, that to identify these knockout alleles, the complete ORFeome would have to be screened. Although existing technologies are not suited for this, emerging massively parallel sequencing technologies [30], and microarray-based enrichment procedures [31,32] provide promising avenues in this direction. Keeping this in mind, archiving frozen sperm samples of mutant F_1 animals, which can be recovered by intracytoplasmic sperm injection (ICSI) [33,34] becomes very attractive.

The identification of a wide range of potentially interesting missense mutations as well as the retrieval of four novel candidate knockout models resulting from the introduction of premature stop codons by screening only a small set of about 300 animals illustrates the power of the approach. Furthermore, the ENU-based approach has the potential to generate allelic series (multiple mutations in the same gene) in different animals, which can facilitate the identification of novel gene functions.

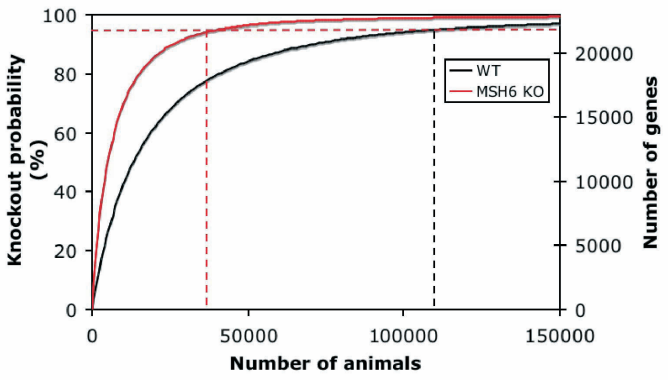

Figure 3: Probability of gene knockouts in MSH6-deficient and wild-type rats. The chance to retrieve a knockout for any given gene and the total number of genes that will be knocked out when all genes would be screened for mutations is plotted as a function of the number of mutant F_1 animals for wild-type (WT, 40 mg ENU/kg, black line) and MMR-deficient (*Msh6*$^{-/-}$, 30 mg ENU/kg, red line) rats. The red dashed lines show the number of animals needed to knockout 95% of all genes and any given gene with 95% chance. The use of a MSH6-deficient background reduces the number of animals ~2.5-fold.

For example, hyper- and hypomorphic mutations provide information related to gene dosage effects and residues important for specific protein interactions or enzymatic functions can be identified.

Currently, the 4 knockouts and 45 selected missense mutations are being crossed to the next generation and bred to homozygosity for subsequent phenotypic analysis. As homozygous MSH6-deficiency, which could occur in later generations, would result in further accumulation of novel mutations, this outcross is also used to eliminate the mutant *msh6* allele from the genetic background by genotype-assisted breeding. In addition, further outcrossing to wild-type background should be performed and littermates should be included as control animals in phenotypic characterization experiments to minimize confounding effects of background mutations. Although such effects should never be ignored, estimates do indicate that this potential problem should not be exaggerated [35], especially if outcrossing is combined with marker-assisted selection for which the outbred Wistar background is well-suited.

CONCLUSIONS

We have significantly improved the target-selected mutagenesis gene knockout technology for rats by taking advantage of deficiency of specific MMR function. As mutagen-driven target-selected mutagenesis approaches have become popular in organisms for which no ES cell-based techniques are available, such as medaka [36], but also plants like rice [37], Lotus [38] and maize [39], the improvements described here may equally well be applied to other species for optimization of gene knockout efficiency.

METHODS

Animals and ENU mutagenesis protocol. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Sciences according to the Dutch legal ethical guidelines. Experiments were designed to minimize the number of required animals and their suffering. Male MSH6 knockout rats (*Msh6*^{1Hubr}) of 12 weeks of age were given three weekly intraperitoneal injections of 1, 5, 10, 20, 25, 30 and 35 mg ENU /kg bodyweight. Preparation of ENU (Isopac; Sigma) was done as described [5]. Injected males were monitored for fertility by breeding with one or two females, starting 3 weeks after the last mutagenesis treatment. Pups from these early matings were counted, but not analyzed. Ten weeks after the last injection, fertile males of the highest-dosed fertile group were kept on a weekly breeding scheme with two females to produce F_1 progeny for mutational analysis.

Animals were housed under standard conditions in groups of two to three per cage per gender under controlled experimental conditions (12-h light/dark cycle, 21±1°C, 60% relative humidity, food and water *ad libitum*).

Genomic DNA isolation, PCR and sequencing conditions. At two weeks of age F_1 progeny were uniquely tagged by ear clipping and the resulting tissue fragments were lysed as described [5], followed by phenol/chloroform (1:1, vol/vol) extraction. DNA was precipitated by adding 300 µl isopropanol, mixing and centrifuging for 20 min, at 21,000 *g* at 4°C. The supernatant was discarded and pellets were washed with 70% ethanol and dissolved in 500 µl 10 mM Tris-HCl (pH 8.0). 768 pre-selected amplicons were amplified using a nested PCR setup as described [5]

with the following modifications. The first PCR reaction was carried out in 2 x 384 wells plates per F_1 animal sample, in a total volume of 5 µl, and every well contained a unique set of primers (0.2 µM of each). After thermocycling, the PCR1 reactions were diluted with 20 µl H_2O and 1 µl was used as template for the second round of PCR, which was carried out in the same 2 x 384 wells format as the first PCR, however, with different sets of primers which are internal to the first set (nested). Sequencing reactions contained 0.1 µl BigDye (v3.1; Applied Biosystems), 1.9 µl BigDye Dilution Buffer (Applied Biosystems) and 0.4 µM of the forward primers used for the PCR2 reaction in a total volume of 5 µl. After thermocycling purified sequencing products were analyzed on a 96-capillary 3730XL DNA analyzer (Applied Biosystems), using the standard RapidSeq protocol on 36 cm array. Sequences were analyzed for the presence of heterozygous mutations using PolyPhred [22] and in-house developed software. All candidate mutations were verified in independent PCR and sequencing reactions.

Project management and primer design using LIMSTILL. All resequencing projects were designed and 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 [40] and parameters are set to design primers with an optimal melting temperature of 58°C.

Knockout probability calculation. Calculation of the frequency of introducing a premature stop codon with different mutation spectra is an integrated feature of LIMSTILL (http://limstill.niob.knaw.nl). The chance of generating a knockout-type allele was calculated for all 768 working amplicons by comparing the probability of nonsense mutations and splicing donor/acceptor mutations divided by all changes in coding sequence plus splicing donor/acceptor sites (first two and last two nucleotides of each intron), all corrected for the wild-type and mutant mutation spectra. To calculate the knockout probability both the mutation frequency and corresponding spectrum were taken into account. Rat Ensembl Build 47.34q was downloaded from ftp.ensembl.org. For every gene the longest transcript (22,959 in total) was used to calculate the total number of coding nucleotides. Wistar animals treated with 40 mg/kg ENU represent the most optimal wild-type group and were compared with F_1 progeny from *msh6*$^{-/-}$ animals that was born 14 – 17 weeks after the last ENU injection.

Statistical analysis. For calculating statistical differences in the mutation spectrum data the chi-squared test was used. $P < 0.05$ was considered to be statistically significant.

ACKNOWLEDGMENTS

The authors thank S. Linsen for bioinformatic help and statistical analyses.

REFERENCES

1. Jacob HJ, Kwitek AE (2002) Rat genetics: attaching physiology and pharmacology to the genome. Nat Rev Genet 3: 33-42.
2. Lazar J, Moreno C, Jacob HJ, Kwitek AE (2005) Impact of genomics on research in the rat. Genome Res 15: 1717-1728.
3. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6: 507-512.
4. Stemple DL (2004) TILLING - a high-throughput harvest for functional genomics. Nat Rev Genet 5: 145-150.
5. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
6. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of

knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
7. Justice MJ, Carpenter DA, Favor J, Neuhauser-Klaus A, Hrabe de Angelis M, et al. (2000) Effects of ENU dosage on mouse strains. Mamm Genome 11: 484-488.
8. Augustin M, Sedlmeier R, Peters T, Huffstadt U, Kochmann E, et al. (2005) Efficient and fast targeted production of murine models based on ENU mutagenesis. Mamm Genome 16: 405-413.
9. Concepcion D, Seburn KL, Wen G, Frankel WN, Hamilton BA (2004) Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. Genetics 168: 953-959.
10. Michaud EJ, Culiat CT, Klebig ML, Barker PE, Cain KT, et al. (2005) Efficient gene-driven germ-line point mutagenesis of C57BL/6J mice. BMC Genomics 6: 164.
11. Quwailid MM, Hugill A, Dear N, Vizor L, Wells S, et al. (2004) A gene-driven ENU-based approach to generating an allelic series in any gene. Mamm Genome 15: 585-591.
12. Sakuraba Y, Sezutsu H, Takahasi KR, Tsuchihashi K, Ichikawa R, et al. (2005) Molecular characterization of ENU mouse mutagenesis and archives. Biochem Biophys Res Commun 336: 609-616.
13. Cuppen E, Gort E, Hazendonk E, Mudde J, van de Belt J, et al. (2007) Efficient target-selected mutagenesis in Caenorhabditis elegans: toward a knockout for every gene. Genome Res 17: 649-658.
14. Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13: 524-530.
15. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RH, et al. (2003) Efficient target-selected mutagenesis in zebrafish. Genome Res 13: 2700-2707.
16. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
17. Claij N, van der Wal A, Dekker M, Jansen L, te Riele H (2003) DNA mismatch repair deficiency stimulates N-ethyl-N-nitrosourea-induced mutagenesis and lymphomagenesis. Cancer Res 63: 2062-2066.
18. Jiricny J (2006) The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7: 335-346.
19. Duckett DR, Drummond JT, Murchie AI, Reardon JT, Sancar A, et al. (1996) Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. Proc Natl Acad Sci U S A 93: 6443-6447.
20. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
21. Smits BM, Mudde J, Plasterk RH, Cuppen E (2004) Target-selected mutagenesis of the rat. Genomics 83: 332-334.
22. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25: 2745-2751.
23. Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. Mamm Genome 11: 478-483.
24. Bhanot OS, Grevatt PC, Donahue JM, Gabrielides CN, Solomon JJ (1992) In vitro DNA replication implicates O2-ethyldeoxythymidine in transversion mutagenesis by ethylating agents. Nucleic Acids Res 20: 587-594.
25. Klein JC, Bleeker MJ, Lutgerink JT, van Dijk WJ, Brugghe HF, et al. (1990) Use of shuttle vectors to study the molecular processing of defined carcinogen-induced DNA damage: mutagenicity of single O4-ethylthymine adducts in HeLa cells. Nucleic Acids Res 18: 4131-4137.
26. Feitsma H, de Bruijn E, van de Belt J, Nijman IJ, Cuppen E (2008) Mismatch repair deficiency does not enhance ENU mutagenesis in the zebrafish germ line. Mutagenesis 23: 325-329.
27. Aitman TJ, Critser JK, Cuppen E, Dominiczak A, Fernandez-Suarez XM,

et al. (2008) Progress and prospects in rat genetics: a community view. Nat Genet 40: 516-522.

28. Luo G, Ivics Z, Izsvak Z, Bradley A (1998) Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc Natl Acad Sci U S A 95: 10769-10773.
29. Kitada K, Ishishita S, Tosaka K, Takahashi R, Ueda M, et al. (2007) Transposon-tagged mutagenesis in the rat. Nat Methods 4: 131-133.
30. Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5: 16-18.
31. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, et al. (2007) Genome-wide in situ exon capture for selective resequencing. Nat Genet 39: 1522-1527.
32. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, et al. (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4: 907-909.
33. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
34. Nakatsukasa E, Inomata T, Ikeda T, Shino M, Kashiwazaki N (2001) Generation of live rat offspring by intrauterine insemination with epididymal spermatozoa cryopreserved at -196 degrees C. Reproduction 122: 463-467.
35. Keays DA, Clark TG, Flint J (2006) Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. Mamm Genome 17: 230-238.
36. Taniguchi Y, Takeda S, Furutani-Seiki M, Kamei Y, Todo T, et al. (2006) Generation of medaka gene knockout models by target-selected mutagenesis. Genome Biol 7: R116.
37. Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, et al. (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7: 19.
38. Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, et al. (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. Plant Physiol 131: 866-871.
39. Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, et al. (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12.
40. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386.

5

SYSTEMATIC GENERATION OF *IN VIVO* G PROTEIN-COUPLED RECEPTOR MUTANTS IN THE RAT

Ruben van Boxtel[1], Bas Vroling[2], Pim W. Toonen[1], Isaac J. Nijman[1], Henk S. van Roekel[1], Mark Verheul[1], Coos Baakman[2], Victor Guryev[1], Gert Vriend[2] and Edwin Cuppen[1],[3]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2]Center for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.
[3]Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, Utrecht, The Netherlands

Adapter from Pharmacogenomics J *in press*

ABSTRACT

G protein-coupled receptors (GPCRs) constitute a large family of cell surface receptors that are involved in a wide range of physiological and pathological processes, and are targets for many therapeutic interventions. However, genetic models in the rat, one of the most widely used model organisms in physiological and pharmacological research, are largely lacking. Here, we applied *N*-ethyl-*N*-nitrosourea (ENU)-driven target-selected mutagenesis to generate an *in vivo* GPCR mutant collection in the rat. A pre-selected panel of 250 human GPCR homologs was screened for mutations in 813 rats, resulting in the identification of 131 nonsynonymous mutations. From these, 7 novel potential rat gene knockouts were established as well as 45 lines carrying missense mutations in various genes, associated with or involved in human diseases. We provide extensive *in silico* modeling results of the missense mutations and show experimental data, suggesting loss-of-function phenotypes for several models, including *Mc4r* and *Lpar1*. Taken together, the approach used resulted not only in a set of novel gene knockouts, but also in allelic series of more subtle amino acid variants, similar as commonly observed in human disease. The mutants presented here may greatly benefit studies to understand specific GPCR function and support the development of novel therapeutic strategies.

INTRODUCTION

Treatment of animals with a mutagenic compound that introduces random mutations in the germ line is a very fast and efficient method for introducing a wide range of mutations in large sets of genes *in vivo*. In rodents, ENU has been shown to be the most potent chemical germ line mutagen [1]. ENU-treatment of male animals causes adducts in the DNA of spermatogonial stem cells, which after several rounds of cell division, result in random point mutations and mutagenized sperm [2]. F_1 animals derived from outcrosses with wild-type females carry random heterozygous ENU-induced mutations in their genome. Subsequently, the DNA of these animals can be screened by a variety of techniques for the presence of mutations in pre-selected genes of interest [3,4], with the goal to identify animals that carry induced variants that affect normal protein function, e.g. by the introduction of a premature stop or by affecting functionally important residues.

The laboratory rat *Rattus norvegicus* is one of the most used model organisms in biomedical research and has been the preferred model for studying human physiology and pathology [5]. As a highly diverged mammalian model (~60 million years with human and 20-40 million years with mouse [6]), the rat is highly complementary to the mouse, enabling phenotypic comparison of gene knockouts in both mammals to better understand the specific gene function in human biology. In addition, in specific cases the rat can have advantages in studying mammalian physiology and biology because of its relative large body size and the availability of well-established behavioral and neurological assays [7]. While most rat knockout models have thus far been generated through ENU-driven approaches, only recently alternative technologies emerged. Transposon-tagged mutagenesis [8], zinc-finger nuclease-mediated knockout generation [9] and the isolation of pluripotent ES cells that potentially can be used for gene targeting [10,11], now provide a range of possibilities for manipulating the rat genome and promises to boost the use of the rat as a versatile genetic model system. ENU-driven target-selected mutagenesis has specific characteristics that make it an attractive technology that is complementary to the other approaches [12]. First, it is a relatively simple technology without any cell or oocyte manipulation steps. Second, it can easily be scaled up for high throughput and is a relatively cheap method, especially in terms of the number of animals used per knockout (in this paper ~100 rats). Thirdly, it offers the possibility to identify (allelic series of) more subtle variation because of amino acid changes that result in hyper- and hypomorphic alleles [3,4]. One of the major disadvantages of the ENU-based approach was its relative inefficiency. However, recently we increased the efficiency by about 2.5-fold by taking advantage of DNA mismatch repair (MMR)-deficiency in the MSH6 knockout rat [13,14], a system known to be involved in repairing ENU-induced lesions in the genome [15]. Further efficiency improvements can be expected by implementing next-generation sequencing technology for mutation discovery. Another drawback of the method is that mutation

generation is random and that only the discovery is done in a targeted fashion. In other words, generation of knockouts is relatively efficient, but obtaining a knockout for a specific gene is still challenging. However, ENU-driven target-selected mutagenesis is a versatile technology for the systematic generation of large catalogs of knockouts and allelic variants of gene families or eventually all protein-coding genes. The latter approach in combination with efficient cryopreservation and rederivation protocols would generate a unique genome-wide resource for knockouts as well as mutant alleles reflecting human genetic variation.

Here, we applied the improved ENU-driven target-selected mutagenesis method for generating a unique resource of *in vivo* GPCR mutant rat models, consisting of both knockouts as well as (allelic series of) missense mutations. GPCRs are 7 transmembrane (TM) receptors, which regulate many cellular processes, including the senses of taste, smell, and vision, and control a myriad of intracellular signaling systems in response to external stimuli. Importantly, many diseases are linked to GPCRs and they represent by far the largest class of targets for current drugs as well as for the development of novel small-molecule medicines [16]. Moreover, because of their role in the regulation of cellular function they are arguably one of the best-studied classes of proteins, although for many GPCRs their ligand as well as biological function remains to be elucidated. Furthermore, genetically altered GPCR animal models are scarce, especially in non-murine species, The use of a random mutagenesis approach for the generation of GPCR mutants is in principle very well suited for understanding *in vivo* receptor function as new insights can be obtained by completely knocking out specific receptors, but also by changing functionally important residues, e.g. involved in ligand binding or second messenger signal transduction. Importantly, the high structural conservation between the different GPCRs allows for confident prediction of possible effects of amino acids changes. We systematically applied the ENU-driven target-selected mutagenesis approach to a set of about 250 rat GPCRs that have clear orthologs to human GPCRs. In total, we identified 131 nonsynonymous mutations in 99 different GPCRs, including 7 novel potential knockout alleles and 45 missense mutants that were predicted to affect specific GPCR function or stability of folding of the protein. Characterization of selected models demonstrates that ENU target-selected mutagenesis is a powerful and efficient approach for *in vivo* functional studies on G protein-coupled receptors.

RESULTS

ENU target-selected mutagenesis setup

The target rat GPCR genes for mutation screening were selected based on one-to-one orthology with human GPCRs (as defined in Ensembl database), where odorant receptors were excluded. GPCRs are ideal genes for mutation screening by PCR-based resequencing in an ENU target-selected mutagenesis setup because these genes are often encoded by only a single long exon, which maximizes the information content per target

amplicon. Although the chance of identifying mutations in splice site residues, which often results in a knockout allele, is decreased, this is compensated by a higher number of nonsynonymous mutations. The genes of interest were screened using a nested PCR amplification setup followed by dideoxy resequencing [3]. Although different methods for mutation retrieval can be used, like a yeast-based assay [4], CEL1-based nuclease cleavage [17] or Mu transposase-based detection [18], resequencing is considered to be the golden standard since it is equally sensitive towards all types of point mutations and is well suited for scaling and automation. After a first round of amplicon testing a panel of 486 different amplicons covering 250 different GPCRs for screening was established (**Table S1**).

MSH6-deficient males (*msh6*$^{-/-}$) were mutagenized with a predetermined optimal dose of 3 weekly treatments of 30 mg per kg bodyweight of ENU [14], which yielded 18 fertile founders (**Table 1**). Subsequently, mating the ENU-treated *msh6*$^{-/-}$ males with untreated females generated a mutant F_1 population, harboring random heterozygous ENU-induced mutations. Only F_1 animals were screened that were generated after a full cycle of spermatogenesis (>60 days after mutagenesis) to prevent retrieval of chimaeras. Genomic DNA of the F_1 animals was isolated from a tail clip that was collected at 1 to 2 weeks of age and screened for the ENU-induced mutations in the preselected panel of genes-of-interest. F_1 animals carrying interesting candidate mutations were weaned and the mutations were reconfirmed in independent assays. In total, we screened 813 F_1 animals, covering 139 Mb of DNA (**Table 1**) and identified 193 unique mutations, resulting in a mutation rate of 1 per 720 kb, which is in agreement with the previously described increased mutation frequency in *msh6*$^{-/-}$ rats [14].

Table 1: ENU mutation efficiency

Number of GPCR genes screened	250
Fertile founders[a]	18
Screened F_1 animals[b]	813
Screened base pairs (bp)	139 x 10^6
Nonsynonymous mutations	131
Nonsense	9
Missense	122
Synonymous	32
Non-coding	30
Total mutations	193
Mutation rate	1 per 720 kb

[a]MSH6-deficient male animals were treated 3 times weekly with 30 mg/kg bodyweight ENU. Founders were considered to be fertile if at least one nest was produced more than 10 weeks after the last ENU treatment. [b]Only F_1 animals were screened that were born at least 10 weeks after the last ENU treatment. GPCR, G protein-coupled receptor.

ENU-induced mutations

Out of the 193 ENU-induced mutations, 163 are located in coding sequences and 131 result in nonsynonymous changes in 99 different GPCRs (Supplementary **Table 2**). We identified 9 mutations that cause the introduction of a premature stop codon in the open reading frame (ORF). These represent 5.5% of all coding mutations (**Table 1**) and corresponds nicely with the expected percentage of knockout alleles when considering codon usage in the rat and the mutation spectrum in a MMR-deficient background [14]. 122 mutations were identified that cause an amino acid change (missense). As expected, none of the non-coding mutations mutated a splicing donor/acceptor site.

In silico analysis of the mutations

In order to categorize and prioritize the nonsynonymous mutations, the effects of the mutations on protein function was analyzed *in silico* by evaluating available experimental mutation data, analyzing residue conservation patterns in multiple sequence alignments and studying homology models of the mutated receptors. Since the structure of a protein is directly related to its function, the best way of estimating the effects of a mutant is by studying the protein structure itself. Unfortunately, there is very limited structural data available for GPCRs. However, based on alignment data and the few experimentally resolved protein structures [19,20,21] it is possible to build structure models of most class A GPCRs.

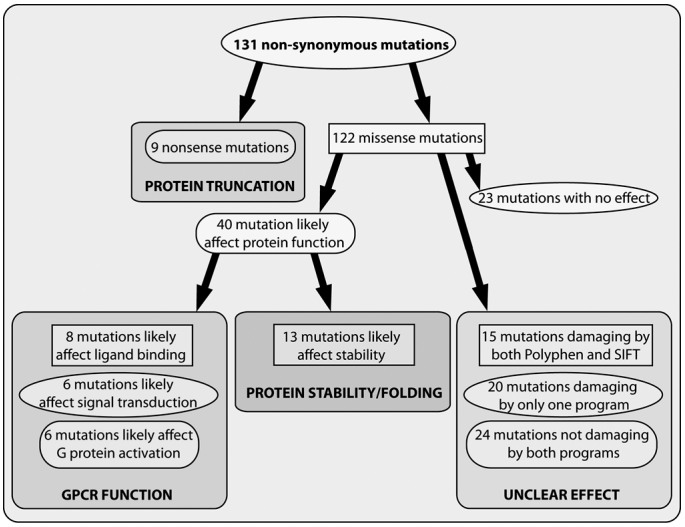

Figure 1: Systematic *in silico* analysis of the identified ENU-induced mutations. All mutations were grouped according to their predicted affect on GPCR function. Mutations that are likely to affect protein function can be further categorized depending on their effect on GPCR function. The group for which no predictions can be made by lack of structural data were analyzed with PolyPhen [23] and SIFT [24] software.

