Natural and Induced

Genetic Variation in the Rat

Bart M. G. Smits

Natural and Induced Genetic Variation in the Rat

Natuurlijk Voorkomende en Geïnduceerde Genetische Variatie in de Rat
(met een samenvatting in het Nederlands)

Proefschrift ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof. Dr. W.H. Gispen, ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op woensdag 30 november 2005 des middags te 12:45 uur

door

Bartholomeus Mathijs Godefridus Smits
geboren op 28 juni 1978, te Roosendaal en Nispen

Promotor: Prof. Dr. R.H.A. Plasterk

Copromotor: Dr. E. Cuppen

Cover: Artistic impression of the rat knockout project.
On the right, a large library of F1 mutagenized rats is depicted, from which DNA samples are extracted. By high-throughput resequencing, induced heterozygous point mutations are discovered.

ISBN: 90-393-4094-3

Printed by: Wöhrmann Print Service

The work described in this thesis was performed at the Hubrecht Laboratory of the Royal Netherlands Academy of Arts and Sciences (KNAW), within the Graduate School for Developmental Biology, Utrecht, The Netherlands. This study was financially supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program (IOP) on Genomics and NV Organon.

Voor mijn ouders,
Martyn,
en Marijntje.

Contents

CHAPTER 1

Introduction:
Rat Genetics: The Next Episode

Adapted from invited review Trends in Genetics

Rat Genetics: The Next Episode

Bart M. G. Smits and Edwin Cuppen*

Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

*Corresponding Author

With the complete rat genome sequence available and many QTLs identified for many aspects of a variety of complex human diseases, it is now time for the next stage in rat genetic research: the identification and causal affirmation of underlying polymorphisms and genes. Recent developments in the area of Single Nucleotide Polymorphism (SNP)-based genetic markers and technologies to manipulate the rat genome will undoubtedly be important tools in this next episode.

Rat as a model for complex human disease

The laboratory rat is one of the most extensively studied model organism for various aspects of human health and disease, like physiology, toxicology, and neurobiology, as well as for drug development. Some diseases are best mimicked in the rat, in particular, neurodegenerative diseases and disorders affecting higher brain function, like schizophrenia, anxiety, depression, and addiction. In addition, rats are well suited for surgical manipulations due to their relatively large size and many models for neurological disorders are actually established by surgical procedures [1].

Selective breeding and thorough characterization has resulted in the establishment of over 200 rat inbred strains that specifically resemble many features of human common disease [2-4]. Most phenotypes in these models are complex traits, meaning that multiple functional elements contribute to the establishment of the pathological state. Many mapping studies have been performed to dissect the underlying genetic basis, resulting in the identification of nearly 1,000 Quantitative Trait Loci (QTLs), ranging in size from 10 to over 190 Kb (http://rgd.mcw.edu/objectSearch/qtlQuery.jsp) and containing hundreds or even up to a thousand genes. Although several functional components have recently been discovered in such regions, the vast majority will need additional efforts to be deciphered [5]. Recent developments in rat genetics and genomics are likely to play a crucial role in this process.

Genome sequence

In April 2004, rat joined mouse and human as the third vertebrate for which the complete genome sequence (more than 90 % coverage) has been determined [6]. All these genomes encode similar numbers of genes, although some genes (around 400), mostly involved in pheromone production, immunity, chemosensation, detoxification, or proteolysis, appear to be rat-specific. Almost all known human disease genes have orthologs in rat. Moreover, the eutherian core of the

rat genome, as defined by orthologous alignments to mouse and human, comprises around 40% and contains the vast majority of exons and known regulatory elements. Around 30% of the genome aligns only with mouse, suggesting the presence of rodent specific segments and reflecting their more recent common ancestry [6].
The complete sequence of a third mammalian genome is primarily of great value for comparative genomics [7]. Comparative genomics intends to integrate different types of genomic data from multiple species. The comparative value is manifested at nucleotide level, by allowing alignments between the three species, as well as on phenotypic level. A versatile tool to explore comparative genomics between rat, mouse, and human is VCMap (Virtual Comparative Map; http://rgd.mcw.edu/VCMAP), which enables investigators to link the physiology of rat with the genetics of mouse and rat, and the clinical significance of human [8]. The availability of the rat genome will undoubtedly accelerate genetic studies in this model, as it allows researchers to accurately determine positions of genetic markers and QTLs, and the identity of candidate genes located in these intervals. However, genetic or phenotypic resemblance alone does not guarantee a common underlying molecular mechanism. Experimental verification, employing the benefits of model organisms, remains essential.