The 9 nonsense mutations that introduce a premature stop codon in the ORF will result in truncated versions of the proteins and are most likely to result in complete functional knockouts of the genes (**Fig. 1**). The 122 missense mutations can be grouped in amino acid changes that likely affect protein function or stability, and changes that will have little or no effect (Supplementary **Table 2**). For 40 mutations we predicted that they are likely to affect receptor function and for 23 that they are likely to have no effect (**Fig. 1**). For 59 mutations no good predictions could be made, which was largely because of the position of these mutations in loop regions, where sequence conservation is low and where very little structural information is available.

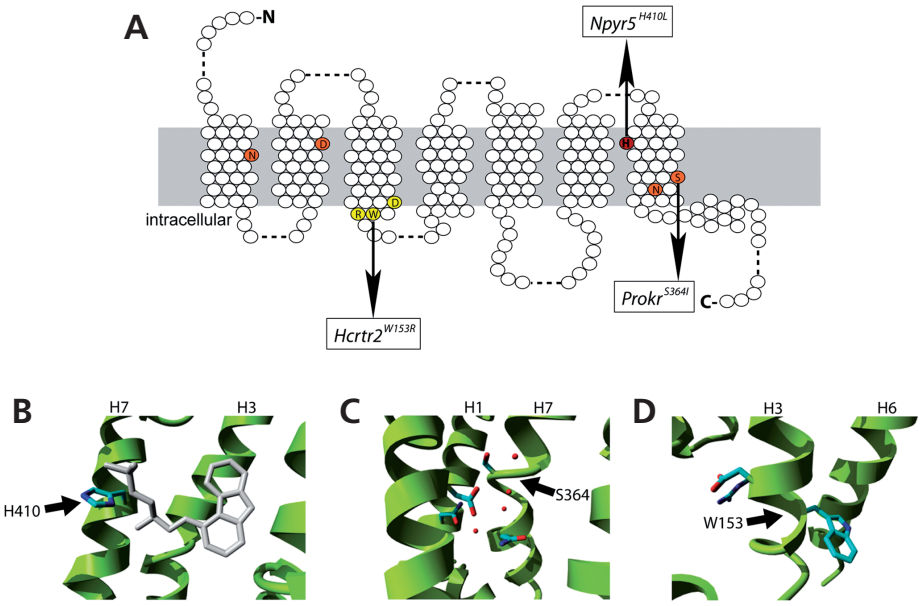

Figure 2: Illustrations of the mutant structural environments in homology models of the mutated receptors. (**A**) Schematic overview of a consensus GPCR with in red the mutation shown in (B), in orange the ionic pocket [22] and mutated residue in (C) and in yellow the (D/E)R(Y/W) motif and mutated residue depicted in (D). (**B**) An example of a mutation that is predicted to affect ligand binding. The mutant H410L in the neuropeptide receptor NPY5R is located in the putative ligand-binding pocket. The structure of the co-crystallized ligand of the β2-adrenergic receptor is shown in grey. Although the NPY5R receptor binds a different class of ligands the binding site location is expected to be similar. Substituting the histidine for leucine is likely to change ligand-binding affinity. (**C**) The mutant S364I in the prokineticin 2 (PROK2) is located just above the ionic pocket, which is involved in signal transduction from the ligand binding site to the G-protein binding site. A number of structural waters are located in this pocket. The substitution of the serine for isoleucine is likely to disrupt the ionic pocket because of steric constraints, a major change in hydrophobicity and loss of interactions with structural waters. (**D**) The mutant W153R in the hypocretin (orexin) receptor 2 (HCRTR2) is located in the (D/E)R(Y/W) motif, which is the most conserved part of the GPCR family and involved in receptor activation and subsequent G-protein coupling. The substitution of trypthophan for arginine will disrupt receptor activation.

The damaging mutations can be further categorized in mutations that affect specific GPCR functions, like ligand binding, signal transduction, G-protein activation, or protein stability. Eight mutations were predicted to affect the process of ligand binding, because they involved a residue located in the putative ligand-binding pocket (**Fig. 2A** and **B**). Another 6 mutations are likely to affect signal transduction, for example, the mutant *Prok2*^{S364I}, which involves a residue that is part of the ionic lock (**Fig. 2A** and **C**) [22]. 13 mutations were identified that are likely to affect G-protein activation by changing residues involved in the binding of the G-protein, like the W153R mutation in the highly conserved (D/E)R(Y/W) motif of *Hcrtr2* (**Fig. 2A** and **D**). Finally, a total of 13 mutations were identified that are likely to affect protein stability, for example, by changing a hydrophobic residue that sticks into the lipid bilayer into a hydrophilic one.

For mutations with unclear predictions we made use of PolyPhen [23] and SIFT [24] software to predict the effects of the amino acid changes (**Fig. 1**). Fifteen mutations were predicted to be damaging by both programs, 20 were predicted to be damaging by one of the programs and 24 were predicted not to be damaging by either program.

Archiving the mutants

F_1 animals carrying interesting mutations were outcrossed with untreated animals to establish the mutant lines and generate more heterozygous carriers. Mutants that were predicted to have no effect on GPCR function were archived by cryopreserving sperm of male carriers from either the F_1 or F_2 generation. The resulting resource can be used to revive these lines by ICSI [18] at a later stage. In a limited number of cases we were unable to cross F_1 animals to the next generation, partly because of fertility problems, which is more commonly observed in F_1 animals derived from ENU-mutagenized founders, or as a result of a dominant effect of the induced mutation. For example, the mutation of an aspartic acid (D74V) in the second TM domain of *Agtr1a*, which is part of the ionic pocket [22] and a highly conserved residue in GPCR super family, resulted in an extremely high blood pressure-like phenotype (increased liver, heart and dilated veins), lower bodyweight and severe testis atrophy, causing sterility and eventually death (data not shown). Completely in line with these observed *in vivo* effects, a similar amino acid change D74N was shown previously to affect AGTR1A function *in vitro* in COS cells [25].

The mutants that were successfully crossed to the next generation were genotyped for carrying the mutation of interest as well as contra-selected for the mutation in *Msh6*. All F_1 animals are heterozygous for the latter mutation as the ENU-mutagenesis was done in *msh6*$^{-/-}$ males that were subsequently outcrossed to wild-type females. Although no adverse effects are expected in a heterozygous *Msh6* background, in later generations the mutation could become homozygous and result in the accumulation of more mutations and cancer [13]. Therefore, we systematically selected F_2 animals to eliminate this mutation from the lines.

Table 2 lists the rat mutants that were crossed to next generations and for which living carriers are available. These include 7 mutant lines with protein truncations and contain

Table 2: *In vivo* GPCR rat mutants[a]

Category[b]	Gene
Protein truncation	*Ccr4, Gpr19, Gpr65, Gpr84, Htr1f, Il8rb, Mc4r*
Ligand binding	*Adra1b, Nmur2, Npy5r, P2ry1, P2ry13, Tacr1*
Signal transduction	*Fzd6, Galr1, Htr4, Il8rb*[c], *Prokr2, Ptafr,*
G-protein activation	*Edg2, Eltd1*[d], *Gpr120, Gpr68, Hcrtr2, Lhcgr*[d], *Mrgprd*[d], *Sstr2, Sstr5*
Protein stability/folding	*Bdkrb2, Chrm5, Cx3cr1, Fshr, Fzd7, Gnrhr, Gpr4, Gpr85, Mc5r, Mtnr1b, Smo*
Unclear effect[e]	*Drd3, Ffar3, Gpr116, Gpr142, Gpr15, Gpr182, Gpr56, Grm5, Htr2a, Lpar4, P2ry4, Xcr1*^{R137C}, *Xcr1*^{R218W}

[a]These mutant ratlines are crossed out to at least the F_2 generation and living carriers are available. [b]The categories are based on expert interpretation of structural information and bioinformatic predictions unless stated differently. [c]This mutation is linked to *Il8rb*^{C307X}. [d]These mutations were predicted to result in an increased constitutive activity of the receptors. [e]Structural information was not available for these protein domains but bioinformatic analysis by both Polyphen [23] and SIFT [24] software predicted that these mutations are likely to have damaging consequences.

well-studied receptors, like *Mc4r*, of which mutations in human have been associated with severe forms of obesity [26], as well as orphan receptors, like *Gpr19* (**Table 2**). Furthermore, 45 animals that carry mutations that likely affect GPCR function were crossed to following generations to establish a mutant line. These include 6 mutations that are predicted to affect ligand binding, 6 that may affect signal transduction and 9 that may affect G-protein activation. Additionally, 11 mutations that are predicted to affect protein stability and 13 with unclear effects on GPCR function or stability, but that were predicted to be damaging by both PolyPhen [23] and SIFT [24], were crossed to next generations.

Functional consequences of ENU-induced mutations

Mutations that result in the introduction of a premature stop codon most likely represent novel functional gene knockout models for these genes in the rat, similar as has been shown previously for the rat genes *Brca2* [4,27], *Apc* [28], *Sert* [29], *Msh6* [13] and *Pmch* [30]. Six of the nonsense mutants that were isolated in this screen cause protein truncations within or before the 7th TM domain and therefore lack the entire C-terminus, including the 8th helix, which is important for GPCR stability and function. The remaining ENU-induced premature translational stop was identified in the 8th helix of MC4R (*Mc4r*^{K314X}), four amino acids before the palmitoylated cysteine residue (**Fig. 3A**). Most likely, this mutation results in a complete loss of receptor function, since the two isoleucines residues, which are located after the mutated residue (**Fig. 3A**) were shown previously to be essential for localizing MC4R to the plasma membrane [31]. In addition, C-terminally truncated versions of another GPCR, namely the

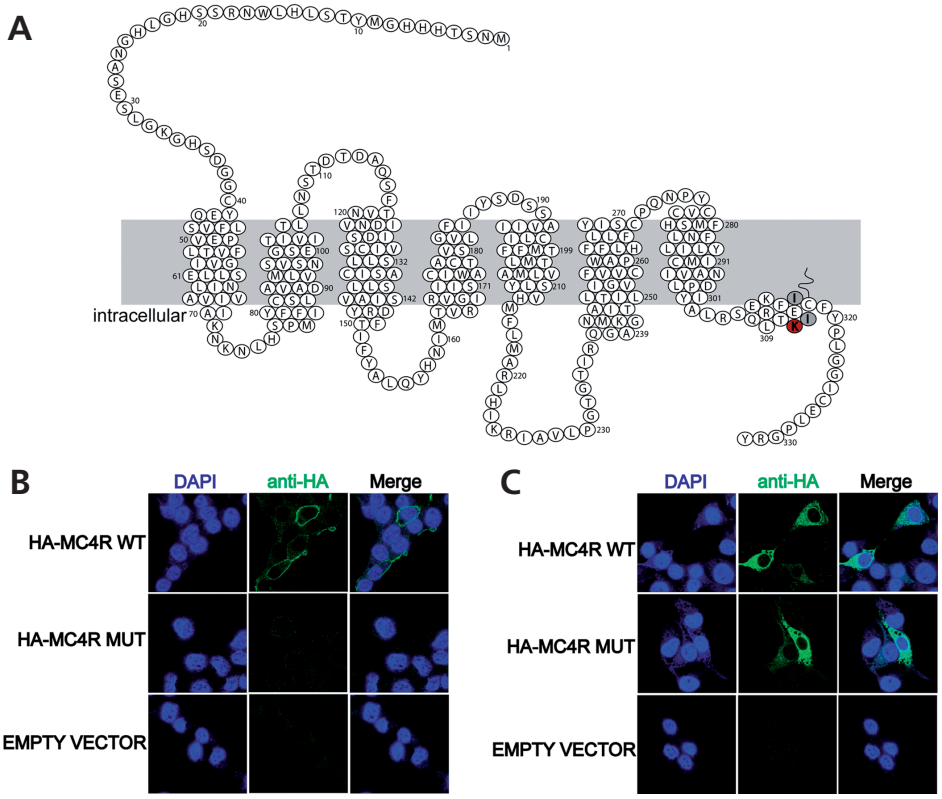

Figure 3: MC4R^{K314X} fails to localize to the plasma membrane *in vitro*. (**A**) Schematic overview of MC4R in the rat. Red indicates the location of the ENU-induced premature translational stop. Grey indicates two isoleucine residues that were shown previously to be essential for membrane localization.[31] (**B**) *In vitro* protein localization assays in transfected COS cells reveals plasma membrane localization for wild-type MC4R, but not for the mutated version of MC4R. Membrane localization was detected using N-terminally HA-tagged fusion constructs and extracellular availability of the HA tag in intact cells. (**C**) Both wild-type and mutant fusion proteins can be detected in fixed and permeablized COS cells, indicating that the mutant fusion protein is expressed, but fails to properly insert into the plasma membrane.

lysophophatidic receptor LPAR1 also fail to localize to the plasma membrane [32], indicating the importance of the C-terminus for correct membrane expression. To test this hypothesis, we expressed N-terminally haemagglutinin (HA)-tagged MC4R with and without the ENU-induced mutation in COS cells. After transfection, intact non-permeabilized cells were incubated with an antibody against the HA-tag, which can only bind if the HA-MC4R fusion protein is correctly incorporated into the plasma membrane. Indeed, HA-tagged wild-type MC4R was clearly detectable at the plasma membrane, whereas no mutant receptor could be detected in the same assay (**Fig. 3B**). To test whether the mutant form was expressed at all in these cells, we detected the protein in fixed and permeabilized transfected cells and showed the presence of

approximately equal amounts of expression of both wild-type and mutant HA-MC4R fusion proteins (**Fig. 3C**). This shows that *Mc4r*^{K314X} is still expressed *in vitro*, but fails to localize to the plasma membrane, which is likely to affect normal receptor function. In line with these predictions, *Mc4r*$^{K314X/K314X}$ rats display a major increase in body weight as well as in the amount of peritoneal and subcutaneous fat (**Fig. S1**), which is a comparable phenotype reported for traditional knockout mouse models [33] suggesting loss of receptor function.

Obviously, for missense mutations it is more difficult to robustly predict an affect. However, to confirm the value of the stringent bioinformatic predictions that we implemented, a mutation in *Lpar1* that results in the change of a methionine into an arginine in the 8th helix (**Fig. 4A**) and that was predicted to be deleterious for protein function, was analyzed. Interestingly, aberrant lysophosphatidic (LPA) signaling in humans has been associated with carcinogenesis in humans [34] and specifically LPAR1 knockout mice show phenotypic changes observed in psychiatric disease [35]. In the GPCR class A family the hydrophobicity of the affected residue is highly conserved and it is analogous to the phenylalanine of the NPxxY(x)$_6$F motif in Rhodopsin. This residue sticks into a hydrophobic pocket and contacts the tyrosine of the same domain, which is important for the folding of the 8th helix. The amino acid change in our mutant will disrupt the hydrophobic interactions and additionally, an arginine is too big to fit in this pocket and will most probably result in incorrect packing of the 8th helix (**Fig. 4B**). As a consequence, the environment of G-protein binding will be disturbed by the mutation in *Lpar1*^{M318R}, since it is thought that this helix interacts with the G-protein [36]. Indeed, homozygous mutant rats showed LPAR1 loss-of-function phenotypes, like craniofacial disorder (**Fig. 4C**) and smaller size (**Fig. 4D**), which is comparable to the phenotype found in *Lpar1* knockout mice [37]. However, we also observed significant differences in the phenotypes between the two animal models. In contrast to mice, no neonatal lethality was observed in rats since homozygous mutants were born at the expected frequencies (29.5% ± s.e.m. 8.7; n=4) and no hematomas were observed [37]. The milder phenotype in the rat could be explained by the nature of the mutation, which is only changing a single amino acid residue instead of truncating or deleting the protein. However, species-specific differences can also not be ruled out. To study the molecular consequences of the nonsynonymous substitution, we analyzed the membrane expression *in vitro* using N-terminally HA-tagged versions of mutant and wild-type LPAR1 isoforms. We observed strongly reduced, yet detectable, membrane incorporation as compared with wild-type LPAR1 (**Fig. 4E**), while expression levels of both fusion proteins were equal (**Fig. 4F**). These *in vitro* data suggest that the ENU-induced mutation results in a hypomorphic allele, which could explain the mild phenotype observed *in vivo*. More detailed experiments indicated that the decreased level of *in vitro* membrane expression of *Lpar*^{M318R} is most likely the result of increased spontaneous membrane internalization, which fits with the role of the C-terminus in receptor activation and internalization (R.B., E.C., unpublished results).

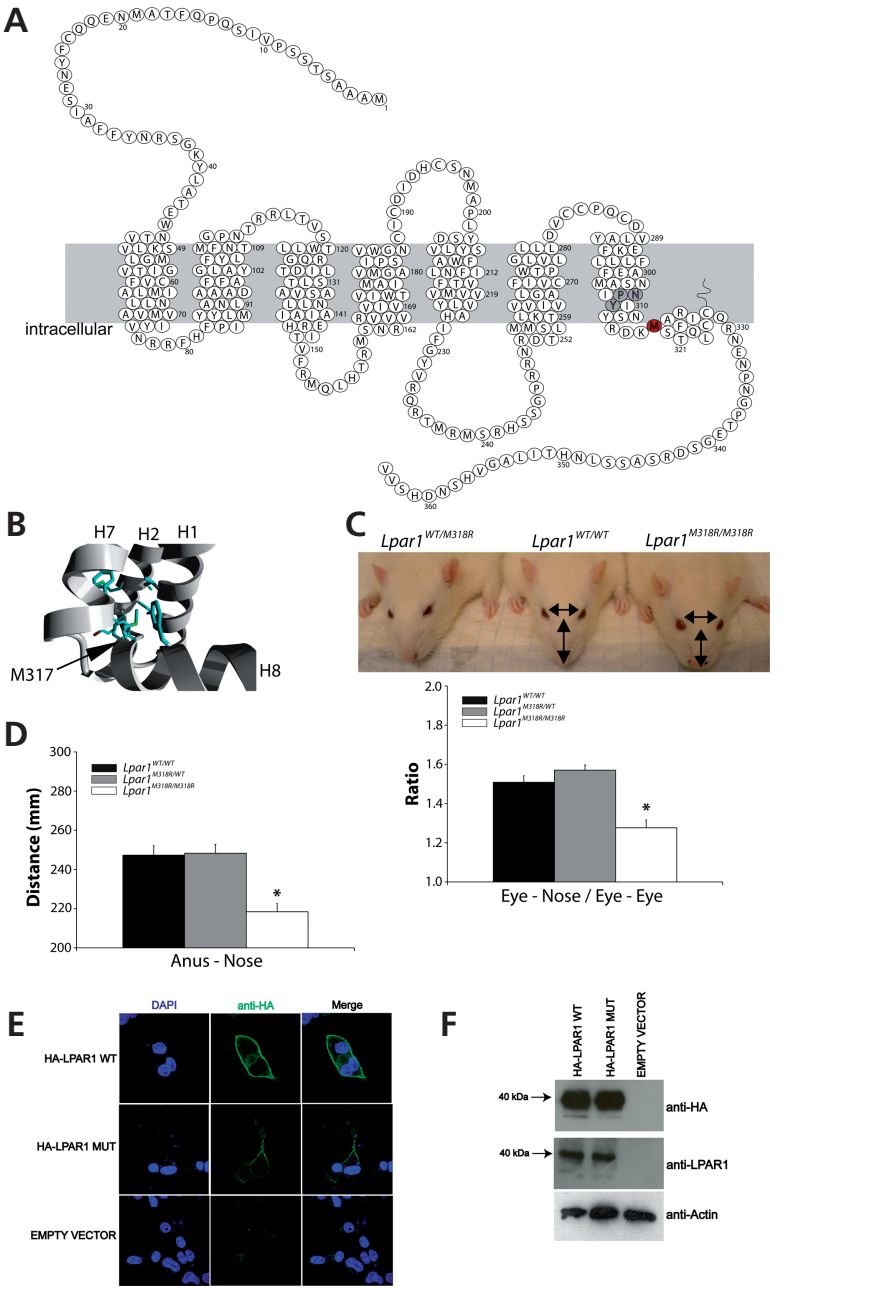

Figure 4: *Lpar1*$^{M318R/M318R}$ **rats show a loss-of-function phenotype.** (**A**) Schematic overview of LPAR1 in the rat. Red indicates the mutated residue, which is located in the 8th helix and grey indicates the NPxxY motif. (**B**) *In silico* analyses of the effect of the mutation in *Lpar1*. The substitution of methionine by arginine is likely to cause a severe disruption of the hydrophobic interface between helix 1, 2 and 7. This is mainly because of the fact that arginine is significantly bigger than methionine, therefore forcing a disruption of the local structure. The fact that a ▷

▷ hydrophobic residue is substituted for a highly hydrophilic residue types only adds to the destabilization of the interface. (**C**) Homozygous mutant *Lpar1* rats show a craniofacial disorder, using a measure independent of overall head size (eye-to-nose tip length / interocular distance), which was also observed in *Lpar1* knockout mice [37]. Error bars show ± s.e.m. and * indicates statistical difference, $p<0.01$ (n=7 each genotype). (**D**) Homozygous mutant rats are smaller. Error bars show ± s.e.m. and * indicates statistical difference, $p<0.01$ (n=7 each genotype). (**E**) LPAR1^{M318R} is still expressed in the plasma membrane *in vitro*, although at much lower levels than wild-type LPAR1. N-terminally HA-tagged wild-type or mutant receptor were transiently expressed in COS cells. Intact cells were incubated with an antibody against HA before fixing and staining the cells. (**F**) Cell lysates of COS cells expressing wild-type or mutant HA-tagged LPAR1 show the comparable protein levels by western blot analysis. Both an antibody against the HA tag as well as one against human LPAR1 was used to show the expression of the fusion proteins.

Finally, this unique resource of GPCR mutants in the rat is not only suited for studying the *in vivo* effect of the mutated receptor in behavior (e.g. *Gpr19*, *Gpr85*), immunology (e.g. *Ilr8b*, *Gpr65*), or metabolism (e.g. *Mc4r*, *Npy5r*), but it can also be employed for the derivation of primary cell cultures and studying the molecular and functional consequences of the mutation *ex vivo*. Indeed, we isolated embryonic fibroblasts from *Lpar*$^{M318R/M318R}$ rats to study the effect of this mutation in an *in vitro* system, without the necessity of creating transgenic cell lines (R.B., E.C., unpublished results).

DISCUSSION

Single nucleotide polymorphisms (SNPs) are the most common form of human genetic variation [38] and this class of variants is mimicked by the action of ENU that results in the introduction of random point mutations in the genome. Therefore, *in vivo* mutants generated by ENU-driven target-selected mutagenesis can be of great relevance for studying the function of genes and gene variants with effects on human physiology and pathology. Here we made use of this approach to generate mutant models for GPCRs in the rat. The strength of this approach is that in a single experiment a wide range of mutants can be isolated for a large set of genes of interest.

Although the genetic toolbox of the rat has very recently expanded significantly with techniques like transposon insertion mutagenesis [8], targeted zinc-finger nucleases-mediated knockout generation [9] and the availability of pluripotent rat ES cells [10,11], ENU-driven target-selected mutagenesis has developed in the past years into a robust and highly efficient technique. In addition, this approach has the unique characteristic that it simultaneously can provide allelic series of knockout and other alleles, like hypo- and hypermorphic mutants. Also the screen described here resulted in multiple nonsynonymous mutant alleles for the same gene (**Table S2**). Such alleles can be highly informative for understanding gene function and the effects of disease-associated variants identified in human. Finally, the technique does not depend on special (ES) cell lines and/or advanced oocyte or embryo manipulation and the created mutants are not 'transgenic' of nature, since no artificial DNA construct is

integrated into the genome. One disadvantage, however, of ENU mutagenesis could be the presence of background mutations. However, this is a complication that should be taken into account in most approaches for the generation of mutant animals, including homologous recombination-based techniques as it has been shown that long-term culturing of ES cells does result in the accumulation of genetic changes [39]. Nevertheless, the presence of background mutations can relatively easily be controlled or overcome by outcrossing heterozygous carriers to the parental strain [40] and the use of wild-type and heterozygote littermates as controls in phenotypic characterization studies.

Although the use of MMR-deficient background for mutagenesis has greatly increased the efficiency of ENU target-selected mutagenesis in the rat [14], further improvements to the approach can and are still being implemented. The availability of

Table 3: Known human disease genes[a]

Gene	Mutation	Category[b]	MIM Morbid Description (Accession)
CX3CR1	I118K	Protein stability/ folding	Human immunodeficiency virus type 1, susceptibility to (609423) Coronary heart disease, susceptibility to (607339) Macular degeneration, age-related (603075)
DRD3	S355P	Unclear effect[c]	Tremor, hereditary essential (190300) Schizofrenia (181500)
FSHR	V488A	Protein stability/ folding	Ovarian hyperstimulation syndrome (608115) Twinning, dizygotic (276400) Ovarian dysgenesis 1 (233300)
GNRHR	I93T	Protein stability/ folding	Fertile eunuch syndrome (228300) Hypogonadotropic hypogonadism (146110)
GPR56	R96H	Unclear effect[c]	Polymicrogyria, bilateral frontoparietal (606854)
HTR2A	N54D	Unclear effect[c]	Major depressive disorder (608516) Anorexia nervosa, susceptibility to (606788) Schizofrenia (181500) Obsessive-compulsive disorder 1 (164230) Alcohol dependence (103780)
LHCGR	I446N	G-protein activation	Leydig cell hypoplasia, type 1 (238320) Precocious puberty, male-limited (176410)
MC4R	K314X	Protein truncation	Obesity (601665)
PROKR2	S364I	Signal transduction	Kallmann syndrome 3 (244200)
SSTR5	V226A	G-protein activation	Pituitary adenoma, growth hormone-secreting (102200)

[a]According to OMIM database (http://www.ncbi.nlm.nih.gov/omim/). [b]The categories are based on expert interpretation of structural information and bioinformatic predictions unless stated differently. [c]Structural information was not available for these protein domains but bioinformatic analysis by both Polyphen [23] and SIFT [24] software predicted that these mutations are likely to have damaging consequences.

an archive of frozen F_1 rats, which we and others [18] are currently generating can in principle be screened almost infinitely, and will be of great benefit to the rat research community. Additionally, the availability of next-generation sequencing platforms [41], combined with microarray-based genomic enrichment [42], provides promising avenues for further increases in the efficiency of the ENU target-selected mutagenesis approach by rigorously scaling of the targeted mutation discovery effort.

Taken together, we demonstrate that ENU-driven target selected mutagenesis is a highly effective and feasible approach for generating a unique and expandable resource of GPCR mutants in the rat. We established 7 novel potential genetic knockout rat models and over 40 missense mutant lines, including amino acid changes in very conserved GPCR motifs like the (E/D)R(Y/W) motif and the ionic pocket, demonstrating the specific power of random ENU mutagenesis *in vivo*. Selection of the most promising models was aided by extensive bioinformatic analysis, which will also be instrumental for the efficient design of molecular characterization strategies. Notably, at least 10 mutant lines concerned genes, of which polymorphisms in the human orthologs are known to be involved in disease processes (**Table 3**). Furthermore, many of the affected genes have been associated with one or more diseases in recent gene and genome-wide association studies, e.g. *GPR85* in a GWAS study for attention-deficit/hyperactivity disorder (ADHD) [43], illustrating the relevance of these rat models for studying human disease (*see* http://geneticassociationdb.nih.gov, http://www.ncbi.nlm.nih.gov/gap and http://www.genome.gov/gwastudies). Finally, all rat models described here will be made available to the community through the international rat knockout consortium (www.knockoutrat.org).

MATERIALS AND METHODS

Animals and ENU target-selected mutagenesis protocol. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Sciences according to the Dutch legal ethical guidelines. Experiments were designed to minimize the number of required animals and their suffering. ENU treatment of male MSH6 knockout rats ($Msh6^{1Hubr}$) was done as described [14]. Animals were housed under standard conditions in groups of two to three per cage per gender under controlled experimental conditions (12-h light/dark cycle, 21±1°C, 60% relative humidity, food and water *ad libitum*). Genes of interest were screened using PCR amplification followed by capillary sequencing as described [14].

Project management and primer design using LIMSTILL. The resequencing experiments were designed and 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, and entry, archiving and primary interpretation of mutations. The primer design application within LIMSTILL is Primer3-based [44] and parameters are set to design primers with an optimal melting temperature of 58°C. The sequences of the primers used in this project are available upon request.

Mutant effect prediction. Mutation data was retrieved from the GPCRDB [45], which contains a large number of mutants that were obtained from the tinyGRAP database [46] and mutations that were automatically extracted from literature by the software package MuTeXt [47]. The mutants from the GPCRDB that were used in the analyses include mutants at the same position

and in the same receptor as the novel mutants as well as mutants at corresponding residue positions in related proteins. Each mutant in the GPCRDB contains references to literature. The literature describing these mutations was manually checked for relevance and descriptions of the mutant. If the mutation in the mutated receptor or highly homologous receptor has already been described in literature and the results of the experiments are interpreted correctly, we can be reasonably certain of the effects of the mutation on protein function.

When no good mutation data was available residue conservation in the multiple sequence alignments was analyzed. By analyzing residue conservation in multiple sequence alignments we can estimate if the mutant residue is likely to be tolerated. If residue conservation at the position of the mutant is high it is very likely that the mutant has a detrimental effect. If there is little variability the properties of the residues, i.e. charge or aromaticity, are likely to be important. When there is a lot of variability the effect of the mutant is probably largely determined by size constraints if the residue is located on the inside of the protein, or effects because of changes in hydrophobicity if the residue is located at the outside of the protein. Alignments of the primary protein family as well as the superfamily were used. The alignments were obtained from the GPCRDB.

Mutants are best studied in the context of a receptor structure, where the local environment of the mutated residue can shed light on its function and tolerated substitutions. For this, we used the homology models of the receptors that were mutated. Homology models of the receptors were built automatically using in-house software. The recently resolved crystal structures [19,20,21] were used as templates from which the software automatically detected the best for each model. The alignments of the GPCRDB were used to align the model with the template.

The effects of the mutations were estimated by manually combining and interpreting the results of the mutant literature searches and the analyses of alignments and homology models and was performed by experts in this field of research.

***In vitro* fusion protein expression studies.** Wild-type and mutant receptors were N-terminally HA tagged and cloned into the expression vector pcDNA3.1 (Invitrogen). The receptor fusion proteins were expressed in COS-7 cells, which were seeded on a coverslip, and 24 hours after transfection the cells were placed on ice and incubated with DMEM-buffered HEPES containing 0.2% fatty acid-free bovine serum albumin (DHB) for 15 minutes. Subsequently, the cells were incubated for 1 hour with a polyclonal rabbit anti-HA (Abcam Inc, Cambrigde) on ice in DHB at a 1:250 dilution. The cells were methanol-fixed and washed thoroughly with PBS and incubated for one hour in blocking buffer (1% BSA in 0.1% PBS-Tween) at room temperature. The cells were washed three times with PBS and incubated for 1 hour with a secondary anti-rabbit antibody conjugated with FITC (Abcam Inc, Cambrigde) at room temperature in the dark. After three times washing with PBS the coverslips were mounted using Vectashield with DAPI (Brunschwig Chemie, Amsterdam) and analyzed using confocal microscopy. For western blotting COS-7 cells were lysed 24 hours after transfection and the proteins were separated on a SDS gel (10% acrylamide gradient, Bio-Rad) and transferred to a nitrocellulose membrane. The membrane was incubated for 1 hour at room temperature with either a 1:4,000, 1:500 or 1:2,000 dilution of respectively a polyclonal rabbit anti-HA antibody (Abcam Inc, Cambrigde), a polyclonal rabbit anti-human EDG2 (LPAR1) antibody (Abcam Inc, Cambrigde) or a polyclonal rabbit anti-actin antibody (Sigma Aldrich) in blocking buffer followed by an incubation for 1 hours with peroxidase-conjugated, anti-rabbit IgG diluted 1:2,000 in blocking buffer at room temperature. Protein bands were detected by using the enhanced chemiluminescence detection method (ECL, Amersham Biosciences).

REFERENCES

1. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, et al. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A 76: 5818-5819.
2. Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. Mamm Genome 11: 478-483.
3. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
4. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
5. Jacob HJ, Kwitek AE (2002) Rat genetics: attaching physiology and pharmacology to the genome. Nat Rev Genet 3: 33-42.
6. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci U S A 100: 1056-1061.
7. Jacob HJ (1999) Functional genomics and rat models. Genome Res 9: 1013-1016.
8. Kitada K, Ishishita S, Tosaka K, Takahashi R, Ueda M, et al. (2007) Transposon-tagged mutagenesis in the rat. Nat Methods 4: 131-133.
9. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, et al. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325: 433.
10. Buehr M, Meek S, Blair K, Yang J, Ure J, et al. (2008) Capture of authentic embryonic stem cells from rat blastocysts. Cell 135: 1287-1298.
11. Li P, Tong C, Mehrian-Shai R, Jia L, Wu N, et al. (2008) Germline competent embryonic stem cells derived from rat blastocysts. Cell 135: 1299-1310.
12. Gondo Y (2008) Trends in large-scale mouse mutagenesis: from genetics to functional genomics. Nat Rev Genet 9: 803-810.
13. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
14. van Boxtel R, Toonen PW, Verheul M, van Roekel HS, Nijman IJ, et al. (2008) Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics 9: 460.
15. Claij N, van der Wal A, Dekker M, Jansen L, te Riele H (2003) DNA mismatch repair deficiency stimulates N-ethyl-N-nitrosourea-induced mutagenesis and lymphomagenesis. Cancer Res 63: 2062-2066.
16. Ma P, Zemmel R (2002) Value of novelty? Nat Rev Drug Discov 1: 571-572.
17. Smits BM, Mudde J, Plasterk RH, Cuppen E (2004) Target-selected mutagenesis of the rat. Genomics 83: 332-334.
18. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
19. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, et al. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289: 739-745.
20. Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, et al. (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature 450: 383-387.
21. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, et al. (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 454: 486-491.
22. Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, et al. (2001) Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem 276: 29171-29177.
23. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25: 2745-2751.

24. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31: 3812-3814.
25. Hines J, Fluharty SJ, Yee DK (2003) Structural determinants for the activation mechanism of the angiotensin II type 1 receptor differ for phosphoinositide hydrolysis and mitogen-activated protein kinase pathways. Biochem Pharmacol 66: 251-262.
26. Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, et al. (2003) Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. N Engl J Med 348: 1085-1095.
27. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.
28. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
29. Homberg JR, Olivier JD, Smits BM, Mul JD, Mudde J, et al. (2007) Characterization of the serotonin transporter knockout rat: a selective change in the functioning of the serotonergic system. Neuroscience 146: 1662-1676.
30. Mul J, Yi CX, van den Berg SA, Ruiter M, Toonen P, et al. (2009) Pmch expression during early development is critical for normal energy homeostasis. Am J Physiol Endocrinol Metab.
31. VanLeeuwen D, Steffey ME, Donahue C, Ho G, MacKenzie RG (2003) Cell surface expression of the melanocortin-4 receptor is dependent on a C-terminal di-isoleucine sequence at codons 316/3J Biol Chem 278: 15935-15940.
32. Urs NM, Kowalczyk AP, Radhakrishna H (2008) Different mechanisms regulate lysophosphatidic acid (LPA)-dependent versus phorbol ester-dependent internalization of the LPA1 receptor. J Biol Chem 283: 5249-5257.
33. Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, et al. (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell 88: 131-141.
34. Mills GB, Moolenaar WH (2003) The emerging role of lysophosphatidic acid in cancer. Nat Rev Cancer 3: 582-591.
35. Harrison SM, Reavill C, Brown G, Brown JT, Cluderay JE, et al. (2003) LPA1 receptor-deficient mice have phenotypic changes observed in psychiatric disease. Mol Cell Neurosci 24: 1170-1179.
36. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, et al. (2003) Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci U S A 100: 2290-2295.
37. Contos JJ, Fukushima N, Weiner JA, Kaushal D, Chun J (2000) Requirement for the lpA1 lysophosphatidic acid receptor gene in normal suckling behavior. Proc Natl Acad Sci U S A 97: 13384-13389.
38. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-236.
39. Liang Q, Conte N, Skarnes WC, Bradley A (2008) Extensive genomic copy number variation in embryonic stem cells. Proc Natl Acad Sci U S A 105: 17453-17456.
40. Keays DA, Clark TG, Flint J (2006) Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. Mamm Genome 17: 230-238.
41. Ansorge WJ (2009) Next-generation DNA sequencing techniques. N Biotechnol 25: 195-203.
42. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, et al. (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4: 907-909.
43. Anney RJ, Lasky-Su J, O'Dushlaine C, Kenny E, Neale BM, et al. (2008) Conduct disorder and ADHD: evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. Am J Med Genet B Neuropsychiatr Genet 147B: 1369-1378.
44. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386.
45. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, et al. (2003) GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 31: 294-297.
46. Beukers MW, Kristiansen I, AP IJ, Edvardsen I (1999) TinyGRAP database: a bioinformatics tool to mine G-protein-

coupled receptor mutant data. Trends Pharmacol Sci 20: 475-477.
47. Horn F, Lau AL, Cohen FE (2004) Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. Bioinformatics 20: 557-568.

SUPPLEMENTARY DATA

For Tables S1 and S2 readers are referred to the The Pharmacogenomics Journal website (http://www.nature.com/tpj).

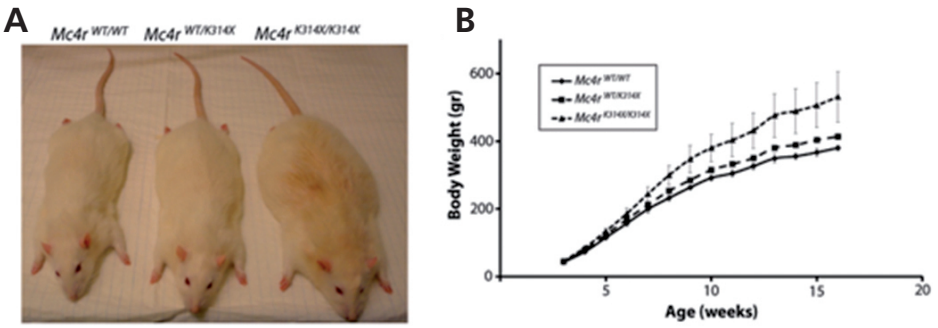

Figure S1: *Mc4r*$^{K314X/K314X}$ show increased body weight. (**A**) Representative picture of a *Mc4r*$^{WT/WT}$, *Mc4r*$^{WT/K314X}$ and *Mc4r*$^{K314X/K314X}$ rat. (**B**) Growth curve of *Mc4r*$^{WT/WT}$, *Mc4r*$^{WT/K314X}$ and *Mc4r*$^{K314X/K314X}$ male rats. Error bars display s.e.m. (n=2 for each genotype).

6

NEXT GENERATION REVERSE GENETICS: TOWARDS GENOME-WIDE COLLECTIONS OF GENE KNOCKOUTS AND MUTANTS USING NEXT-GENERATION SEQUENCING

Ruben van Boxtel[1], Isaac J. Nijman[1], Michal Mokry[1], Pim W. Toonen[1] and Edwin Cuppen[1,2]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2]Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, Utrecht, The Netherlands

ABSTRACT

Chemical-based target-selected mutagenesis is an effective method for generating *in vivo* mutant models that has been applied successfully in a wide range of organisms, including plants, invertebrates and vertebrates. However, its efficiency is limited by costs and throughput of the targeted mutation discovery step. We developed a protocol based on genomic enrichment using microarray-based capturing of multiplexed barcoded samples, followed by SOLiD-based next-generation sequencing, which results in highly efficient mutation discovery with high sensitivity and specificity. In a single enrichment and sequencing run 770 genes in 20 rats were screened for induced mutations. Besides all known variants, novel knockout and missense alleles were retrieved at the expected frequency, with an overall false-positive rate of less than 1 in 6 million basepairs. Eventually, this approach can be used in any species of interest for the generation of genome-wide collections of knockouts as well as allelic series of missense variants.

INTRODUCTION

Target-selected mutagenesis, which is also known as TILLING (Targeting Induced Local Lesions In Genomes) [1,2], has become a universal and versatile method for the generation of genetic mutants, especially in species for which no ES cell-based homologues recombination-driven techniques are available, like the rat [3,4], zebrafish [5], medaka [6], C. elegans [7] and a wide range of plant species, including Arabidopsis [8] and maize [9]. The principle of the approach is that mutations are randomly introduced in a large set of individuals (thousands to several thousands), followed by targeted discovery of mutations in genes of interest, e.g. by nuclease-based [10] or Mu transposase-based mismatch detection [11] or PCR-based dideoxy resequencing [5].

In vertebrates, male animals are mutagenized by intraperitonal injection of *N*-ethyl-*N*-nitrosourea (ENU), which has been shown to be the most potent germ line mutator [12], while in plants ethyl methanesulfonate (EMS) treatment of seeds, is the preferred method of choice. These chemicals very effectively introduce random point mutations, which after outcrossing results in a heterogeneous pool of F_1 individuals all carrying independent heterozygous mutations. Subsequently, the DNA of these F_1 individuals is screened for heterozygous mutations in genes of interest, with the objective to identify induced mutations that affect protein function, e.g. by introducing a premature stopcodon or by changing a functionally important residue.

An advantage of the chemical-driven target-selected mutagenesis approach is that it not only results in the retrieval of genetic knockouts, but also in allelic series of amino acid changes. In this sense, ENU-induced mutations mimic the most common form of human genetic variation, namely single nucleotide polymorphisms (SNPs), making the resulting genetic models highly relevant for studying aspects of complex human diseases. In addition, mutants generated using target-selected mutagenesis are not transgenic since no artificial DNA is introduced, which is especially relevant for crop species.

The efficiency of the approach essentially depends on the induced mutation frequency and the efficiency of mutation discovery. While the first aspect is largely dependent on species and genetic background [3,13], mutation rates can, at least in the rat, be increased by about 2.5 fold when using a mismatch repair deficient background [14], but further improvements are unlikely to be implemented for this step. However, mutation discovery remains a tedious process that is currently largely limited to sequential analysis of individual genes. The recently emerging massively parallel sequencing technologies combined with microarray-based capture of genomic regions of interest now provide the tools for also boosting the efficiency of the mutation discovery step. Here, we show that a combination of an optimized microarray-based genomic enrichment procedure followed by SOLiD-based next-generation sequencing allows for the highly efficient discovery of ENU-induced mutations in the rat in a multiplexed fashion for more than a thousand genes in parallel.

RESULTS

Sample preparation and targeted multiplexed enrichment

To validate the next-generation forward genetic procedure, we made use of a previously generated living repository of mutant F_1 rats that were generated from mutagenesis of mismatch repair deficient *msh6*$^{-/-}$ male founder rats [14,15]. The animals that were used here were screened previously by traditional dideoxy resequencing of PCR-amplified genomic segments (exons) and carried known mutations [16]. First, we designed custom microarrays (Agilent SurePrint 244K) to enrich for genomic regions of interest [17]. These designs included the regions containing the known mutations as well as all conserved non-odorant G protein-coupled receptors plus an additional ~500 kb of sequence encoding MIM morbid genes (*see* **Table S1** for complete target design). In total, the complete coding sequence of 770 genes was covered, adding up to 1.4 Mb of genomic sequence.

For the first experiment, the genomic DNA of 10 F_1 animals was sheared into smaller fragments (between 50 – 150 bp) and cloned between shortened P1 and internal adapters (IA) (**Fig. 1**). During the ligation-mediated PCR (LM-PCR) amplification step, barcodes compatible with AB/SOLID sequencing were introduced between IA and P2 adapter, unique for each animal. Next, the 10 samples were pooled equimolarly and subjected to multiplex enrichment on a single microarray slide and AB/SOLiD sequencing in a single full slide run. Approximately 16.6 ±3.8 million reads were obtained for every animal. In contrast to standard non-multiplexed enrichment, where we routinely obtain between 60 and 90% of reads on target, we only acquired about 35% of the reads on target across all samples using multiplexed enrichment. Nevertheless, an average coverage of about 200x for the targeted regions was obtained for all samples, which is sufficient for reliable variant discovery (**Table 1**).

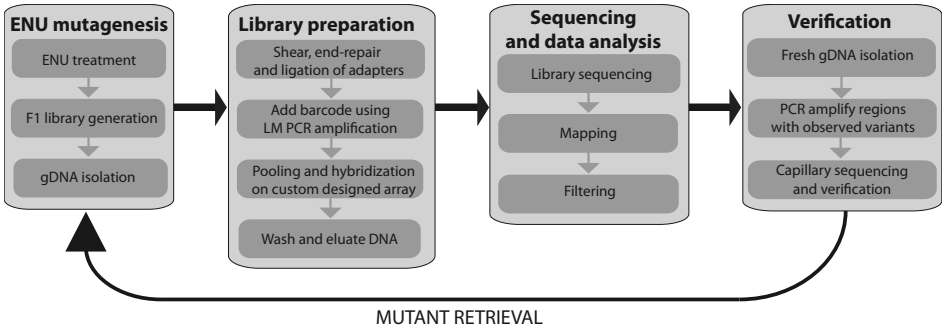

Figure 1: Experimental design. The next-generation reverse genetics procedure can be divided into 4 different experimental steps. First, mutagenized individuals are outcrossed to generate an F_1 library of which DNA is isolated. Second, the DNA is used to generate a sequencing library. SOLiD barcodes are introduced in order to pool F_1 animals before the multiplex enrichment for genes of interest. Next, the library is sequenced and the data is analyzed. Finally, the identified variants are verified by an independent assay, e.g. traditional capillary sequencing.

Table 1: Sequencing statistics for enrichment of a multiplexed pool of 10 barcoded rat DNA samples for a target region of ~1.4 Mb using standard conditions

Rat	On target (%)	Total bp covered ≥ 20x[a]	Coverage of regions of interest		Fold coverage[a]	
			≥ 1x (%)	≥ 20x (%)	Mean	Median
1	35	2.418.005	99,28	93,88	153	138
2	35	2.570.707	99,33	94,87	185	166
3	35	2.580.781	99,29	94,83	180	163
4	34	2.697.970	99,34	95,07	197	177
5	34	2.661.490	99,33	95,15	188	171
6	35	2.877.912	99,40	95,73	220	199
7	35	3.421.266	99,45	96,19	317	280
8	35	2.935.474	99,43	95,93	228	206
9	35	2.856.780	99,40	95,60	214	194
10	35	2.411.699	99,25	93,97	155	140
Average	35	2.743.208	99,35	95,12	204	183

[a]including non-targeted regions (e.g. exon-flanking intronic sequences)

The relatively low enrichment specificity might be due to the long barcode-containing 3' adaptors (52 nt) that could potentially allow carry-over of non-targeted DNA molecules by cross-hybridization of two independent DNA library molecules [18]. Therefore, we designed specific primers to block the barcode sequences before and during hybridization-based enrichment (**Fig. 2A**). New libraries were prepared for a second set of 20 animals using the blocking primers for the enrichment (**Fig. 2B**). Now, the enrichment efficiency was increased to an average of 60% of the reads on target, resulting in 188x average coverage per animal from a single slide AB/SOLiD run (**Table 2**).