QTLs and mapping

Years of comprehensive characterization of inbred strains for a diversity of traits have paved the path for geneticists, whose turn it is now to reveal the genetic nature of these multigenic traits. This started off with the identification of QTLs, which was initially done by whole-genome scanning [9,10]. QTLs were identified and roughly mapped to a large chromosomal region by correlating genetic markers with trait values in the mapping population. The resulting intervals are commonly over 10 cM or 15% of a chromosome/linkage group.
Currently, a systematic, powerful way to map and further dissect QTLs can be found in the generation of consomic, congenic, and Recombinant Inbred (RI) strains from two inbred strains differing in the trait of interest [5,11,12]. Panels of consomic strains, in which each chromosome is introgressed into the isogenic background of another inbred strain, can be constructed by marker-assisted selection [13]. They can be used to assign QTLs to chromosomes or to rapidly develop congenic strains, in which parts of chromosomes are introgressed into another inbred background, also by marker-assisted selection. Subsequently, a panel of congenic lines harboring overlapping pieces of a QTL region can be used to accurately determine the QTL boundaries. RI strains are basically randomly bred congenic strains that can be used similarly to map QTLs [14]. The major advantage of congenics or RI strains is that recombination events isolate potential causal factors in an isogenic background, thereby making parts of the complex disease phenotype monogenic.
Genetic markers play an important role in the mapping of QTLs and the development of consomic and congenic strains. Hence, detailed maps of rat microsatellite markers or SSLPs (Simple Sequence Length Polymorphisms) have been developed [15-17], as well as automated, multiplexing genotyping technology for SSLP genome scanning [18]. Although currently over 17,000 entries are stored in the SSLP database of RGD (Rat Genome Database: http://rgd.mcw.edu), the major drawback is the relatively moderate density between strains of interest in a particular QTL region. SSLP-based genome scanning will identify QTL intervals of 10 cM or more, containing hundreds of genes. At this point, the identification of the genes that contribute to a QTL, especially those that contribute to less than 5% to the trait, is the most challenging part, mainly because of locus heterogeneity, epistasis, low penetrance, and limited statistical power [19,20]. The use of a reasonable number of congenic animals (~1,000) could reduce a QTL interval to roughly 1-2 cM [11], which would still contain about 20 to 40

candidate genes. Many QTL mapping projects in the rat become idle at this point as two bottlenecks need to be solved to proceed.
First, the number of recombination events has to be increased to decrease the candidate gene region. To create sufficient recombination events, sample size must be high enough. In mouse, several QTL fine-mapping strategies have focused on production of mapping populations with increased recombination events based on various principles, like AIL (Advanced Intercross Lines) [21], the use of HS (Heterogenous Stocks) [22], large-scale production of RI strains [23], the use of RIST (Recombinant Inbred Segregation Test) [24], Yin-Yang crosses [25], outbred stocks [26], and more [5,24].
Second, the amount of informative genetic markers between the strains of choice has to be increased. Polymorphic SSLP-based markers are usually becoming very rare at these resolutions, requiring the switch to other types of markers, such as Single Nucleotide Polymorphisms (SNPs). Although the biallelic nature of SNPs makes them less informative compared to the multiallelic SSLPs, the ease of detection, frequent occurrence, and versatility justifies their exploitation for mapping experiments. Besides potential high density, which is necessary for mapping experiments, SNP markers tend to occur closer to genes than SSLPs [27]. Assuming that the functional elements of QTLs are often associated with genes, the chance to find close linkage with a SNP or multiple coupled SNPs (haplotype) will be larger as compared to SSLPs. In addition, it is highly likely that most of the causal variants are SNPs themselves. Both experimental and *in silico* SNP discovery tactics have been followed to get an impression on SNP characteristics and frequencies in the rat [28-30]. These studies are biased towards more constrained coding regions in the genome, having an average SNP frequency of 1 per 630 bp [29]. Although this frequency will be much less for an arbitrary combination of inbred strains, mostly depending on their ancestral relationship, their density is larger than for SSLPs. For example, Zimdahl et al., (2004) found over 4,100 candidate coding SNPs between strains WKY and BN (1 in 1,100 bp). The SSLP repository of RGD contains a bit less than 4,100 SSLPs for WKY genome-wide. The SSLP map is already saturated, but the SNP collection available via dbSNP, the central SNP repository of NCBI (http://www.ncbi.nlm.nih.gov/SNP) already contains 43,000 entries and this number is expected to increase dramatically in the near future.

Towards the identification of causal polymorphisms

To refine the regions that contains the functional variant, mapping experiments at very high resolution have to be carried out. The generation of genome-wide high-density SNP maps and the construction of a rat haplotype map may assist this process [31-35]. Haplotype blocks are sets of coupled polymorphisms that reflect a common ancestral origin and can be deduced from high-density SNP genotyping datasets. Although these blocks can potentially be hundreds of SNPs large, only a limited number of markers have to be genotyped to infer the complete block structure and genotype for every marker. In addition, for every polymorphism or haplotype block the distribution over a wide set of inbred strains can be determined (Strain Distribution Pattern) and correlated with the genotype distribution patterns. In theory, blocks of haplotype sharing are smaller than in RI lines or congenics, thereby potentially increasing resolution. Although haplotype-based mapping has been used successfully in mouse [32,34-37], it remains to be proven if the organization of the rat genome is comparable to the mouse and if the structure is useful for mapping. If haplotype blocks are large, the resolution of haplotype-based approaches will be poor, making all the variants in a block a potential causal polymorphism. If the blocks are small, the advantage of determining the complete block structure in many strains, followed by haplotype-based genotyping in the (sub-)strains of interest may be limited as compared to conventional approaches [31].