Mutation discovery

To identify heterozygous mutations in the sequencing data, we developed a bioinformatic pipeline that takes into account allele frequency and coverage on both DNA strands. Using standard parameters all 31 known mutations in the 30 F_1 animals that were screened in this study were picked up except for one. Manual inspection of the mutation that was missed revealed that the coverage of the mutant allele was below the threshold level, but also that this mutation was flanked by a linked SNP two basepairs downstream of the mutation. The presence of two mismatches between the probe and the mutant allele may result in decreased capturing efficiency compared to DNA fragments originating from the wild-type allele. Furthermore, the presence of two additional polymorphisms may also affect mapping efficiency of the short read (50-mer) sequencing data, which would further increase the reference bias.

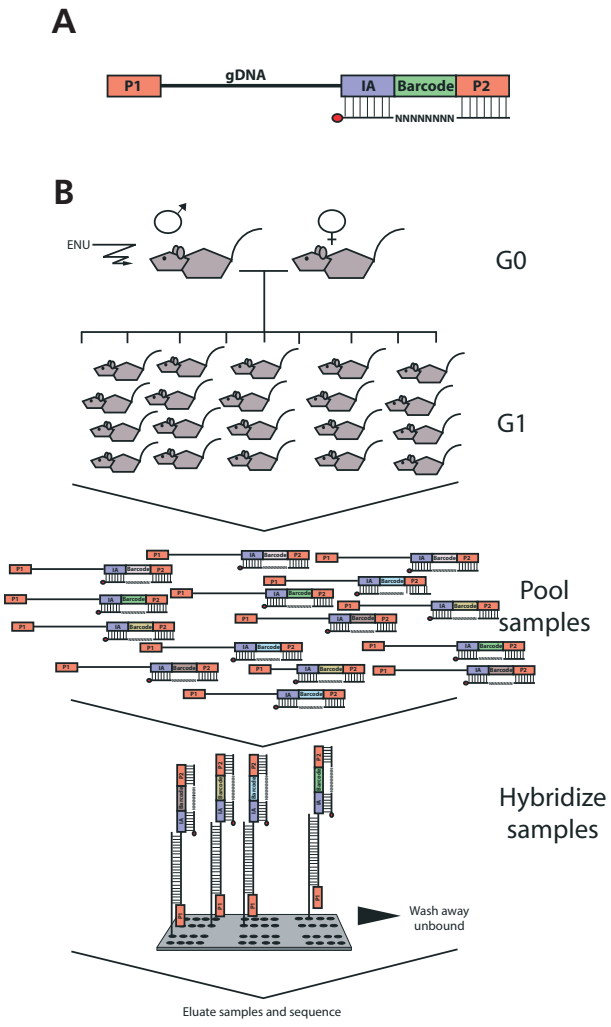

Figure 2: Sequencing library preparation. (**A**) Specific blocking primer to increase hybridization efficiency. The primer is complementary to the internal adapter (IA) and the P2 adapter with degenerate bases in between that hybridize to the different barcodes. (**B**) The DNA of 20 rats was used to generate one sequencing library by adding individual-specific barcodes to the DNA fragments, pooling the DNA samples followed by enrichment for genomic regions of interest by microarray-based hybridization.

Besides the known 31 mutations, an additional 133 novel heterozygous variations were identified using standard settings (**Table 3**). PCR amplicons were designed to confirm these mutations by capillary sequencing. For 21 variants, we failed to design a working amplicon or capillary sequencing failed due to low sequence complexity (e.g. simple repeats). The fact that these loci could be screened by AB/SOLiD sequencing, can be considered a specific advantage of the current approach compared to traditional PCR-based dideoxy sequencing. A total of 100 variations were readily confirmed, while 7 mutations could not be confirmed by capillary sequencing. However, inspection of the allele frequencies in combination with sequence coverage in the SOLiD data (**Fig. 3**) suggests that most of these mutations are real mutations that may be false negatives in traditional PCR-based dideoxy resequencing.

Table 2: Sequencing statistics for enrichment of a multiplexed pool of 20 barcoded rat DNA samples for a target region of ~1.4 Mb using hybridization blocking primers

			Coverage of regions of interest		Fold coverage[a]	
Rat	On target (%)	Total bp covered ≥ 20x[a]	≥ 1x (%)	≥ 20x (%)	Mean	Median
11	62	1.569.935	99,36	96,39	185	180
12	61	1.591.579	99,41	96,70	180	178
13	61	1.557.874	99,40	96,79	177	175
14	61	1.561.642	99,36	96,40	179	175
15	61	1.634.892	99,37	97,09	229	225
16	61	1.515.813	99,42	97,27	252	247
17	61	1.575.129	99,18	93,24	104	101
18	59	1.595.990	99,35	95,25	135	132
19	59	1.692.246	99,30	95,32	152	146
20	58	1.485.761	99,34	96,33	199	191
21	58	1.580.952	99,36	96,48	195	189
22	59	1.506.383	99,37	96,83	214	210
23	59	1.552.695	99,36	96,34	186	181
24	59	1.546.423	99,35	96,53	193	189
25	59	1.562.846	99,41	96,94	204	201
26	59	1.407.492	99,24	95,12	157	151
27	59	1.614.455	99,36	96,29	183	179
28	59	1.501.378	99,32	95,52	158	151
29	59	1.554.242	99,36	96,28	181	177
30	59	1.546.249	99,43	97,43	293	285
Average	59	1.537.312	99,36	96,38	196	191

[a]including non-targeted regions (e.g. exon-flanking intronic sequences)

Of the 100 confirmed variations, 30 were observed in more than one F_1 animals and are therefore believed to be common single nucleotide polymorphisms (SNPs) in the outbred rat strain (Wistar) that was used. The 75 remaining variants are considered to be ENU-induced mutations and include 2 nonsense, 1 translational start site, 53 missense, 14 silent, and 5 non-coding mutations (**Table 3**). Notably, the nonsense mutation in *Grm2* (*Grm2*^{C407X}) and the translational start site mutation in *Prlhr* (*Prlhr*^{1MX}) were observed in 2 different F_1 animals, suggesting clonal origin from the same mutagenized spermatogonial stem cell or mutational hotspots [7]e¤. The *Grm2*^{C407X} mutation results in a truncated form of the G protein-coupled receptor GRM2, which lacks all 7 transmembrane domains and is thus likely to represent a functional knockout allele. The other nonsense mutation introduces a premature stopcodon in the N-terminus

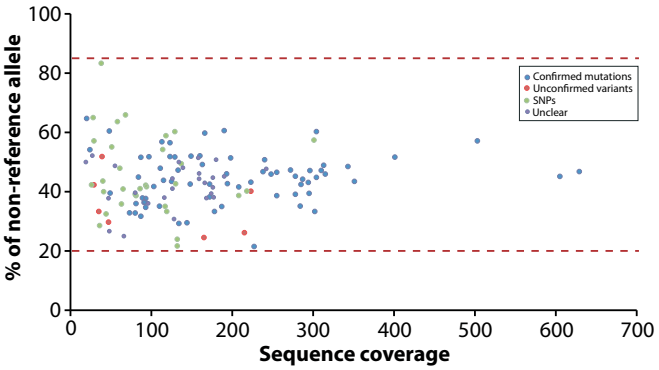

Figure 3: Reliability plot of identified variants. The allele frequencies of the identified variants are plotted against sequence coverage in the SOLiD data. The red dashed lines indicate the upper and lower allele frequency boundaries between which variants are considered to be heterozygous. False positive variants tend to have a low percentage of non-reference alleles and a low coverage.

of the peroxisomal membrane protein PXMP3 (*Pxmp3*^{Y106X}) and is also expected to result in a complete loss-of-function. Similarly, *Prlhr*^{1MX} will most probably result in a complete functional knockout allele as the first ATG triplet within *Prlhr* is now out of frame and causes an aberrant transcript.

DISCUSSION

Here, we demonstrate the possibilities of using microarray-based genomic selection combined with AB/SOLiD next-generation sequencing in reverse genetic screens by assaying a total ~58 Mb divided over 30 animals and identifying 75 chemically induced mutations. This translates into a mutation frequency of 1 in 775 kb, which is similar to results obtained in previous screens in this genetic background [14,16]. This indicates that the next-generation sequencing-based approach used here is equally sensitive as traditional capillary dideoxy sequencing. Moreover, with only 7 potential false positives (7% of all candidate mutations), the overall false positive rate in our experiments is less than 1 in 6 million base pairs surveyed. Although this rate may go slightly up when lower coverage sequencing is used, the obtained rates are much better than routinely obtained by high-throughput capillary dideoxy resequencing (~20% of the candidate mutations can not be confirmed in traditional screens across various species that we have screened, E.C. personal communication). Although we may still miss mutations, especially for regions with lower capturing efficiency and consequently lower sequencing coverage or in the direct neighborhood of linked polymorphisms, we found that high-throughput PCR-based resequencing also has a significant false negative rate as we did identify several mutations in genes that were previously screened by capillary sequencing (*see* **Table S2**).

The results described here provide a solid proof of principle for the use of next-generation sequencing technologies for target-selected reverse genetics procedures and demonstrate a very significant improvement in the efficiency of the target-selected mutagenesis or TILLING approach in general. Previously, it was already shown that a collection of only 5,000 mutagenized *C.elegans* individuals contains one or more knockout alleles for virtually every gene and that a missense allele for every 10th amino acid is present [7]. However, for identification of these alleles, the complete exome would have to be screened. Now, with continuously increasing throughput of NGS platforms and efficient parallelizable in solution tools for whole exome enrichment [17,19] these genome-wide collections of knockout mutants as well as extensive allelic series come within reach for a wide range of species. The next challenge will be to generate large libraries of mutant F_1 individuals, which is not equally attainable for every species. Archives of cryopreserved sperm are currently being established for the rat, which can be used for rederivation of the mutant strains by intracytoplasmatic sperm injection (ICSI) [11]. Eventually, mutant models (knockout as well as allelic series of missense muations) for genes of interest in a range of species may be ordered from pre-screened public collections, thereby facilitating functional genomics research and assisting interpretation of personal genomes.

MATERIALS AND METHODS

Animals and ENU target-selected mutagenesis protocol. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Sciences according to the Dutch legal ethical guidelines. Experiments were designed to minimize the number of required animals and their suffering. ENU treatment of male MSH6 knockout rats (Msh6^{1Hubr}) was done as described [14]. Animals were housed under standard conditions in groups of two to three per cage per gender under controlled experimental conditions (12-h light/dark cycle, 21±1°C, 60% relative humidity, food and water *ad libitum*).

Library preparation. Genomic DNA of each F_1 individual was fragmented for 6 minutes using a Covaris S2 sonicator (6 x 16 mm AFA fiber Tube, duty cycle: 20%, intensity: 5, cycles/burst: 200, frequency sweeping). After fragmentation, fragments were blunt-ended and phosphorylated at the 5' end using End-it Kit (Epicentre) according to the manufacturer's instructions, followed by ligation of double-stranded short adapters (adapter 1: pre-annealed duplex of 5'-CTA TGG GCA GTC GGT GAT-3' and 5'-ATC ACC GAC TGC CCA TAG TTT-3' and adapter 2: pre-annealed duplex of 5'-CGC CTT GGC CGT ACA GCA G-3' and 5'-GCT GTA CGG CCA AGG CG-3'; all oligo's were acquired through Integrated DNA Technologies (Coralville, IA) and pre-annealing was done by mixing complementary oligonucleotides at 500 µM concentration and running on thermocycler with the following program: 95˚C for 3 min, 80˚C for 3 min, 70˚C for 3 min, 60˚C for 3 min, 50˚C for 3 min, 40˚C for 3 min and 4˚C hold). Ligation was performed using Quick ligation kit (New England Biolabs) with 1 µg of fragmented DNA, 750 nM adaptor 1 and adaptor 2, 150 µl of 2x Quick ligation buffer, and 5 µl Quick Ligase in a total volume of 300 µl. Samples were purified on Ampure beads (Agencourt) and amplified using 400 µl of Platinum PCR Supermix with 750 nM of both amplification PCR primers (P1_short: 5'-CTA TGG GCA GTC GGT GAT-3' and P2: 5'-CTG CCC CGG GTT CCT CAT TCT CTN NNN NNN NNN CTG CTG TAC GGC CAA GGC G-3', where N represent unique barcode sequence for each library), 2.5 U of Pfu DNA polymerase (Stratagene) and 5 U Taq DNA polymerase (Bioline). Before ligation-mediated amplification, the PCR sample was incubated at 72˚C for 5 minutes in

PCR mix to perform nick translation on non-ligated 3'-ends. After 6 cycles of amplification, the library DNA was purified on Ampure beads and the quality was checked on a gel for the proper size range and the absence of adapter dimers and heterodimers. Ten (pool1) and 20 (pool2) individual samples have been pooled together in equal parts and size selected on 4% agarose gel for 125-175bp fraction.

Capture array design. We extracted the exonic sequence from 770 genes from Ensembl (build53; *see* **Table S2**) with a footprint of 1.392.385 bp. A custom PERL script is used to design 60bp oligos within a sliding window of 10 bp. Within each window the most optimal probe is selected based on melting temperature and absence of homopolymer stretches. To exclude potentially repetitive elements from the design, all probes were compared to the reference genome using BLAST and those returning more than one hit (as defined by a 60% match of probe sequence) were discarded from the design. From the requested regions, 97.5% (1.357.860 bp) could be covered by 239.110 probes matching these criteria. Probes were synthesized on custom 244k Agilent arrays with randomized positions.

Enrichment hybridization and elution. Prior to hybridization size selected library was amplified using 10 PCR cycles in 1000 µl of Platinum PCR Supermix with 750 nM of both amplification PCR primers (P1_short: 5'-CTATGGGCAGTCGGTGAT-3' and P2_short: 5'-CCG GGT TCC TCA TTC TCT-3') and 5 U of Pfu DNA polymerase to produce a sufficient amount of library DNA necessary for enrichment. Amplified library DNA was subsequently purified using a MinElute Reaction Cleanup Kit (Qiagen). Amplified DNA was mixed with 10x weight excess of Hybloc rat DNA (Applied Genetics Laboratories, Melbourne, FL), 10x weight excess of both blocking oligos in case of pool2 (block1: 5'-CCC CGG GTT CCT CAT TCT CTR NGN KRN RNN CTG CTG TAC GGC CAA GGC G/3ddC/-3' and block2: 5'-CGC CTT GGC CGT ACA GCA GNN YNY MNC NYA GAG AAT GAG GAA CCC GGG G/3ddC/-3') and concentrated using a speedvac to a final volume of 12.3 µl. DNA was mixed with 31.7 µl Nimblegen aCGH hybridization solution and denatured at 95°C for 5 minutes. After denaturing the sample was hybridized for 65 hours at 42°C on a 4-bay MAUI hybridization station using an active mixing MAUI AO chamber (MAUI). After hybridization, the array was washed using the Nimblegen Wash Buffer Kit according to the user's guide for aCGH hybridization. Elution was performed using 800 µl of elution buffer (10 mM Tris pH 8.0) in an Agilent Microarray Hybridization Chamber at 95°C for 30 minutes. After 30 minutes the chamber was quickly disassembled and elution buffer collected into a separate 1.5ml tube. Eluted library DNA was concentrated in a speedvac to a final volume of 50 µl and amplified with a limited number of PCR cycles (13 cycles) with full-length primers (amp-P1: 5'-CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG AT-3' and amp-P2: 5'-CTG CCC CGG GTT CCT CAT TCT-3'), to introduce full length adapter sequences required for SOLiD sequencing.

AB/SOLiD sequencing. To achieve clonal amplification of library fragments on the surface of sequencing beads, emulsion PCR (emPCR) was performed according to the manufacturer's instructions (Applied Biosystems). 1500 pg of double stranded library DNA was added to 5.6 ml of PCR mix containing 1x PCR Gold Buffer (Applied Biosystems), 3000 U AmpliTaq Gold, 40nM emPCR primer 1, 3 µM of emPCR primer 2, 3.5 mM of each deoxynucleotide, 25mM MgCl2 and 1.6 billion SOLiD sequencing beads (Applied Biosystems). PCR mix was added to SOLiD ePCR Tube containing 9 ml of oil phase and emulsified using ULTRA-TURRAX Tube Drive (IKA). The PCR emulsion was dispensed into 96-well plate and cycled for 60 cycles. After amplification the emulsion was broken with butanol, beads were enriched for template-positive beads, 3'-end extended and covalently attached onto sequencing slides. Both pool1 and pool2 were sequenced on a single sequencing slide on the AB/SOLiD system version 3 to produce 50 bases long reads and barcodes.

Mapping of sequencing data and SNP calling. Sequencing reads were mapped against the reference genome (Rat Ensembl Build 56.34x) using the Maq package [20], which allows mapping in SOLiD color space corresponding to dinucleotide encoding of the sequenced DNA with following settings: number of maximum mismatches that can always be found -n 3, threshold on the sum of mismatching base qualities -e 150. Raw variant positions were called by

the Maq package and filtered using custom scripts (available upon request). For stringent variant calling we used the following filtering settings: 1) positions with lower than 20x and higher than 5000x coverage were excluded, 2) each of non-reference alleles had to be supported by at least 3 independent reads (as determined by different read start positions) separately on positive and negative strand with quality > 10, 3) the non-reference allele should account for at least 20% of the reads covering the polymorphic position, and 4) the ratio between + and – strand reads should be between 1/9 and 9. Positions that passed these filtering settings were considered as candidate variant. Since the ENU-induced mutations are heterozygous in the assayed F_1 animals, a variant was qualified as heterozygous when the fraction of non-reference alleles was between 20% and 85%.

REFERENCES

1. Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135: 630-636.
2. Stemple DL (2004) TILLING--a high-throughput harvest for functional genomics. Nat Rev Genet 5: 145-150.
3. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
4. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
5. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RH, et al. (2003) Efficient target-selected mutagenesis in zebrafish. Genome Res 13: 2700-2707.
6. Taniguchi Y, Takeda S, Furutani-Seiki M, Kamei Y, Todo T, et al. (2006) Generation of medaka gene knockout models by target-selected mutagenesis. Genome Biol 7: R116.
7. Cuppen E, Gort E, Hazendonk E, Mudde J, van de Belt J, et al. (2007) Efficient target-selected mutagenesis in Caenorhabditis elegans: toward a knockout for every gene. Genome Res 17: 649-658.
8. Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13: 524-530.
9. Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, et al. (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12.
10. Till BJ, Burtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32: 2632-2641.
11. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
12. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, et al. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A 76: 5818-5819.
13. Justice MJ, Carpenter DA, Favor J, Neuhauser-Klaus A, Hrabe de Angelis M, et al. (2000) Effects of ENU dosage on mouse strains. Mamm Genome 11: 484-488.
14. van Boxtel R, Toonen PW, Verheul M, van Roekel HS, Nijman IJ, et al. (2008) Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics 9: 460.
15. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
16. van Boxtel R, Vroling B, Toonen P, Nijman IJ, van Roekel H, et al. (2010) Systematic generation of in vivo G protein-coupled receptors mutants in the rat. Pharmacogenomics J in press.
17. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, et al. (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4: 907-909.

18. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, et al. (2009) Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat Protoc 4: 960-974.
19. Ansorge WJ (2009) Next-generation DNA sequencing techniques. N Biotechnol 25: 195-203.
20. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

SUPPLEMENTARY DATA

Tables S1 and S2 are available upon request.

7

A MUTATION IN HELIX 8 OF THE RAT LYSOPHOSPHATIDIC ACID RECEPTOR 1 CAUSES A LOSS-OF-FUNCTION PHENOTYPE THROUGH CONSTITUTIVE ARRESTIN-MEDIATED DESENSITIZATION

Ruben van Boxtel[1] and Edwin Cuppen[1,2]

[1]Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2]Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, Utrecht, The Netherlands

ABSTRACT

G protein-coupled receptors (GPCRs) regulate many physiological processes by controlling a myriad of intracellular signaling systems in response to external stimuli. GPCR signaling is tightly regulated by a variety of fine-tune mechanisms, including β-arrestin-mediated desensitization and internalization. Here, a novel ENU-induced lysophosphatidic acid receptor 1 (LPAR1) rat mutant is described that carries a missense mutation (M318R) in the cytoplasmatic helix 8. This mutation results in a loss-of-function phenotype *in vivo*, as well as impaired LPA-induced ERK1/2 phosphorylation. Moreover, *in vitro* analysis revealed decreased cell surface expression because of spontaneous internalization, which was mediated by agonist-independent recruitment of β-arrestin. In line with this, substitution of the analogous residue in the prototypic β2-adrenergic receptor (ADRB2) resulted in a similar phenotype. Taken together, our experiments show that missense mutations in the helix 8 of GPCRs can result in complete loss-of-function phenotypes *in vivo* by constitutive arrestin-mediated desensitization of the receptor.

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute one of the largest membrane-bound protein families, which are characterized by 7 membrane spanning domains with an extracellular N-terminus and intracellular C-terminus. GPCRs sense a wide variety of extracellular signals, including hormones, neurotransmitters and lipids, and thereby regulate many cellular processes. Moreover, this receptor family represents by far the largest class of targets for current drugs as well as for the development of novel small-molecule medicines [1]. Classical GPCR signaling is thought to result from ligand-dependent stabilization of an active conformation of the receptor [2]. The activated receptor can subsequently interact with heterotrimeric G proteins, resulting in the dissociation of this complex into a G_{α} subunit and $G_{\beta\gamma}$ dimmer that have an independent capacity for downstream signaling. Given the important role in regulating essential cellular function, many mechanisms have been evolved for fine-tuning receptor signaling, like desensitization, even in continuing presence of agonist. Desensitization is initiated by agonist-dependent phosphorylation of the C-terminus of the receptor, usually carried out by GPCR kinases (GRKs), followed by recruitment of β-arrestins [3]. β-arrestins silence activated receptors by sterically inhibiting further G protein coupling, by acting as a scaffold for enzymes that attenuate second messengers, like the degradation of cAMP by PDE4, and by mediating receptor internalization via clathrin-coated pits [4]. More recently it has been appreciated that β-arrestins can initiate intracellular signaling pathways, including MAP kinase ERK, in a GPCR dependent matter [4]. An extra layer of complexity in GPCR signaling was added when selective agonist were discovered, capable of signaling via only the G protein pathway or β-arrestin pathway, which can have distinct cellular outcomes [5].

Lysophosphatidic acid (LPA) mediates a variety of biological processes, including cell proliferation, migration and survival, by signaling via specific GPCRs [6]. Currently, at least 5 different LPA receptors are known (LPAR1-5) of which LPAR1 is the most widely expressed in both humans and rodents [7]. Aberrant LPA signaling has been associated with a variety of human disease, including cancer [8]. Indeed, it was recently shown that overexpression of autotaxin (ATX), the primary enzyme producing LPA, or LPAR1-3 *in vivo* increased late-onset mammary tumorigenesis, invasion and metastases [9]. Furthermore, abundant expression of LPAR1-3 in both the embryonic as well as adult brain, suggest an important role for LPA signaling in brain development [7]. This is exemplified by the observation that LPA enhances cortical growth and folding *ex vivo*, which was absent in LPAR1 and LPAR2 double knockout mouse brains [10]. Targeted disruption of LPAR1 in mice resulted in increased perinatal lethality, reduced body size, craniofacial disorder and increased apoptosis in sciatic nerve Schwann cells [11], demonstrating nonredundant function for this receptor *in vivo*. Furthermore, an independently generated LPAR1-deficient mouse model displayed phenotypic changes observed in psychiatric disorders, including a deficit in prepulse inhibition and changed levels of the neurotransmitter serotonin, as well as craniofacial aberrancies [12].

The laboratory rat *Rattus norvegicus* is one of the most used model organism in biomedical research and is well-suited for studying human disease [13], including cancer [14,15,16] and disorders affecting higher brain function, like schizofrenia, anxiety, depression and addiction. Using ENU-driven target-selected mutagenesis [17], a mutation was identified in *Lpar1* that resulted in the substitution of methionine 318 into an arginine in the putative helix 8 of the receptor [18]. The helix 8 region of GPCRs forms an amphipathic helix that lies parallel to the plasma membrane and is anchored by palmitoylated cysteine groups [19]. Importantly, upon receptor activation helix 8 is believed to move, suggesting functional importance of this region in this process. The hydrophobicity of the mutated residue in LPAR1 is highly conserved in the GPCR class A family. Furthermore, this position is analogous to the phenylalanine of the NPxxY$(x)_{5,6}$F motif found in many GPCRs, including the prototypic Rhodopsin and β2 adrenergic receptor (ADRB2) (**Fig. 1A**). The residue sticks into a hydrophobic pocket and contacts the tyrosine of the same motif, which is important for the folding of helix 8. Computational analyses suggest that the amino acid change in the LPAR1 mutant will disrupt this hydrophobic interaction and additionally, an arginine is too big to fit in this pocket and will most probably result in incorrect packing of helix 8 [18]. Supportive for these predictions was the observation that homozygous mutant animals for the M318R mutation in LPAR1 display an apparent loss-of-function phenotype, characterized by smaller size and a craniofacial disorder [18] comparable as in LPAR1 knockout mice [11]. However, the phenotype observed in *Lpar1*$^{M318R/M318R}$ rats is less severe as in knockout mice, which additionally display high neonatal lethality because of a suckling disorder [11], reflecting either species-specificity or the hypomorphic nature of the M318R mutation.

Here we show that the LPAR1^{M318R} loss-of-function phenotype may result from constitutive desensitization of the receptor by recruiting β-arrestin in an agonist-independent fashion. As a result, the pool of cell surface expressed receptor is decreased resulting in a hypomorphic signaling response upon LPA. Furthermore, mutating the analogous residue in human ADRB2 also seems to result in constitutive desensitization *in vitro*, generalizing the involvement of helix 8 in this process.

RESULTS

Agonist-independent internalization of LPAR1^{M318R} *in vitro*

Lpar1^{M318R} homozygous mutant rats display characteristics of a loss-of-function phenotype, namely smaller size and a craniofacial disorder [18]. Given the importance of the mutated residue in folding of the putative helix 8 and the role of this helix in signal transduction, we set out to investigate the molecular and functional consequences of the mutation in GPCR signaling. First, we tested if LPAR1^{M318R} could still be transported and expressed in the plasma membrane, the site of action of the receptor. Although membrane expression was still observed, it was considerably less

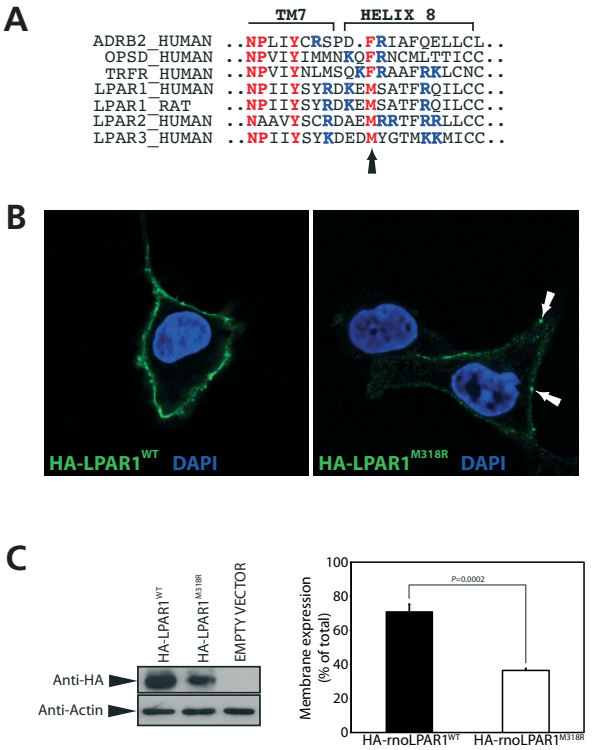

Figure 1: LPAR1^{M318R} results in a decreased pool of cell surface expressed receptors. (**A**) Alignment of amino acid sequence of the 8th helices of different class A GPCRs. Indicated in red is the NPxxY(x)$_{5,6}$F motif of which the tyrosine in TM7 domain and the phenylalanine in helix 8 form a hydrophobic interaction in the inactive state. In the LPAR1-3 the phenylalanine is replaced by a methionine, conserving hydrophobicity. The arrow indicates the location of the amino acid change caused by the mutation. Indicated in blue are the basic amino acids that are commonly found in helix 8 and shown to be involved in receptor phosphorylation by GRKs [21]. Note that the ENU-induced mutation in *Lpar1* causes a change of the methionine into a basic arginine. (**B**) Less membrane expression of LPAR1^{M318R} observed *in vitro*. Intact serum-starved COS-7 cells expressing N-terminally HA-tagged wild type or mutant LPAR1 were incubated with an anti-HA antibody. The antibody can only bind if the cell expresses the receptor in the membrane, because only then is the HA-tag localized outside the cell. Arrow indicates punctuate expression pattern of LPAR1^{M318R} (**C**) The size of the cell surface expressed receptor pool as percentage of the total receptor pool is decreased in cells expressing LPAR1^{M318R}. Right panel shows western blot analysis of the total fraction of COS-7 cells expressing either HA-LPAR1^{WT} or HA-LPAR1^{M318R}. Left panel shows the pool of cell surface expressed receptor as a percentage of total HA-tagged expressed receptors measured by cell surface ELISA.

with the HA-tagged LPAR1^{M318R} as compared to wild type receptor (**Fig. 1B**). Western blot analysis of whole cell lysates revealed a slightly decreased expression level of HA-LPAR1^{M318R} compared to HA-LPAR1^{WT} (**Fig. 1C** *left panel*), which could reflect receptor instability or degradation. Therefore we quantified cell membrane expression using a cell surface ELISA and indeed observed that the pool of cell surface expressed

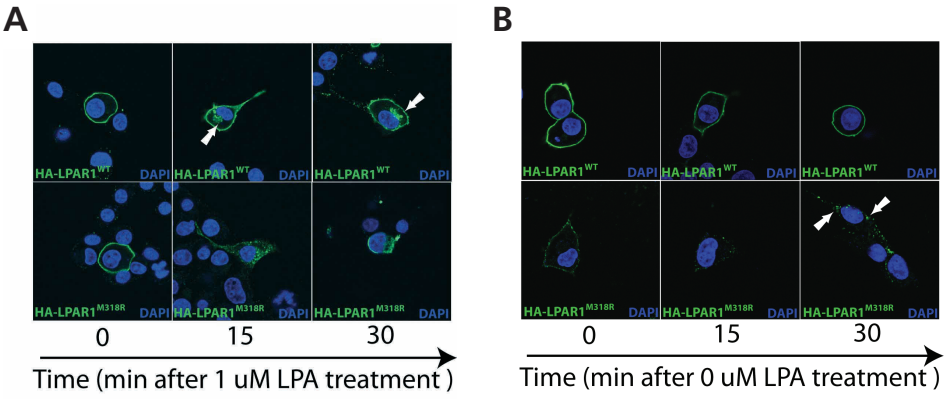

Figure 2: Pulse-labeling assay shows spontaneous internalization of LPAR1^{M318R} *in vitro*. Intact serum-starved COS-7 cells expressing HA-tagged wild type or mutant LPAR1 were incubated with an anti-HA antibody followed by administration of 1 μM LPA (**A**) or medium without LPA (**B**). Subsequently, cells were fixed and permeablized at the indicated time points and stained with a secondary antibody labeled with FITC. After treatment with 1 μM LPA receptor internalization is observed in both cells expressing wild type (*see arrows*) or mutant LPAR1. In contrast only LPAR1^{M318R} internalizes after treatment with medium that does not contain LPA (*see arrows*).

HA-LPAR1^{M318R} as a percentage of the total expressed mutant receptor was significantly decreased compared to HA-LPAR1^{WT} (**Fig. 1C** *right panel*).

Interestingly, cell surface-expressed mutant LPAR1 displayed a punctuate appearance (**Fig. 1B**), which can be indicative for receptor clustering in clathrin-coated pits and internalization [20]. We tested if the difference in membrane expression was the result of aberrant receptor internalization after agonist treatment. Both wild type and mutant receptor internalized after treatment with 1 μM of LPA (**Fig. 2A**). However, mutant LPAR1 also displayed agonist-independent internalization, which was not observed in cells expressing wild type receptor (**Fig. 2B**) and is in line with the punctuate expression pattern of LPAR1^{M318R}.

Importance of helix 8 in agonist-induced receptor phosphorylation

It was recently shown in both the thyrotropin-releasing hormone receptor (TRHR) and ADRB2 that lysine and arginine residues that typically surround the hydrophobic phenylalanine in helix 8 (**Fig. 1A**) play a critical role in agonist-induced receptor phosphorylation by GRKs [21]. It is conceivable that the agonist-independent internalization of LPAR1^{M318R} results from inappropriate receptor phosphorylation, especially because the ENU-mutation not only disrupts the hydrophobic interaction between TM domain 7 and the helix 8 but also introduces an additional arginine. To test this hypothesis we mutated F332 in the helix 8 of the prototypic human ADRB2 using site-directed mutagenesis (ADRB2^{F332R}). First we checked cell surface expression of N-terminally HA-tagged ADRB2^{F332R} in COS-7 cells. Like in LPAR1 the mutation

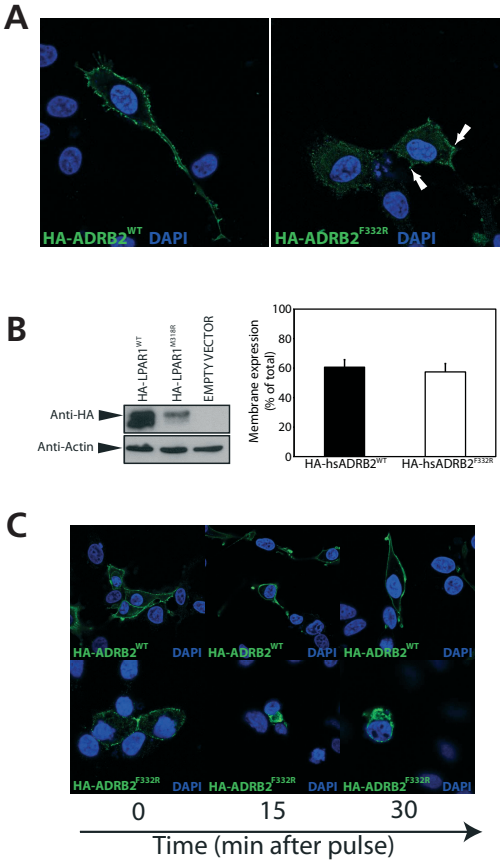

Figure 3: Introduction of the analogous M318R amino acid change in human ADRB2 (F332R) does also result in spontaneous internalization. (**A**) Staining of intact serum-starved COS-7 cells expressing N-terminally HA-tagged wild-type or mutant ADRB2 reveals less cell surface expressed mutant receptor. Arrows indicate punctuate expression pattern of mutant receptor. (**B**) Quantification of observation in (A) using a cell surface ELISA assay. Right panel shows western blot analysis of the total fraction of COS-7 cells expressing either HA-ADRB2^{WT} or HA-ADRB2^{F332R}. Left panel shows the pool of cell surface expressed receptor as a percentage of total HA-tagged expressed receptor. (**C**) Pulse labeling assays demonstrates spontaneous internalization of mutant ADRB2. Intact serum-starved COS-7 cells expressing HA-tagged wild type or mutant ADRB2 were incubated with an anti-HA antibody and subsequently with DMEM. At the indicated time point the cells were fixed, permeablized and stained with a secondary antibody labeled with FITC.

resulted in decreased cell surface expression of ADRB2. However, total HA-ADRB2^{F332R} expression was much lower than HA-ADRB2^{WT} expression as assayed by western blot analysis (**Fig. 3B** *left panel*). In addition, when we quantified cell membrane expression no difference was observed in the size of the pool of cell surface expressed HA-ADRB2^{F332R} as a percentage of the total expressed mutant receptor compared to HA-ADRB2^{WT} (**Fig. 3B** *right panel*). Since the total expressed HA-ADRB2^{F332R} is much lower than wild-type receptor, this could reflect a faster breakdown of the internalized mutant ADRB2 compared to internalized mutant LPAR1 *in vitro*. Notably, cell surface-expressed ADRB2^{F332R} displayed the punctuate appearance as observed in mutated LPAR1 (**Fig. 1B**), which again indicates receptor clustering in clathrin-coated pits. Indeed, ADRB2^{F332R} displayed agonist-independent internalization (**Fig. 3C**) comparable as LPAR1^{M318R}, suggesting a general role of the hydrophobic residue in helix 8 in GPCR internalization.

Next, we determined the effect of the mutation on agonist-induced receptor phosphorylation since this is an important step leading to receptor desensitization

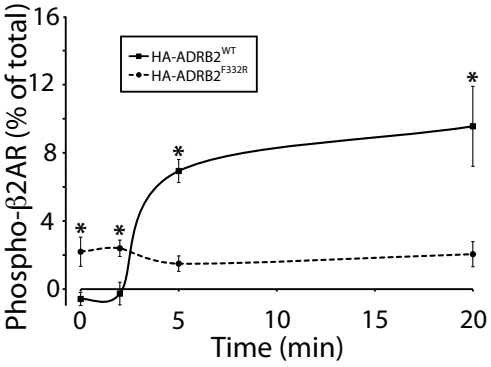

Figure 4: ADRB2^{F332R} shows impaired agonist-induced receptor phosphorylation. Time-response curve of COS-7 cells that transiently express wild type or mutant HA-ADRB2. Cells were treated with 1 μM isoproterenol and fixed at the indicated time points.

and internalization. To quantify this, a phosphoreceptor ELISA assay was performed using an antibody directed against two established GRK sites in the C-terminus of ADRB2 [22]. Treatment of cells expressing HA-ADRB2^{WT} with isoproterenol induced receptor phosphorylation (**Fig. 4**), consistent with previous data [21]. In contrast, mutant ADRB2 failed to show any agonist-induced receptor phosphorylation (**Fig. 4**). Interestingly, we did observe a slight but significantly higher basal phosphorylation level of the mutant receptor when compared to wild-type receptor, which was agonist-independent and suggests a potentially constitutively desensitized receptor state. One explanation for the observation that the level of phosphorylated ADRB2^{F332R} does not increase after agonist treatment as it does in the wild-type situation, is that the mutant receptor already adopts a confirmation suitable for phosphorylation by GRKs without the necessity of agonist. Interestingly, 'only' ~2% of total HA-tagged ADRB2^{F332R} is agonist-independently phosphorylated (**Fig. 4**), while the pool of cell surfaced expressed mutant receptor as percentage of total HA-tagged receptors is equal (**Fig. 3B** *right panel*). This probably means that plasma membrane expression of the mutant receptor is not the major determining factor for the observed level of receptor phosphorylation and it likely reflects the temporality of this desensitization mark, which in case of the mutant receptor is rapidly exchanged for downstream desensitization factors, like recruitment of β-arrestin [3].

Arrestin dependence in spontaneous internalization

It has been demonstrated that LPA-induced internalization of LPAR1 is essentially dependent on at least β-arrestin2 recruitment [23]. To test if β-arrestin2 distribution in the cell is affected by LPAR1^{M318R}, we co-expressed the N-terminally HA-tagged LPAR1 construct with an β-arrestin2-EGFP fusion protein. In wild-type situation, β-arrestin2-EGFP is diffusely located in the cytoplasm (**Fig 5A**). Upon LPA treatment the fusion protein is transiently recruited to the plasma membrane and eventually restores its diffuse expression pattern (**Fig. 5A**), which is consistent with previously reported data [23] and indicates that the β-arrestin2-EGFP fusion protein is functional.

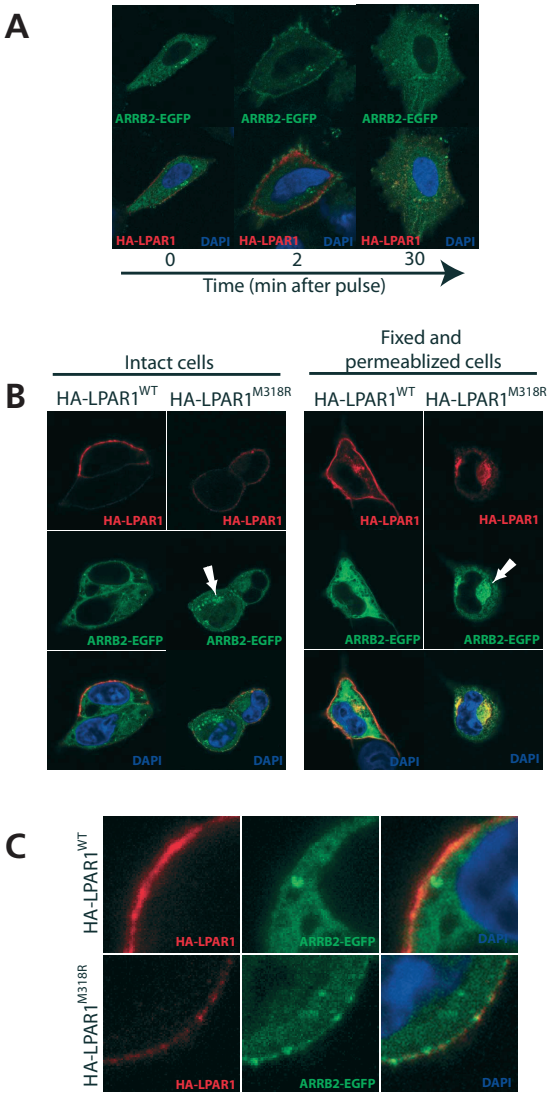

Figure 5: Mutation results in spontaneous recruitment of β-arrestin2 *in vitro*. (**A**) Co-expression of HA-LPAR1^{WT} and β-arrestin2-EGFP shows transiently recruitment of β-arrestin2 to the plasma membrane upon LPA treatment. (**B**) Serum-starved untreated COS-7 cells co-expressing β-arrestin2-EGFP and wild type or mutant HA-LPAR1 were either incubated with an anti-HA antibody while intact (*left panels*) or fixed and permeablized (*right panels*). Cells transfected with mutant HA-LPAR1 show β-arrestin2-EGFP expression in intracellular foci (see *arrows*), which are not observed in cells tranfected with wild-type HA-LPAR1. (**C**) Enlargement of plasma membranes of cells displayed in (B) left panels, clearly shows recruitment of β-arrestin2-EGFP to the cell surface in cells co-transfected with mutant HA-LPAR1, which is not observed in cells co-transfected with wild-type HA-LPAR1.

Co-expressing HA-LPAR1^{M318R} and β-arrestin2-EGFP results in an accumulation of β-arrestin2 in intracellular foci, which is not observed in cells expressing LPAR1^{WT} (**Fig. 5B**). These intracellular foci perfectly co-localized with LPAR1^{M318R} (**Fig. 5B** *right panel*). In addition, an agonist-independent recruitment of β-arrestin2 to the cell surface in cells expressing LPAR1^{M318R} but not in cells expressing LPAR1^{WT} was observed (**Fig. 5C**). Specifically, cell surface recruited β-arrestin2 co-localizes with the punctuate LPAR1^{M318R} expression foci (**Fig. 5C**). This suggests constitutive recruitment of β-arrestin2 to both cell surface and intracellular expressed mutant LPAR1 even in the absence of agonist.

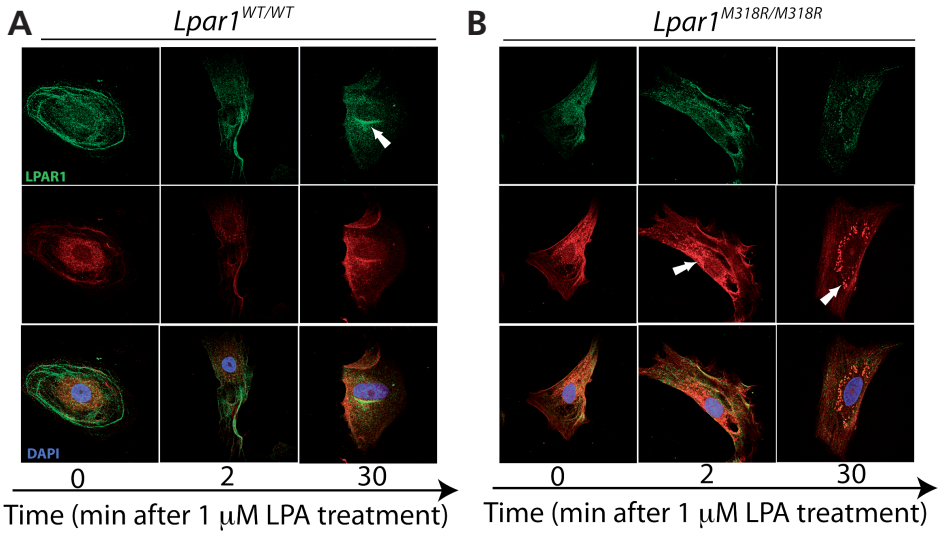

Figure 6: Accumulation of endogenous β-arrestin2 is observed in *Lpar1*$^{M318R/M318R}$ REFs. (**A**) Serum-starved wild-type REFs show transient recruitment of endogenous β-arrestin2 to the cell surface 2 minutes after LPA treatment and accumulation of internalized endogenous LPAR1 30 minutes after LPA treatment (*see arrow*). (**B**) Serum-starved untreated homozygous mutant REFs display a more diffuse LPAR1 expression pattern compared to wild-type REFs. Upon treatment endogenous β-arrestin2 accumulation in intracellular foci is apparent (*see arrows*) that is not observed in wild-type REFs, suggesting prolonged recruitment by mutant LPAR1.

Next, we isolated primary REFs from the LPAR1^{M318R} strain, which endogenously express *Lpar1* (**Fig. S1**). Serum-starved REFs were stained for endogenous β-arrestin2 and LPAR1 different at time points after LPA treatment (**Fig. 6**). Since it is known that it is difficult to raise specific antibodies against GPCRs, the anti-LPAR1 antibody was tested using HA-LPAR1 overexpression (**Fig. S2**). The expression of endogenous LPAR1 in homozygous mutant REFs is more diffuse throughout the cell compared to the more compartmentalized LPAR1 localization in wild-type REFS (**Fig. 6**). Furthermore, there seems to be less LPAR1 in homozygous mutant REFs compared to wild-type REFs, which could reflect increased breakdown of endogenous receptor, as was also observed *in vitro* (**Fig. 1C** *left panel*). Upon LPA treatment, β-arrestin2 is transiently recruited of to the cell surface, followed by diffuse localization of β-arrestin2 and accumulation of intracellular LPAR1 in wild-type REFs (**Fig 6A**). *Lpar1*$^{M318R/M318R}$ REFs showed an accumulation of β-arrestin2 in foci near the nucleus upon LPA treatment, which also appears to co-localize with LPAR1 and persist up to 30 minutes after the treatment (**Fig. 6B**). However, accumulation of β-arrestin2 was already observed without the addition of agonist in cells overexpressing LPAR1^{M318R} (**Fig. 5**), which is more difficult to see in REFs probably because of low expression of endogenous β-arrestin2 and LPAR1. Nevertheless, these observations support the hypothesis that M318R may result in

constitutive recruitment of β-arrestin2 to LPAR1 and rules out nonspecific effects in the *in vitro* studies because of overexpression of LPAR1 and β-arrestin2.

Impaired LPA-induced ERK1/2 activation in REFs

One of the pathways activated by LPA is the mitogen-activated protein (MAP) kinase cascade via pertussis toxin (PTX)-sensitive G_i in a tyrosine kinase-dependent manner [24]. This pathway is thought to be the main mediator of the stimulatory effect of LPA on cell proliferation. Interestingly, it has been shown that LPA can also stimulate the MAP kinase pathway via a G protein-independent manner by β-arrestin2 recruitment [25], which can serve as a scaffold for cytoplasmic signaling complexes [4]. To test the effect of LPAR1^{M318R} on the LPA-induced MAP kinase-signaling cascade, serum-starved homozygous mutant and wild-type primary REFs were stimulated with LPA and ERK1/2 phosphorylation was measured. LPA treatment results in a rapid and transient ERK1/2 phosphorylation response (**Fig. 7**), consistent with previous findings [25]. In contrast, homozygous mutant REFs show an attenuated LPA-induced ERK1/2 phosphorylation response (**Fig. 7**), demonstrating that M318R in LPAR1 diminishes LPA signaling in primary REFs. Possibly, the increase of spontaneous internalization via β-arrestin2 in LPAR1^{M318R}, results in less cell surface expressed receptor and a hypomorphic response to LPA treatment. This would be in line with the loss-of-function phenotype that we observed *in vivo* in homozygous mutant rats [18]. However, if the phenotype is caused by agonist-independent recruitment of β-arrestin2, which can stimulate the MAP kinase signaling cascade independent of G protein activation, we would expect increased β-arrestin2-dependent ERK1/2 phosphorylation in the homozygous mutant REFs. To test this hypothesis, REFs pretreated with pertussis toxin (PTX) and the experiment was repeated. PTX completely abolishes LPA-induced

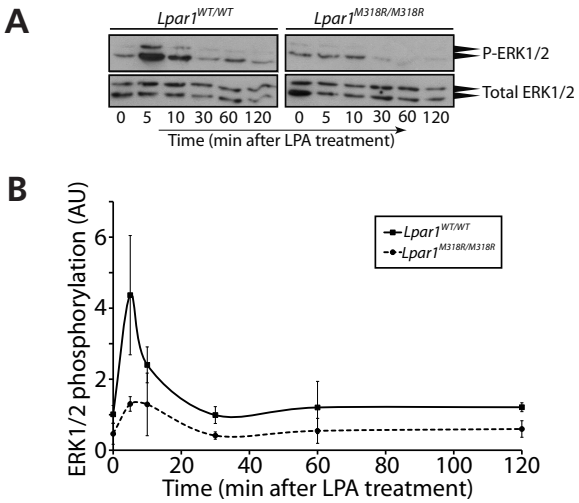

Figure 7: *Lpar1*$^{M318R/M318R}$ REFs display a hypomorhic ERK1/2 phosphorylation pattern. (**A**) Examples of the stimulation of the MAP kinase pathway of wild type and mutant REFs after 1μM LPA treatment. (**B**) Quantification of ERK1/2 phosphorylation response of wild type and homozygous mutant REFs after 1 μM LPA treatment. Data shown are independent experiments of REFs isolated from two different embryos per genotype.

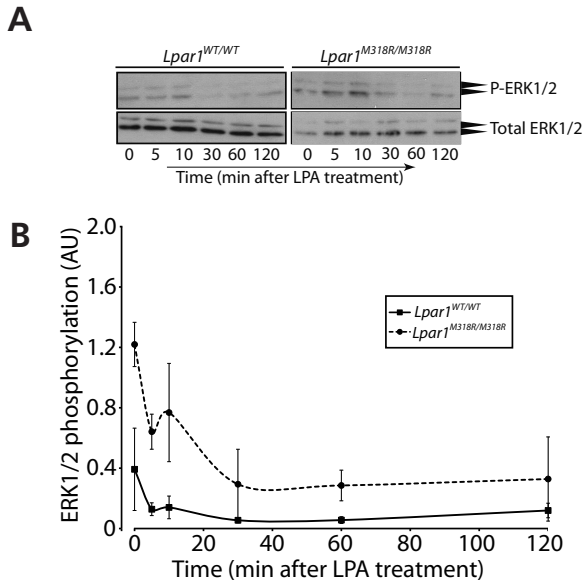

Figure 8: *Lpar1*$^{M318R/M318R}$ REFs show an increased G_i and agonist independent basal ERK1/2 phosphorylation level. REFs were pretreated with PTX before LPA treatment and thereby blocking G_i protein and agonist-induced ERK1/2 phosphorylation. (**A**) Examples of the stimulation of the MAP kinase pathway of wild type and mutant REFs after 1μM LPA treatment. (**B**) Quantification of ERK1/2 phosphorylation response of wild type and homozygous mutant REFs after 1 μM LPA treatment. Data shown are independent experiments of REFs isolated from two different embryos per genotype.

ERK1/2 phosphorylation in wild-type serum-starved REFs (**Fig. 8**). Clearly, a higher basal level of G protein-independent ERK1/2 phosphorylation is observed in homozygous mutant REFs, which is independent of agonist treatment (**Fig. 8**). This probably reflects β-arrestin-dependent MAP kinase activation and further supports the idea of constitutive recruitment of β-arrestin2 as a result of the mutation in LPAR1.

DISCUSSION

Point mutations that affect functionally important gene products, like GPCRs, are often associated with human disease. The underlying molecular mechanisms causing the phenotype, however, are often complex and not properly appreciated. Here we show that an ENU-induced missense mutation in helix 8 of LPAR1 that disrupts a highly conserved hydrophobic interaction between the TM7 domain and helix 8, results in a loss-of-function phenotype *in vivo* [18], but seems to do so by β-arrestin-mediated constitutive desensitization. This causes impaired LPA signaling as shown in primary REFs, probably because of the decrease of cell surface expressed receptor and consequently increased breakdown or impaired recycling. In line with these observations, homozygous mutant animals show an apparent loss-of-function phenotype, which was characterized by a craniofacial disorder and smaller size [18]. Although the LPA responsiveness of primary *Lpar1*$^{M318R/M318R}$ REFs was severely impaired, a small level of LPA-induced ERK1/2 phosphorylation was still observed. This can be the result of either functional redundancy with other LPA receptors expressed in fibroblast (**Fig. S1**) or because the pool of cell surface expressed LPAR1 is much smaller

at the time LPA is administrated when compared to wild-type fibroblasts. Importantly, we observed an increased basal level of ERK1/2 phosphorylation independent of LPA treatment and G-protein activation. Although this increase may seem trivial, it could have profound effects in developmental, physiological and disease processes *in vivo* given the importance of ERK1/2 phosphorylation in a variety of cellular processes. In addition, it has been shown that activation of the G protein pathway can result in contradictory cellular effects compared to activation of the β-arrestin pathway [26]. Specifically, it has been shown that β-arrestin signaling can inhibit GPCR-mediated apoptosis [27], which could possibly explain the lack of postnatal lethality in the *Lpar1*$^{M318R/M318R}$ rat [18] that was observed in LPAR1 knockout mice [11].

Previously, mutations that change the arginine in the highly conserved DRY motif have been shown to result in constitutively desensitized GPCRs [20]. Importantly, a naturally occurring loss-of-function mutation in the DRY motif of the human vasopressin type II receptor, which is associated with nephrogenic diabetis insipidus results in a constitutive arrestin-mediated desensitization [28], underlining the importance of understanding this mechanism in GPCR-related pathophysiology. Here we demonstrate that interruption of the hydrophobic interaction between the NPxxY motif and helix 8, which is also highly conserved in class A GPCRs may result in constitutive desensitization. Interestingly, both motifs are suggested to play significant roles in G protein activation [29], indicating functionally overlapping binding sites of β-arrestin and the G protein.

Specific features arguing for constitutive desensitization as a result of disrupting the hydrophobic interaction between the NPxxY motif and helix 8 were not only observed in rat LPAR1^{M318R}, but also in the prototypical human ADRB2^{F332R}. It has been stated that a constitutive desensitized receptor should demonstrate four characteristics, namely inappropriate phosphorylation, abnormal intracellular localization, a greater uncoupling from G protein compared to wild-type receptor and abnormal association with β-arrestins [30]. The mutation described here displays all these features as jointly demonstrated in rat LPAR1 and human ADRB2. Notably, since it was recently shown that basic arginine and lysine residues in helix 8 are important for agonist-induced GPCR phosphorylation by GRKs and consequently recruitment of β-arrestin [21], it is possible that the molecular and functional effects observed in the receptor mutants is a combination of inappropriate proper folding of helix 8 and increased receptor phosphorylation by GRKs.

The *Lpar1*$^{M318R/M318R}$ rat is a unique model to study the *in vivo* loss-of-function of the receptor in this model organism. Especially the involvement of LPAR1 in human psychiatric disorders could potentially be well studied in this genetically modified rat model. Although the mutation does not result in a complete loss of receptor function, like in the case of knockout animals, amino acid changes associated with human disease are also unlikely to result in complete loss-of-function. Single nucleotide polymorphisms are the most common form of naturally occurring genetic variation

in human [31] and when located in coding regions they are most likely to result in missense mutations. Since the mutation described here may associate helix 8 with constitutive desensitization, which was to date only described to be associated with mutations found in the DRY motif, it is likely more human disease related mutations in genes encoding GPCRs will be found to result in constitutive desensitization. Importantly, the LPAR1^{M318R} rat represents an *in vivo* model for a GPCR that may be constitutively desensitized mediated by β-arrestin. In the growing understanding of the complexity of GPCR signaling, including biased agonism [5] and β-arrestin-mediated signal transduction [4], this mutant rat model helps to unravel the consequences of GPCR constitutive arrestin-mediated desensitization *in vivo*.

MATERIALS AND METHODS

Animals and primary REF isolation. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Sciences according to the Dutch legal ethical guidelines. Experiments were designed to minimize the number of required animals and their suffering. Heterozygous carriers were mated and at E13.5 embryos were isolated. After washing the embryo thoroughly, the head and visceral organs were removed and used for DNA isolation and genotyping. The embryos were minced and treated with trypsin to get a single cell suspension. Rat embryonic fibroblasts were grown in DMEM supplemented with 10% FCS.

***In vitro* fusion protein expression studies.** Wild type and mutant receptors were N-terminally HA-tagged by cloning into the expression vector pcDNA3.1 (Invitrogen). β-arrestin2 was fused with EGFP by cloning its cDNA sequence isolated from brain into pEGFP-N1 (Invitrogen). The receptor fusion proteins were expressed in COS-7 cells, which were seeded on a coverslip, and 24 hours after transfection the cells were placed on ice and incubated with DMEM-buffered HEPES containing 0.2% fatty acid-free bovine serum albumin (DHB) for 15 minutes. Subsequently, the cells were incubated for 1 hour with a polyclonal rabbit anti-HA (Abcam Inc, Cambrigde) on ice in DHB at a 1:500 dilution. For cell surface expression analysis the cells were immediately methanol-fixed and washed thoroughly with PBS and incubated for one hour in blocking buffer (1% BSA in 0.1% PBS-Tween) at room temperature. For pulse labeling assays, the cells were washed with PBS followed by incubation with or without agonist in DMEM at 37°C, 5% CO_2. At the indicated time points the cells were methanol-fixed and washed thoroughly with PBS and incubated for one hour in blocking buffer (1% BSA in 0.1% PBS-Tween) at room temperature. The cells were washed three times with PBS and incubated for 1 hour with a secondary anti-rabbit antibody conjugated with FITC (Abcam Inc, Cambrigde) at room temperature in the dark. After three times washing with PBS the coverslips were mounted using Vectashield with DAPI (Brunschwig Chemie, Amsterdam) and analyzed using confocal microscopy. For western blotting COS-7 cells were lysed 24 hours after transfection and the proteins were separated on a SDS gel (10% acrylamide gradient, Bio-Rad) and transferred to a nitrocellulose membrane. The membrane was incubated for 1 hour at room temperature with a 1:5,000 or 1:1,000 dilution of respectively a polyclonal rabbit anti-HA antibody (Abcam Inc, Cambrigde) or a polyclonal rabbit anti-actin antibody (Sigma Aldrich) in blocking buffer followed by an incubation for 1 hours with peroxidase-conjugated, anti-rabbit IgG diluted 1:5,000 in blocking buffer at room temperature. Protein bands were detected by using the enhanced chemiluminescence detection method (ECL, Amersham Biosciences).

Cell surface ELISA. Surface expression of N-terminally tagged receptors was quantified as described [32]. Briefly, the N-terminally tagged receptors were transiently expressed in COS-7 cells and 24 hours after transfection the cells harvested, seeded in 12 well plates and serum-starved overnight. To specifically measure the pool of cell surface expressed receptors, the cells were placed on ice and incubated with DMEM-buffered HEPES containing 0.2% fatty acid-free

bovine serum albumin (DHB) for 15 minutes. Subsequently, the cells were incubated for 1 hour with a polyclonal rabbit anti-HA (Abcam Inc, Cambrigde) on ice in DHB at a 1:500 dilution, rinsed with PBS on ice, fixed with ice-cold methanol:aceton (1:1) and air dried. Cells were then washed 5 minutes with PBS, blocked with RIPA/milk buffer (150 mM NaCl, 50 mM Tris, 1 mM EDTA, 10 mM NaF, 100nM sodium orthovanadate, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, pH 8.0, and 5% nonfat dried milk). To measure the total pool of N-terminally tagged receptors, cells were first fixed with ice-cold methanol:aceton (1:1), air dried and then washed 5 minutes with PBS. Subsequently, the cells were blocked with RIPA/milk buffer (150 mM NaCl, 50 mM Tris, 1 mM EDTA, 10 mM NaF, 100nM sodium orthovanadate, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, pH 8.0, and 5% nonfat dried milk) and incubated with 1:500 anti-HA antibody (Abcam Inc, Cambrigde). All cells were then washed 3 times 5 minutes with PBS and incubated with 1:5,000 peroxidase-conjugated, anti-rabbit IgG diluted 1:5,000 in RIPA/milk buffer. The cells were washed 3 times 5 minutes with PBS and incubated with ELISA TMB reagent (Sigma-Aldrich, St. Louis, MO). The reaction was terminated with stop reagent for TMB substrate (Sigma-Aldrich, St. Louis, MO) and the absorbance was measured at 450 nm. Data was presented as the A_{450} of the pool of cell surface expressed receptors as a percentage of the A_{450} of the total pool of HA-tagged receptors.

Phosphorecepor ELISA. The measurement of agonist-induced ADRB2 phosphorylation was quantified as described [32,33]. The GRK-dependent phosphorylation of ADRB2 was measured using a rabbit antibody against the pSer355 and pSer356 in the human receptor (Santa Cruz Biotechnology, CA) at 1:500 using the cell surface ELISA protocol. The A_{450} of phosphorylated ADRB2 was presented as a percentage of the A_{450} of cell stained for the total pool of HA-ADRB2.

Staining of endogenous LPAR1 and β-arrestin in REFs. REFs were seeded on glass coverslip and allowed to attach overnight. Subsequently, cells were serum-starved overnight in DMEM supplemented with 0.2% fatty acid-free BSA, followed by treatment with 1 μM LPA. At the indicated time points the cells were fixed with ice-cold methanol and dried. After washing with PBS the cells were incubated with blocking buffer (1% BSA and 0.1% Tween in PBS) for 1 hour at room temperature. Cells were incubated with an antibody raised in rabbit against human LPAR1 (Abcam Inc, Cambrigde) and an antibody raised in mouse against human β-arrestin2 (Santa Cruz Biotechnology, CA) both at 1:200 dilution in blocking buffer overnight at 4°C. Cells were washed 3 times 10 minutes with PBS and incubated with anti-rabbit conjugated with FITC and anti-mouse conjugated with Alexa-546 both at 1:200 dilution in blocking buffer for 1 hour at room temperature. After three times washing with PBS the coverslips were mounted using Vectashield with DAPI (Brunschwig Chemie, Amsterdam) and analyzed using confocal microscopy.

ERK1/2 phosphorylation analysis. Primary REFs were seeded in 12-wells plates and allowed to attach overnight. Subsequently, the cells were serum-starved overnight in DMEM supplemented with 0.2% fatty acid-free BSA, and with or without the presence of 100 ng/ml pertussis toxin (PTX). Cells were treated with 1 μM LPA and lysed with RIPA buffer supplemented with complete protease inhibitor cocktail (Roche Diagnostics) and PhosSTOP phosphatase Inhibitor Cocktail (Roche Diagnostics) at the indicated time points. Western blot analysis was performed as described above with a mouse antibody against pT183 and pY185 in human ERK1 and ERK2 (Abcam Inc, Cambrigde) and a rabbit antibody against total rat ERK1 and ERK2 (Cell Signaling Technology, MA).

Statistical analysis. Data are shown as average ± standard deviation of at least triplicate measurements unless indicated different. Where not visible, error bars fell within the symbol size. Data was statistically analyzed, when appropriate, with two-way analysis of variance and Student's unpaired *t* test.

REFERENCES

1. Ma P, Zemmel R (2002) Value of novelty? Nat Rev Drug Discov 1: 571-572.
2. Pierce KL, Premont RT, Lefkowitz RJ (2002) Seven-transmembrane receptors. Nat Rev Mol Cell Biol 3: 639-650.
3. Reiter E, Lefkowitz RJ (2006) GRKs and beta-arrestins: roles in receptor silencing, trafficking and signaling. Trends Endocrinol Metab 17: 159-165.
4. Lefkowitz RJ, Shenoy SK (2005) Transduction of receptor signals by beta-arrestins. Science 308: 512-517.
5. Violin JD, Lefkowitz RJ (2007) Beta-arrestin-biased ligands at seven-transmembrane receptors. Trends Pharmacol Sci 28: 416-422.
6. Moolenaar WH, van Meeteren LA, Giepmans BN (2004) The ins and outs of lysophosphatidic acid signaling. Bioessays 26: 870-881.
7. Choi JW, Herr DR, Noguchi K, Yung YC, Lee CW, et al. (2010) LPA receptors: subtypes and biological actions. Annu Rev Pharmacol Toxicol 50: 157-186.
8. Mills GB, Moolenaar WH (2003) The emerging role of lysophosphatidic acid in cancer. Nat Rev Cancer 3: 582-591.
9. Liu S, Umezu-Goto M, Murph M, Lu Y, Liu W, et al. (2009) Expression of autotaxin and lysophosphatidic acid receptors increases mammary tumorigenesis, invasion, and metastases. Cancer Cell 15: 539-550.
10. Kingsbury MA, Rehen SK, Contos JJ, Higgins CM, Chun J (2003) Non-proliferative effects of lysophosphatidic acid enhance cortical growth and folding. Nat Neurosci 6: 1292-1299.
11. Contos JJ, Fukushima N, Weiner JA, Kaushal D, Chun J (2000) Requirement for the lpA1 lysophosphatidic acid receptor gene in normal suckling behavior. Proc Natl Acad Sci U S A 97: 13384-13389.
12. Harrison SM, Reavill C, Brown G, Brown JT, Cluderay JE, et al. (2003) LPA1 receptor-deficient mice have phenotypic changes observed in psychiatric disease. Mol Cell Neurosci 24: 1170-1179.
13. Jacob HJ (1999) Functional genomics and rat models. Genome Res 9: 1013-1016.
14. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
15. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.
16. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
17. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
18. van Boxtel R, Vroling B, Toonen P, Nijman IJ, van Roekel H, et al. (2010) Systematic generation of in vivo G protein-coupled receptors mutants in the rat. Pharmacogenomics J in press.
19. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, et al. (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem 283: 17753-17756.
20. Wilbanks AM, Laporte SA, Bohn LM, Barak LS, Caron MG (2002) Apparent loss-of-function mutant GPCRs revealed as constitutively desensitized receptors. Biochemistry 41: 11981-11989.
21. Gehret AU, Jones BW, Tran PN, Cook LB, Greuber EK, et al. (2009) Role of helix 8 of the thyrotropin-releasing hormone receptor in phosphorylation by G protein-coupled receptor kinase. Mol Pharmacol 77: 288-297.
22. Seibold A, Williams B, Huang ZF, Friedman J, Moore RH, et al. (2000) Localization of the sites mediating desensitization of the beta(2)-adrenergic receptor by the GRK pathway. Mol Pharmacol 58: 1162-1173.
23. Urs NM, Jones KT, Salo PD, Severin JE, Trejo J, et al. (2005) A requirement for membrane cholesterol in the beta-arrestin- and clathrin-dependent endocytosis of LPA1 lysophosphatidic acid receptors. J Cell Sci 118: 5291-5304.
24. van Corven EJ, Hordijk PL, Medema RH, Bos JL, Moolenaar WH (1993) Pertussis

toxin-sensitive activation of p21ras by G protein-coupled receptor agonists in fibroblasts. Proc Natl Acad Sci U S A 90: 1257-1261.
25. Gesty-Palmer D, El Shewy H, Kohout TA, Luttrell LM (2005) beta-Arrestin 2 expression determines the transcriptional response to lysophosphatidic acid stimulation in murine embryo fibroblasts. J Biol Chem 280: 32157-32167.
26. Kenakin T, Miller LJ (2010) Seven Transmembrane Receptors as Shapeshifting Proteins: The Impact of Allosteric Modulation and Functional Selectivity on New Drug Discovery. Pharmacol Rev.
27. Revankar CM, Vines CM, Cimino DF, Prossnitz ER (2004) Arrestins block G protein-coupled receptor-mediated apoptosis. J Biol Chem 279: 24578-24584.
28. Barak LS, Oakley RH, Laporte SA, Caron MG (2001) Constitutive arrestin-mediated desensitization of a human vasopressin receptor mutant associated with nephrogenic diabetes insipidus. Proc Natl Acad Sci U S A 98: 93-98.
29. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, et al. (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature 455: 497-502.
30. Barak LS, Wilbanks AM, Caron MG (2003) Constitutive desensitization: a new paradigm for g protein-coupled receptor regulation. Assay Drug Dev Technol 1: 339-346.
31. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-236.
32. Gehret AU, Jones BW, Tran PN, Cook LB, Greuber EK, et al. (2010) Role of helix 8 of the thyrotropin-releasing hormone receptor in phosphorylation by G protein-coupled receptor kinase. Mol Pharmacol 77: 288-297.
33. Jones BW, Song GJ, Greuber EK, Hinkle PM (2007) Phosphorylation of the endogenous thyrotropin-releasing hormone receptor in pituitary GH3 cells and pituitary tissue revealed by phosphosite-specific antibodies. J Biol Chem 282: 12893-12906.

SUPPLEMENTARY DATA

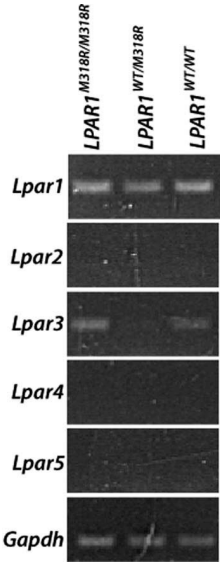

Figure S1: RT-PCR analysis of LPA receptors in primary REFs. RT-PCR was performed using primers for the indicated transcripts of total RNA isolated from wild-type, heterozygous and homozygous mutant REFs.

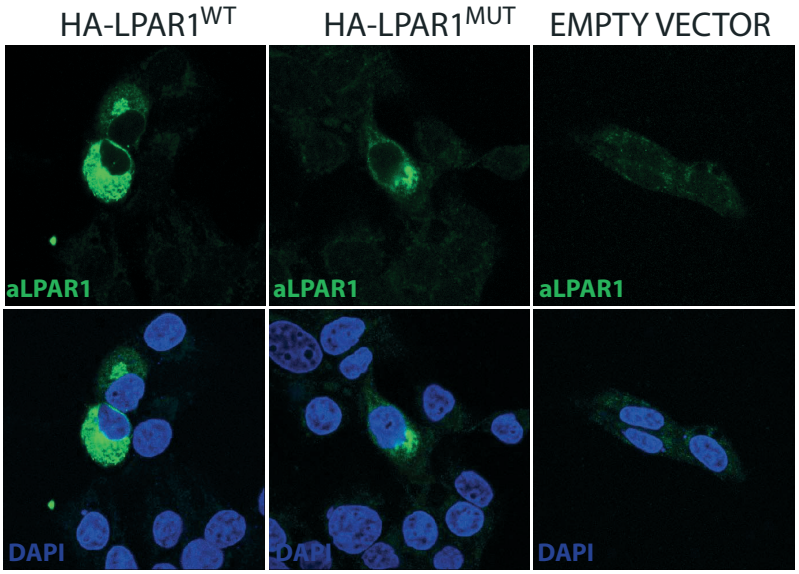

Figure S2: Anti-LPAR1 antibody test. An antibody raised against human LPAR1 was tested using COS-7 cells expressing N-terminally HA-tagged wild-type or mutant LPAR1.

GENERAL DISCUSSION: FUTURE PERSPECTIVES ON ENU TARGET-SELECTED MUTAGENESIS IN THE RAT

Ruben van Boxtel

Hubrecht Institute for Developmental Biology and Stem Cell Research, Cancer Genomics Center, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

PIONEERING RAT REVERSE GENETICS

The laboratory rat is gaining momentum as a mammalian genetic model organism. Traditionally, the rat model has been used extensively to study human physiology and complex disease, strengthening the desire to manipulate its genome [1]. Compared to the mouse model with its seemingly endless possibilities in gene manipulation and standardized mutant phenotyping protocols [2,3], reverse genetics in the rat is still in its early infancy. Nevertheless, in the last decade significant technological breakthroughs, which are reviewed in **Chapter 1**, have enabled the generation of genetically modified rats in an efficient and systematic manner. Already, large collections of knockout animals have been generated and are available for biomedical research (www.knockoutrat.org; http://rgd.mcw.edu). Thus, the question is not *if* rat reverse genetics will be widely implemented in biomedical research, but rather *when* this will happen. To date, only several laboratories worldwide are capable of generating mutant rat models. However, considering the rapid technological improvements in the field of reverse genetics, general application to the generation of genetic mutants in a variety of model organisms, including rats, is to be expected.

ENU target-selected mutagenesis was the first technique used to generate rat knockout models [4,5]. The strength of the approach lies in its simplicity: point mutations are randomly introduced in the male germ line by treating animals with ENU [6] without the need of complicated cell culturing techniques or transgenic animals carrying transposons in their genome or expressing transposases. This technique, like every other reverse genetics approach, has many advantages and disadvantages. Currently, there are three highly complementary approaches successfully applied in the rat, ENU target-selected mutagenesis [4], transposon-tagged mutagenesis [7,8] and zinc-finger nuclease (ZFN)-mediated gene targeting [9] (discussed in **Chapter 1**). In this chapter, the future possibilities of ENU target-selected mutagenesis and its affiliation with other gene manipulation approaches (**Table 1**) applied in rat reverse genetics will be discussed.

SINGLE GENE KNOCKOUT VERSUS LARGE-SCALE MUTAGENESIS

Targeted versus random mutagenesis

Although ENU target-selected mutagenesis has been highly efficient in generating rat mutants, alternative methods have been and are still being developed (**Chapter 1**). A disadvantage of ENU mutagenesis as a reverse genetic tool is that it cannot be targeted to a genetic locus of choice and mutation type cannot be predetermined. If a researcher is interested in one particular gene, random ENU mutagenesis is probably not the preferred method for knocking it out. Even though ENU target-selected mutagenesis is a relatively easy technique, investigators must keep large animal repositories or archives and make large investments to setup high-throughput resequencing methods to identify a mutant allele. Even when these facilities are available, it remains a matter

Table 1: Comparison of available rat mutagenesis techniques

Technique	Targeted or random	Advantages	Disadvantages
ENU mutagenesis target-selected mutagenesis	Random	High mutation efficiency Easily scalable Allows for allelic series	Mutation discovery is relatively laborious Background mutations
Transposon-tagged mutagenesis	Random	Gene insertions easily detectable by reporter gene cassettes Integration site easy to identify	Relatively low mutation efficiency Biased genomic integration pattern
ZFN-mediated gene targeting	Targeted	Allows for gene targeting by NHEJ and theoretically allows for HR High efficiency introducing DSB	Modular assembly of ZF arrays is relatively unsuccessful Commercial ZFNs are expensive

ENU, *N*-ethyl-*N*-nitrosourea; ZFN, zinc-finger nuclease; NHEJ, nonhomologous end joining; HR, homologous recombination; ZF, zinc-finger; DSB, double stranded break.

of chance to identify a knockout allele of that particular gene, which depends on the size of the gene, the ENU-induced mutation frequency and the number of screened F_1 animals. In this case, the use of a targeted approach, like embryonic stem (ES) cell-based homologous recombination (HR)[2] or ZFN-mediated gene targeting [9], can be preferred. In theory, these methods would also allow for the generation of conditional and targeted knock-in alleles, although this has not yet been shown for the rat (**Chapter 1**). Nevertheless, the ability to generate conditional and targeted knock-in alleles has proven to be extremely powerful in mice, making targeted mutagenesis an indispensible genetic tool to model human disease. However, a disadvantage of these targeted approaches is that designing, generating and testing constructs for HR or encoding specific ZFNs is a relatively laborious and time-consuming process for generating a single mutant allele, which is much less amenable for scaling (**Chapter 1**). In addition, a high number of fertilized oocytes have to be injected and relatively many animals have to be generated to isolate knockout alleles for one single gene [9]. Therefore, for large-scale studies, like the generation of knockout alleles for virtually every gene in the rat genome, random mutagenesis, like ENU mutagenesis or transposon-mediated mutagenesis [7,8], is preferred over targeted mutagenesis. These techniques are generally highly efficient in generating large collections of mutant alleles in a limited number of screens. An additional advantage of random mutagenesis is that it can be applied in both phenotype-driven screens as well as genotype-driven screens (**Chapter 1, Chapter 2**), allowing researchers to not only determine specific gene function, but also to investigate the molecular pathway underlying disease phenotypes. Obviously, different technical challenges arise in large-scale mutagenesis

studies compared with targeted mutagenesis approaches, like mutation discovery and mutant archiving, which will be discussed below.

Transposon-mediated versus ENU mutagenesis

Not long after the first ENU-induced rat knockout models were reported, a second random mutagenesis approach was established in the rat [7,8], namely transposon-tagged mutagenesis (**Chapter 1**). Clearly, both approaches are highly complementary with specific advantages and disadvantages (**Chapter 1**). For example, the strength of transposon-tagged mutagenesis is that the transposon can be equipped with reporter cassettes and that the site of insertion can easily be determined [10]. Furthermore, the current transposon-tagged mutagenesis approaches in rats were all based on chromosomal transposition, which are heavily biased by integration site preference [10]. Although this can be beneficial for specifically investigating quantitative trait loci (**Chapter 1**), it makes this technique unsuitable for genome-wide coverage. Nevertheless, these problems can potentially be overcome by applying a 'plasmid-to-genome' delivery approach in cultured spermatogonial stem cells (SSCs) (**Chapter 1**). Importantly, the efficiency of mutagenesis (~ 11% of all F_1 animals show reporter gene expression when Sleeping Beauty transposon is used [7]) is relatively poor compared with ENU mutagenesis, which is a highly efficient germ line mutator. For example, using a DNA mismatch repair deficient background, we have shown that mutation rates of at least 1 mutation every 700,000 bp can be repeatedly reached (**Chapter 4, Chapter 5 and Chapter 6**). If the complete rat transcriptome is 35 Mb, every F_1 animal will carry 50 mutations in the coding DNA, of which 74% will likely be missense and 5% nonsense (**Chapter 4**). This means that every F_1 animal will carry an average of 2–3 nonsense mutations and 37 missense mutations. The major technical challenge, however, is to identify and isolate ENU-induced mutations that affect protein function in genes-of-interest. Because ENU-induced mutation frequencies are high, unwanted background mutations can form phenotypic complications when a mutation of interest is crossed to homozygosity. Nevertheless, as discussed throughout this thesis, these problems can easily be overcome by outcrossing the mutation of interest multiple times to the parental strain. A unique property of ENU mutagenesis is that it generates point mutations in the genome, thereby mimicking the most common form of human genetic variation [11] and more importantly allowing for the generation of allelic series.

Generating allelic series

Allelic series are collections of multiple mutations in the same gene, which are independently identified in different animals. Mutations can result in knockout alleles by insertion of a transposon in the open reading frame (ORF), out-of-frame mutations by error-prone repair of a ZFN-induced double stranded break (DSB) or by introducing an ENU-induced premature stop codon in the ORF. In addition, ENU mutations can cause amino acid changes of functionally important residues, which subsequently can

result in hypo-, hyper- or neomorphic alleles. This can be highly relevant for studying gene function and is nicely exemplified by the collection of ENU-induced G protein-coupled receptors (GPCRs) mutants described in **Chapter 5**. Mutations causing amino acid changes of residues important for correct folding or incorporation into the plasma membrane can result in a more unstable protein and likely represent a hypomorphic allele. These mutant alleles are not only useful for studying gene-dosage effects, but also are important for studying phenotypic effects of hypomorphic alleles of genes that are essential for viability. For example, knockout alleles for the GPCR *Smoothened* (SMO), which is essential for Sonic Hedgehog signaling, result in embryonic lethality [12]. It can be speculated that the phenotype induced by a hypomorphic allele is less severe, allowing study of the function of SMO beyond the process of embryogenesis. In our GPCR mutant collection, a mutation was identified in SMO that was predicted to result in a less stable confirmation because of the incorporation of a proline residue into a transmembrane (TM) domain, which can disrupt the structure considerably (**Chapter 5**). However, it still has to be determined if this mutation results in a viable but partial loss-of-function phenotype *in vivo* by crossing it to homozygosity.

In contrast, ENU-induced mutations resulting in amino acid changes of residues important for maintaining an inactive conformation in the absence of agonist are likely to result in increased constitutive activity and are likely to represent hypermorphic alleles. Notably, constitutively activating mutants (CAMs) have been widely used to study GPCR activation, desensitization and internalization *in vitro*, and importantly, naturally occurring CAMs have been associated with a variety of human diseases [13]. Indeed, CAMs of the luteinizing hormone/choriogonadotropin receptor (LHCGR) have been shown to cause male precocious puberty [14] and Leydig cell tumors [15]. Interestingly, our collection of ENU-induced mutant GPCRs contained a mutant allele of LHCGR, which was predicted by computational analyses to result in an increased constitutive receptor activity, indicating the feasibility of generating mutant alleles that model specific human disease (**Chapter 5**).

Alternatively, the mutation that was identified in LPAR1 resulted in an apparent loss-of-function phenotype *in vivo* (**Chapter 5**) and showed a hypomorphic response upon agonist treatment (**Chapter 7**). Nevertheless, in-depth molecular and functional analyses showed constitutive recruitment of β-arrestin and consequently spontaneous internalization and increased basal mitogen-activating protein (MAP) kinase ERK1/2 signaling, which was agonist independent (**Chapter 7**). This latter observation indicates that the overall hypomorphic response of the mutant LPAR1 is actually the result of a gain-of-function effect of a particular downstream aspect of GPCR signaling. It has only recently been appreciated that GPCR can signal via the G protein as well as β-arrestin with different outcomes [16]. Interestingly, many GPCR activate MAP kinase pathways, both by activating G proteins and β-arrestin but often with opposite cellular effects, such as apoptosis versus cell survival [16]. This is particularly interesting from a pharmacological point of view because agonists have been identified

that are biased upon signaling via either G proteins or β-arrestins [17]. ENU-induced mutants that carry specific amino acid changes that may favor coupling of one pathway over the other can represent indispensible tools for testing biased agonism *in vivo*. Indeed, *in vitro* analyses of angiotensin II type 1 receptor (AT_1) with mutations in the DRY motif showed exclusive signaling via β-arrestins [18], and the LPAR1 mutant suggests that residues in the helix 8 might also influence discrimination between G protein-mediated versus β-arrestin-mediated signaling (**Chapter 7**). In addition, *in vivo* mutants carrying amino acid changes in allosteric sites might also prove to be pharmacologically highly relevant in GPCR research [19] to identify compounds that modulate receptor function instead of completely blocking or activating it.

However, as discussed above, ENU mutagenesis is random, and mutating functionally important residues is dependent on chance. A prerequisite of isolating multiple allelic series of a single gene is the availability of large archives of F_1 animals derived from mutagenized males and methods for high-throughput mutation discovery.

REMAINING TECHNICAL CHALLENGES

Creating large archives of mutant alleles

Even though the germ line mutagenicity of ENU is very high, still many F_1 animals have to be generated to knockout a substantial part of the rat genome. Consequently, a problem arises that is generally associated with large-scale mutagenesis studies, namely generating large repositories of F_1 animals and archiving the nonsynonymous mutants. Clearly, keeping a large living repository of multiple mutant lines is expensive and extremely laborious: every year heterozygous carriers have to be crossed out, genotyped and housed. Therefore a lot of effort has been put in optimizing protocols for archiving frozen rat sperm, which can be revived by intracytoplasmic sperm injection (ICSI) [20]. However, although these techniques are common practice for cryopreserving mouse lines, it remains relatively challenging for freezing rat sperm. Indeed, only a few laboratories worldwide are capable of reviving the mutant lines, which is a prerequisite for archiving large collections of mutants. It is feasible to generate large archives of cryopreserved sperm and frozen genomic DNA of F_1 animals that were generated by outcrossing mutagenized males, which can be screened indefinitely for mutations in genes-of-interest. Such an archive of ~5,000 F_1 animals would contain a large number of knockout alleles (~10,000–15,000 using an MSH6 deficient background) and an even higher number of missense alleles. However, besides the enormous logistic challenge of generating such a library, there is still another significant disadvantage of performing ENU mutagenesis *in vivo*, namely the limited number of strains in which germ line ENU mutagenesis is sufficiently efficient. It has been shown that inbred strains perform poorly in ENU mutagenesis screens because of a combination of increased sensitivity to the toxicity of the mutagen and poor breeding properties [4]. Still, many researchers prefer working with inbred animals because of the homogenous genetic background

to assess the phenotypic consequences of mutant alleles. Notably, mutagenesis using ZFN-mediated or transposon-based technology does allow for the generation of mutant alleles in different rat strains, including inbred [9].

The successful isolation and propagation of pluripotent rat ES cells and multipotent SSCs do offer a possibility to produce frozen archives of mutant alleles without the need to generate large collections of living animals. *In vitro* mutagenesis of ES cells or SSCs may also overcome the problem of strain background preferences observed with *in vivo* ENU mutagenesis, although availability of ES cells or SSCs of different strains is limited. Different mutagenic chemicals can be tested to determine the most optimal one for *in vitro* mutagenesis. For example, although it has been shown that ENU is the most efficient mutagen for treating rodents [21], treatment of cultured cells with ethyl methanesulfonate (EMS) may be preferred because of higher mutation frequency and a more suitable mutation spectrum, primarily G/C to A/T transitions. In addition, RNAi can be easily applied to cultured cells to knockdown DNA repair systems, like DNA mismatch repair or nucleotide excision repair, prior to mutagenesis, allowing for the efficiency of mutagenesis to be optimized without the need for ES cells or SSCs in which specific genes have been knocked out. Cultured cells can be mutagenized in a petri dish, clonally expanded and divided for DNA isolation and cryopreservation. ES cells or SSCs carrying mutations in a gene-of-interest can be easily revived, expanded in culture and placed back in, respectively, blastocysts or recipient males for germ line transmission. Primarily, however, optimal conditions for *in vitro* chemical mutagenesis should be determined, including the influence of the load of DNA damage on the pluripotency of ES cells and the ability of SSCs to generate sperm in recipient males.

Mutation discovery

Another crucial technical challenge in ENU target-selected mutagenesis remains the efficient retrieval of mutations that affect protein function in genes-of-interest. Although every F_1 animal (or colony of mutagenized ES cells or SSCs) contains many nonsynonymous mutations, because of the randomness of ENU, complete exomes have to be resequenced to retrieve these mutations. The recent development of massive parallel sequencing technology provides promising opportunities to achieve this. Indeed, we show in **Chapter 6** that a larger collection of preselected genes can be screened in multiple F_1 animals in a single sequencing run by combining massive parallel sequencing with microarray-based enrichment.

The main question is if it is preferable to put a lot of effort in identifying all nonsynonymous mutations in every individual F_1 animal or colony of mutagenized ES cells or SSCs or to screen a large collection of mutants for all the nonsynonymous mutations in a single or small subset of genes-of-interest. An advantage of knowing all nonsynonymous mutations in a mutant animal is that all the background mutations that are all ENU-induced mutations other than the mutation of interest are known. Isolating the mutation-of-interest will be much easier because outcrosses can be accompanied

by genotyping for 'background' mutations closest to the mutation of interest. Because this can limit the number of outcrosses before assessing the phenotype specifically induced by homozygosity of the mutation-of-interest, it can save considerable time and cost. However, whole exome or genome sequencing is still expensive, and it is not cost-effective for isolating allelic series of a specific gene. In contrast, screening a subset of genes in a large archive will enable the isolation of allelic series of genes-of-interest. This method will be particularly effective for screening frozen archives because these can be screened indefinitely. However, there are still major technical challenges, like pooling multiple samples for next generation sequencing and enriching for genes-of-interest. Although **Chapter 6** provides the proof-of-principle that it is feasible, the method has to be considerably scaled.

CONCLUSIONS

Although the technology to generate rat knockout models is relatively new, these models have already been implemented in a variety of biomedical fields, including tumorigenesis [22,23,24], behavioral biology [25], metabolism [26] and immunology [27] (discussed in **Chapter 1**). Because multiple techniques are currently available to generate knockout rats and large mutagenesis screens are being conducted, it can be expected that rat reverse genetics will become a general tool to investigate specific aspects of human physiology and disease. For the future, the question will be how far we want to push the field of rat reverse genetics. Currently, large-scale mutagenesis screens are conducted, but ZNF-mediated targeted approaches are also used, with the purpose of generating a lot of mutant rat models. Is one of the goals to generate a knockout for every gene in the rat genome, similar to what is being done in mice [3], and are these models worth the effort? Furthermore, is it possible to isolate knockout alleles for every (relevant) gene given the specific biases of all the techniques? For example, can HR take place at every genomic locus and still give rise to germ line transmission, or can ZFN-mediated DSBs be introduced everywhere when the epigenetic landscape is taken into account? ENU mutagenesis is considered to be the most random, unbiased form of mutagenesis, but what is still not known is to what degree. Possibly, the enormous amounts of data from massive parallel sequencing will tell.

ENU target-selected mutagenesis will continue to play an important role in generating rat knockout models. Although mutations cannot be targeted, the technique has proven to be compatible with important technological developments in the field of genetics, including massive parallel sequencing. Its unique property of generating highly efficient single point mutations can have a significant impact on developing mutant animals to understand human physiology and model human disease.

REFERENCES

1. Jacob HJ, Kwitek AE (2002) Rat genetics: attaching physiology and pharmacology to the genome. Nat Rev Genet 3: 33-42.
2. Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6: 507-512.
3. Gondo Y (2008) Trends in large-scale mouse mutagenesis: from genetics to functional genomics. Nat Rev Genet 9: 803-810.
4. Smits BM, Mudde JB, van de Belt J, Verheul M, Olivier J, et al. (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. Pharmacogenet Genomics 16: 159-169.
5. Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nat Biotechnol 21: 645-651.
6. Noveroske JK, Weber JS, Justice MJ (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. Mamm Genome 11: 478-483.
7. Kitada K, Ishishita S, Tosaka K, Takahashi R, Ueda M, et al. (2007) Transposon-tagged mutagenesis in the rat. Nat Methods 4: 131-133.
8. Lu B, Geurts AM, Poirier C, Petit DC, Harrison W, et al. (2007) Generation of rat mutants using a coat color-tagged Sleeping Beauty transposon system. Mamm Genome 18: 338-346.
9. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, et al. (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325: 433.
10. Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, et al. (2009) Transposon-mediated genome manipulation in vertebrates. Nat Methods 6: 415-422.
11. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-236.
12. Zhang XM, Ramalho-Santos M, McMahon AP (2001) Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node. Cell 105: 781-792.
13. Parnot C, Miserey-Lenkei S, Bardin S, Corvol P, Clauser E (2002) Lessons from constitutively active mutants of G protein-coupled receptors. Trends Endocrinol Metab 13: 336-343.
14. Shenker A, Laue L, Kosugi S, Merendino JJ, Jr., Minegishi T, et al. (1993) A constitutively activating mutation of the luteinizing hormone receptor in familial male precocious puberty. Nature 365: 652-654.
15. Liu G, Duranteau L, Carel JC, Monroe J, Doyle DA, et al. (1999) Leydig-cell tumors caused by an activating mutation of the gene encoding the luteinizing hormone receptor. N Engl J Med 341: 1731-1736.
16. Lefkowitz RJ, Shenoy SK (2005) Transduction of receptor signals by beta-arrestins. Science 308: 512-517.
17. Violin JD, Lefkowitz RJ (2007) Beta-arrestin-biased ligands at seven-transmembrane receptors. Trends Pharmacol Sci 28: 416-422.
18. Wei H, Ahn S, Shenoy SK, Karnik SS, Hunyady L, et al. (2003) Independent beta-arrestin 2 and G protein-mediated pathways for angiotensin II activation of extracellular signal-regulated kinases 1 and Proc Natl Acad Sci U S A 100: 10782-10787.
19. Kenakin T, Miller LJ (2010) Seven Transmembrane Receptors as Shapeshifting Proteins: The Impact of Allosteric Modulation and Functional Selectivity on New Drug Discovery. Pharmacol Rev.
20. Mashimo T, Yanagihara K, Tokuda S, Voigt B, Takizawa A, et al. (2008) An ENU-induced mutant archive for gene targeting in rats. Nat Genet 40: 514-515.
21. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, et al. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A 76: 5818-5819.
22. Amos-Landgraf JM, Kwong LN, Kendziorski CM, Reichelderfer M, Torrealba J, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. Proc Natl Acad Sci U S A 104: 4036-4041.
23. Cotroneo MS, Haag JD, Zan Y, Lopez CC, Thuwajit P, et al. (2007) Characterizing a rat Brca2 knockout model. Oncogene 26: 1626-1635.

24. van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, et al. (2008) Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis 29: 1290-1297.
25. Homberg JR, Olivier JD, Smits BM, Mul JD, Mudde J, et al. (2007) Characterization of the serotonin transporter knockout rat: a selective change in the functioning of the serotonergic system. Neuroscience 146: 1662-1676.
26. Mul J, Yi CX, van den Berg SA, Ruiter M, Toonen P, et al. (2009) Pmch expression during early development is critical for normal energy homeostasis. Am J Physiol Endocrinol Metab.
27. Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, et al. (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5: e8870.

ADDENDUM

Nederlandse samenvatting
Acknowledgements
Curriculum Vitae
Publication list

NEDERLANDSE SAMENVATTING

Het breken van de genetische code

Genetische informatie is opgeslagen in het DNA van elke cel in een 4-letterige code, waarin alle instructies voor het normaal functioneren van een organisme liggen opgeslagen. Als er een fout zit in deze code dan kan dat leiden tot ziekte. Inmiddels kennen we de volgorde van de genetische sequentie van de mens, maar dit betekent niet dat we de code ook begrijpen. Om de menselijke genetische code te breken, maken we gebruik van model organismen. Door de genetische sequenties te bepalen van andere organismen en deze te vergelijken met die van de mens, kunnen we de overeenkomsten en verschillen bepalen en speculeren over het functioneel belang hiervan. Uit deze studies kan bijvoorbeeld blijken dat een genetisch element, zoals een gen dat codeert voor een eiwit, geconserveerd is in vele organismen. Wanneer deze organismen evolutionair zowel dicht als ver van de mens staan, kan dit betekenen dat dit element een belangrijke functie vervult. Immers als het element een essentiële fysiologische rol speelt, zal een verandering in de code die de functie beïnvloedt niet compatibel zijn met leven. Er ligt dan een evolutionair negatieve druk[1] op veranderingen in de code van dit element die de normale functie hiervan beïnvloedt. Echter kunnen sequentie verschillen in genetische elementen tussen organismen de fenotypische[2] verschillen tussen die organismen mogelijk verklaren. Om deze hypotheses over het functioneel belang van genetische elementen te testen, wordt gebruikt gemaakt van gerichte genetische modificaties in model organismen.

Over het algemeen kunnen technieken om genetische elementen te manipuleren onderverdeeld worden in twee klassen welke afhankelijk zijn van de onderzoeksvraag: fenotype- of gen-gedreven genetica. Klassieke fenotype-gedreven genetica begint met het vaststellen van een fenotype, een meetbaar verschil dat voorkomt in de populatie van een bepaald model organisme. Door middel van moleculaire biologische technieken kunnen vervolgens de genetische elementen die ten grondslag liggen aan het fenotype van interesse geïdentificeerd worden. In het algemeen wordt de genetische variatie in een populatie verhoogd door onwillekeurige mutagenese[3], bijvoorbeeld middels het toedienen van een chemische stof die puntmutaties[4] aanbrengt zoals ENU. **Hoofdstuk 2** behandelt de technische details van een dergelijke fenotype-gedreven screen.

Gen-gedreven genetica begint met het muteren van een bepaald gen, gevolgd door het vaststellen van fenotypische gevolgen van de mutatie. Deze methode heeft veel aan populariteit gewonnen sinds de volgorde van de genetische sequenties van vele organismen vastgesteld is. Homologe recombinatie in embryonale stamcellen

[1] Dit betekent dat deze verandering op den duur zal verdwijnen uit de populatie.

[2] Het fenotype is het totaal aan waarneembare eigenschappen (kenmerken) van een organisme.

[3] Onwillekeurige mutagenese is het aanbrengen van mutaties die overal in het genoom gelokaliseerd kunnen zijn zonder (of zo min mogelijk) dat deze locatie ergens door beïnvloed wordt.

[4] Een puntmutatie is een verandering van een enkele base.

in de muis is hiervoor een erg succesvolle techniek gebleken (**Hoofdstuk 1**). Echter is deze techniek (nog) niet beschikbaar voor alle model organismen. Dit komt door het ontbreken van pluripotente[5] embryonale stamcellen. Die verschillende model organismen kunnen wel, afhankelijk van de onderzoeksvraag, een voordeel kunnen hebben ten opzichte van de muis. Daarom zijn er alternatieve methoden bedacht om genetische elementen van interesse in een model organisme naar keuze te modificeren.

De rat als genetisch model organisme

In tegenstelling tot het genetisch model organisme de muis, wordt de rat als model organisme traditioneel vooral gebruikt in fysiologische, farmacologische en toxicologische studies. Beide organismen zijn zoogdieren en staan dus relatief dicht bij de mens en zijn bovendien makkelijk te fokken. De muis is momenteel het meest gebruikte genetische zoogdier model organisme. Dit is een gevolg van de zeer uitgebreide technische mogelijkheden met betrekking tot genetische modificatie. Het gebruik van ratten als model organisme heeft echter enkele noemenswaardige voordelen. Deze hebben bijna allemaal te maken met de relatief grote omvang van het dier wat microchirurgie en manipulatie makkelijker maken (zie **Hoofdstuk 1** voor enkele voorbeelden). Ook zijn sommige menselijke ziekten beter te modeleren in de rat, voornamelijk psychische aandoeningen maar ook pathofysiologische aandoeningen, zoals hypertensie. Deze voordelen maakt het rattenmodel complementair aan het muizenmodel, dat weer andere voordelen heeft ten opzichte van de rat. Door middel van fokken en selecteren op natuurlijk voorkomende fenotypische kenmerken zijn er vele rattenstammen gegenereerd die ieder model staan voor een humane complexe aandoening. Echter waren de technische middelen tot gerichte genetische manipulatie in de rat nog lange tijd erg beperkt. **Hoofdstuk 1** behandelt recente technische ontwikkelingen en mogelijkheden die het veld van gen-gedreven genetica in de rat heeft doorgemaakt.

Chemische mutagenese als middel voor gen-gedreven genetica in de rat

Een van de eerste technieken, die gebruikt werd om genetisch gemodificeerde ratten te genereren, maakt gebruikt van onwillekeurige chemische mutagenese van de mannelijke spermatogoniale stamcellen door een behandeling met *N*-ethyl-*N*-nitrosourea (ENU). Deze stof is zeer efficiënt in het introduceren van puntmutaties in snel delende cellen, zoals spermatogoniale stamcellen. Mannelijke ratten die behandeld zijn met ENU worden gekruist met onbehandelde vrouwelijke ratten. Hieruit ontstaat een F_1 populatie[6], waarvan elk individu unieke onwillekeurige heterozygote[7] mutaties in hun

[5] De mogelijkheid om alle cellen in een organisme te vormen, maar niet extra-embryonaal weefsel.
[6] De eerste generatie individuen die ontstaat na het kruizen van gemutageniseerde mannen met onbehandelde vrouwen.
[7] Van elk gen heeft een individu twee kopieën (allelen) namelijk een van de vader en een van de moeder (met uitzondering van genen die gelokaliseerd zijn op de X en Y chromosomen in

genoom dragen. Vervolgens wordt DNA afgenomen van alle F_1 dieren en worden genen van interesse gescreend. Er wordt hierbij gezocht naar mutaties die de normale functie van het genproduct[8] beïnvloedt, zoals de introductie van een premature translationele stopcodon[9] wat meestal tot een genetische knockout[10] leidt. Als er een interessante mutatie wordt gevonden in een F_1 dier dan wordt deze een aantal keer uitgekruist en vervolgens wordt de mutatie homozygoot gekruist. Dit betekent dat een individu in beide allelen van het gen de mutatie draagt. Hierdoor kan het fenotypisch effect van de mutatie in dit individu bestudeerd worden. In **Hoofdstuk 2** wordt een gedetailleerde beschrijving van de ENU doelgerichte mutagenese techniek gegeven.

Efficiëntere ENU mutagenese in DNA mismatch repair deficiënte rat

Een van de eerste genetische ratten knockout modellen dat gegenereerd werd door middel van ENU doelgerichte mutagenese was voor een gen genaamd *Msh6*. Dit gen codeert voor het DNA mismatch repair (MMR) component MSH6, dat samen met MSH2 verkeerde basen combinaties en kleine insertie en deletie lussen (IDL's) herkent[11]. Deze fouten kunnen optreden tijdens het repliceren van DNA, voordat een cel gaat delen. Zij kunnen ernstige gevolgen kunnen hebben voor het correct functioneren van genproducten, doordat ze bijvoorbeeld een codons genen muteren die coderen voor functioneel belangrijke aminozuren in eiwitten. Het is dan ook erg belangrijk dat deze fouten gecorrigeerd worden. **Hoofdstuk 3** laat zien dat de MSH6 knockout rat een verhoogde spontane tumorvorming vertoont, doordat dit model geen verkeerde basen combinaties en IDL's kan corrigeren.

---

mannelijke individuen). Als er verschillende allelen bestaan in een populatie dan kan het zo zijn dat een individu beiden allelen draagt. Dat individu is dan heterozygoot voor het verschil in dat gen. In het geval van het werk in dit proefschrift worden de genetische verschillen door middel van chemische mutagenese geïntroduceerd.

[8] Het eiwit waar het gen voor codeert.

[9] Eiwitten bestaan uit aminozuren. In een gen wordt een aminozuur gecodeerd door een triplet basen (een codon; drie letters van de genetische code). De laatste aminozuur triplet wordt gevolgd door een stopcodon (TAA, TAG, TGA), dat niet meer codeert voor een aminozuur, maar aangeeft dat het eiwit volledig gesynthetiseerd is. Een mutatie kan een aminozuur in het midden van een eiwit veranderen in een stopcodon, wat een premature stopcodon heet en tot gevolg heeft dat het eiwit niet meer of onvolledig gevormd wordt.

[10] Een dier waarbij een gen niet meer codeert voor een functioneel eiwit.

[11] Een DNA molecuul bestaat uit twee strengen. In beide strengen liggen complementaire bases (de vier letters in de genetische code) tegenover elkaar: een adenine (A) paart altijd met een thymine (T), en een guanine (G) altijd met een cytosine (C). Tijdens het proces van DNA replicatie, dat voor een celdeling plaatsvindt, worden de strengen gescheiden en worden er twee nieuwe DNA moleculen gevormd met de twee oorspronkelijke enkele strengen als mal. Het kan echter voorkomen dat er foutjes optreden tijdens de replicatie, zoals het verkeerd paren van basen of het vergeten of extra toevoegen van een base op een locaties waar veel dezelfde basen achter elkaar liggen. Hiervoor wordt de term IDL gebruikt.

Interessant genoeg worden de genetische fouten die ENU veroorzaakt en resulteren in mutaties, ook (in ieder geval deels) door het DNA MMR systeem herkend. **Hoofdstuk 4** beschrijft het gebruik van DNA MMR deficiëntie in de MSH6 knockout rat om de efficiëntie van ENU doelgerichte mutagenese te verhogen. Hoewel MSH6 knockout ratten gevoeliger zijn voor de toxiciteit van ENU, wat resulteert in afgenomen fertiliteit, resulteren lagere doses tot een verhoogde mutatie frequentie. Bovendien verandert het mutatie spectrum op dergelijke wijze dat de kans dat een ENU geïnduceerde mutatie in een MSH6 deficiënte achtergrond een premature stopcodon veroorzaakt groter is vergeleken bij ratten met wild-type[12] achtergrond.

G eiwit gekoppelde receptoren mutanten

De efficiëntie van ENU doelgerichte mutagenese in een DNA MMR deficiënte achtergrond is getest door in een grote groep genen, die coderen voor G eiwit gekoppelde receptoren (GPCR's), te zoeken naar interessante mutaties in F_1 dieren. **Hoofdstuk 5** beschrijft de resultaten van dit experiment. Er is een aantal redenen waarom genen die coderen voor GPCR's interessant zijn om te screenen door middel van ENU doelgerichte mutagenese. GPCR's zijn de grootste familie van membraan gebonden receptoren die een groot spectrum van signaal stoffen kan binden, zoals hormonen en neurotransmitters, maar ook licht en protonen. Dit maakt dat deze soort receptoren erg belangrijke functies vervult in de menselijke fysiologie. Daarnaast zijn zij ook vaak betrokken bij pathologie. Bovendien vormen GPCR's de grootste groep van receptoren waar de huidige medicatie zich op richt. Ten slotte is de typische structuur van zeven transmembrane domeinen met een extracellulaire N-terminus[13] en een intracellulaire C-terminus[14] erg geconserveerd tussen de verschillende GPCR's. Dit maakt bioinformatische voorspellingen betreffende eventuele effecten van de ENU geïnduceerde mutaties mogelijk. Het genereren van GPCR mutanten door middel van ENU doelgerichte mutagenese in een DNA MMR deficiënte achtergrond blijkt erg effectief. Er zijn meerdere knockout allelen gevonden voor genen met een bekende functie, zoals de melanocortin 4 receptor (MC4R). Daarnaast zijn er ook knockout allelen gevonden voor receptoren met nog onbekend ligand[15] en functie.

Een van de grote voordelen van ENU mutagenese is dat het de mogelijkheid biedt allelische series te genereren. Dit betekent dat er verschillende ongelinkte mutaties gevonden kunnen worden in hetzelfde gen, die verschillende functionele

[12] Met een wild-type achtergrond wordt bedoelt dat de individuen de twee oorspronkelijke (dus zonder geïnduceerde mutaties) allelen van een gen dragen. Het genproduct behoort dan normaal te functioneren.

[13] De N-terminus of amino-terminus is het eerste aminozuur van een eiwit of polypeptide. Deze bevat een ongebonden aminogroep.

[14] De C-terminus of carboxyl-terminus is het laatste aminozuur van een eiwit of polypeptide. Deze bevat een ongebonden carboxylgroep.

[15] Het signaal molecuul dat de receptor bindt en de activiteit beïnvloedt.

gevolgen kunnen hebben. Er kunnen dus knockout allelen worden gevormd door het introduceren van premature stopcodons. Echter kunnen mutaties die aminozuur veranderingen (missense) veroorzaken als gevolg hebben dat een eiwit minder tot expressie komt of verminderde katalytische activiteit heeft (hypomorfisch). Anderzijds kunnen mutaties van aminozuren die belangrijk zijn om het eiwit in een inactieve staat te houden, resulteren in een actievere conformatie (hypermorfisch). Verder kunnen mutaties ook aminozuren veranderen die betrokken zijn voor eiwit-eiwit interacties. Dit kan verstrekkende functionele gevolgen hebben voor de functie van deze eiwitten. In de screen naar GPCR mutanten zijn er missense mutanten gevonden, die door middel van bioinformatische analyses in al deze categorieën ingedeeld kunnen worden. Voor een van deze mutanten, de LPAR1^{M318R}, wordt door middel van moleculaire, functionele en biologische analyses aangetoond dat de mutatie inderdaad leidt tot een functioneel effect (**Hoofdstuk 7**).

Identificatie van ENU geïnduceerde mutaties

Er zijn verschillende methoden toe te passen om ENU geïnduceerde mutaties te identificeren (deze staan vermeld in **Hoofdstuk 2**). Lange tijd was het zoeken naar ENU geïnduceerde mutaties in genen van interesse in F_1 dieren door middel van capillair sequensen[16] de meest efficiënte manier om alle mutaties in die genen te identificeren. Echter recentelijk hebben er grote technologische ontwikkelingen plaatsgevonden op het gebied van DNA sequensen. Deze kunnen samengevat worden onder de noemer "massief parallel sequensen". Dit stelt onderzoekers in staat grote stukken DNA – tot hele genomen en van meerdere individuen – snel te sequensen. **Hoofdstuk 6** beschrijft de toepassing van deze revolutionaire techniek voor ENU doelgerichte mutagenese. In deze studie wordt aangetoond dat het mogelijk is een grote groep genen van interesse, waarvoor specifiek verrijkt wordt, te bekijken in meerdere dieren in een enkel experiment. Uiteindelijk biedt deze toepassing de mogelijkheid om complete exomen[17] of zelfs complete genomen van F_1 individuen te sequensen. Anderzijds kan er ook door middel van deze techniek zeer snel en efficiënt gezocht worden naar allelische series van enkele genen van interesse in een zeer grote groep F_1 dieren.

Conclusies

Het veld van gen-gedreven rattengenetica heeft recentelijk vele belangrijke technologische ontwikkelingen doorgemaakt. In **Hoofdstuk 8** wordt bediscussieerd wat de voor- en nadelen zijn van chemische mutagenese ten opzichte van deze nieuwere technieken. Ook worden de mogelijkheden van chemische mutagenese in het model organisme de rat besproken. Zo kan massief parallel sequensen een grote

[16] Capillair sequensen is een manier om de base volgorde van een stuk DNA te bepalen. Het principe van deze techniek bestaat uit het specifiek labellen van de basen, gevolgd door het scheiden van de gelabelde fragmenten op grootte.

[17] Alle base tripletten in een genoom die coderen voor aminozuren.

impact hebben op het screenen van grote verzamelingen van genen, met name als dit gecombineerd wordt met bevroren archieven van F_1 dieren of gemutageniseerde embryonale stamcellen. Hoe groter het aantal gemutageniseerde genomen, hoe groter de kans dat je gen van interesse vaak gemuteerd wordt. Op deze manier kan chemische mutagenese een belangrijke rol spelen. Dit is juist omdat het onwillekeurig zeer efficiënt puntmutaties in het gehele genoom aanbrengt. Ten eerste bootst dit de meest voorkomende vorm van genetische variatie in de mens na, namelijk enkele basenparen verschillen. Ten tweede kan de mogelijkheid tot het genereren van allelische series van genen van interesse een zeer krachtig middel zijn om de functies van genen te onderzoeken.

ACKNOWLEGDEMENTS

Zo, na vier jaar is dit proefschrift voltooid en het was me niet gelukt zonder de steun van alle bijzondere mensen om me heen. Graag richt ik deze woorden tot hen als dankbetuiging.

Uiteraard wil ik allereerst Edwin bedanken voor alle steun, tijd en moeite die hij in me heeft gestoken om me wetenschappelijk te vormen. Ik heb enorm veel van je geleerd de laatste vier jaar, zowel in het lab, achter de computer, op congressen, tijdens besprekingen op je kantoor en gesprekken tijdens de vele sociale bezigheden. Ik ben je dankbaar dat je me de kans hebt gegeven om promotie onderzoek te mogen doen in je lab. Dat je me enerzijds de juiste kant op gestuurd hebt zodat ik dit boekje kon voltooien, maar me ook genoeg vrijheid hebt geven dat ik me heb kunnen vormen tot de onderzoeker die ik nu ben. Het was me een waar genoegen voor je te mogen pipetteren!

Cheers to all (formal) members of the Cuppen Group! The last 4 years were unforgettable and I owe that in a very large part to you. Allereerst wil ik het "rat knockout screen team" bedanken. Pim, Mark en Henk, we waren een geweldig team! Ook al hebben we (nog) geen geschiedenis geschreven, we waren wel succesvol en hebben in dat proces heel veel lol getrapt. Ook alle (ex-)AIO's wil ik bedanken: Harma, Joram (met een m), Sam, Michal, Jos en Sebastiaan (alias Sirbastard). Ik dank jullie voor alle brainstorm sessies, journal clubs, geintjes en voor het feit dat het grootste deel van jullie het uit heeft kunnen houden een kamer met mij te delen. Ook wil ik de "bioinformagicians" bedanken voor al hun onmisbare hulp en begeleiding: Victor (I don't know if I will ever beat you at the table tennis table, or did I already?!), Ies, Marieke, Frans-Paul alias "Long Frans" en Elzo. Ik wil ook nog speciaal de studenten bedanken die de pech hadden door mij begeleid te worden: Henk (de eerste!), Marc, "je weet tóch" Cheung, Sharon, Milou, Sumanth en last but not least Dylan. Ik geloof dat ik de reputatie heb nogal veeleisend te zijn (jullie hadden mij ook niet té serieus moeten nemen), maar ik hoop dat ik "streng doch rechtvaardig" was. Ik heb in ieder geval veel lol met jullie gehad! Verder veel dank aan Judith, Jose, Nico, Ewart, Esther en iedereen die ik vergeten ben om te noemen. Ook veel dank aan alle Hubrechters buiten de Cuppen Groep, met speciale groet aan de dierenverzorgers, Jeroen, Harry en Stieneke.

Heel bijzonder was het werken met ratten, nog bijzonderder was het samenwerken met rattenmensen: Joram, Pim, Judith en Dylan (ja Dyleronio, voor mij hoor jij ook bij de 'gang') het was een voorrecht om met jullie vele uurtjes door te brengen met deze unieke diertjes. We hebben gelachen, gehuild, gezeurd, gevochten, lol getrapt, scheten gelaten, gezweet en pikante foto's genomen. Alles kon en wat in stal gebeurde, bleef in

de stal! Lief en leed hebben we gedeeld in dat benauwde, donkere hol en daar ben ik jullie zeer dankbaar voor.

Ten slotte wil ik mijn thuisfront bedanken: Michelle jij bent mij alles en zonder jou ben ik niets. Lieve en Veerle , jullie zijn mijn motivatie, mijn inspiratiebronnetjes. Ik wil mijn ouders bedanken voor alle liefde, voor een veilige thuishaven, voor de schoppen onder m'n kont, de draaien om mijn oren en vooral ook voor veel steun. Dank aan mijn zussen Wendy en Judith: we lijken zo veel op elkaar en toch verschillen we zoveel, we zijn een raadsel voor genetici! Ik wil Judith nog speciaal bedanken voor alle hulp aan de artistieke vormgeving van dit proefschrift: je bent geweldig! Dank aan Henk, Desiree, Stefan, Emile, Lauren, Mika en alle andere familieleden (en dat zijn er heel erg veel). Ook wil ik het niet-familiaire thuisfront bedanken voor alles: Harald en Jos, Jasper, Mike, Mark, Kelvin en Sumeyra en natuurlijk Róbert. Als ik iemand vergeten ben: mijn nederige excuses.

CURRICULUM VITAE

Ruben van Boxtel was born on November 8, 1979 in Kaatsheuvel, The Netherlands. In 1999 he graduated from the Dr. Mollercollege in Waalwijk after which he started his study Biology at the Radboud University Nijmegen in that same year. He performed his first internship at the department of Cellular Animal Physiology at the Radboud University Nijmegen under supervision of dr. Geert Corstens in the group of Prof. dr. Eric Roubos. Ruben performed his second internship at the department of Cell Biology at the University of Córdoba in Spain under supervision of dr. David Cruz García in the group of Prof. dr. María del Mar de Malagon. In 2006 he graduated, receiving a M.Sc. degree in Biology. In April that same year he began working as PhD student in the group of Prof. dr. Edwin Cuppen (Genome Biology) at the Hubrecht Institute in Utrecht. The results of this work are described in this thesis. In September 2010, Ruben will start as a postdoctoral researcher in the group of Prof. dr. Paul Coffer (Molecular Immunology) at the UMC in Utrecht.

PUBLICATION LIST

van Boxtel R, Cuppen E. The emerging genetic rat model. *Submitted*.

van Boxtel R[*], Nijman IJ[*], Mokry M[*], Toonen P, de Bruijn E, Cuppen E. Targeted genomic enrichment and next-generation sequencing for gene-driven reverse genetic approaches. *Submitted*.

[*]Contributed equally

van Boxtel R, Cuppen E. Generation of genetically modified rodents using random ENU mutagenesis. Methods Mol Biol. *In press*.

van Boxtel R, Vroling B, Toonen P, Nijman IJ, van Roekel H, Verheul M, Baakman C, Guryev V, Vriend G, Cuppen E. Systematic generation of in vivo G protein-coupled receptor mutants in the rat. Pharmacogenomics J. *In press*.

van Boxtel R, Gould MN, Cuppen E, Smits BM. ENU mutagenesis to generate genetically modified rat models. Methods Mol Biol. 2010;597:151-67.

van Boxtel R, Toonen PW, Verheul M, van Roekel HS, Nijman IJ, Guryev V, Cuppen E. Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics. 2008 Oct 7;9:460.

van Boxtel R, Toonen PW, van Roekel HS, Verheul M, Smits BM, Korving J, de Bruin A, Cuppen E. Lack of DNA mismatch repair protein MSH6 in the rat results in hereditary non-polyposis colorectal cancer-like tumorigenesis. Carcinogenesis. 2008 Jun;29(6):1290-7.

Corstens GJ, **van Boxtel R**, van den Hurk MJ, Roubos EW, Jenks BG. The effects of disruption of A kinase anchoring protein-protein kinase A association on protein kinase A signalling in neuroendocrine melanotroph cells of Xenopus laevis. J Neuroendocrinol. 2006 Jul;18(7):477-83.