An alternative approach to get to candidate genes is the use of integrated genome-wide transcriptional profiling in combination with linkage analysis in RI strains. Taking advantage of already existing physiological QTLs (pQTL) and the newly identified expression level QTLs (eQTL), this approach yielded an impressive list of highly likely candidate genes for hypertension [38], although it does not directly identify causal polymorphisms. Although identification of the responsible gene is already an important finding, the final stage would be the identification and causal proof of the underlying polymorphisms (Fig. 1). Discriminating functional from non-functional polymorphisms based on sequence and gene information alone is usually not sufficient. Therefore, genetic evidence is necessary. Consequently, tools to manipulate the genome are needed. Although such tools are lagging far behind the mouse, recent developments, like transgenesis, RNAi-mediated gene silencing, knockout technology, and nuclear transfer, are promising developments that are expected to meet the emerging needs. This introduction will describe the current state-of-the-art of these recent advances and in their importance for rat genetic research.

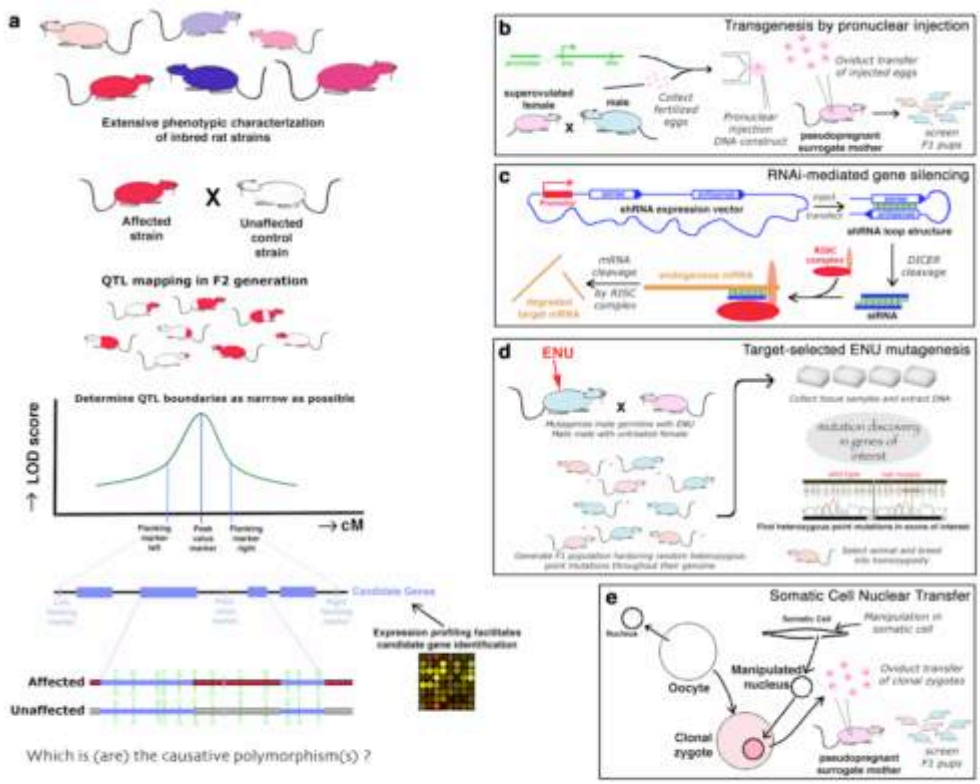


Figure 1: Candidate polymorphism identification and overview of genetic follow-up experiments.
a) Schematic representation of QTL-mapping and candidate gene identification. Extensive phenotypic characterization of inbred rat strains has identified many complex traits, included complex diseases in the rat. An affected strain is crossed with an unaffected control strain. The F1 offspring of such a cross will be heterozygous at all chromosomal positions. The gamete stem cells of the F1 animals will undergo meiosis, and the resulting haploid gametes will carry chromosomes that are a mosaic of the genotypes in the two parents. The F2 progeny are genotyped using genetic markers that are polymorphic between the two parental strains and phenotyped. Through statistical techniques, a correlation is made between a particular genotype and the phenotype that is measured. QTL boundaries can be determined. There are tricks to narrow down QTL regions even more, like the use of congenic animals that have a gene or small region

introgressed from the one strain into the other. Once the size of the QTL is sufficiently reduced, candidate genes can be picked. Genome-wide expression profiling in relevant tissues may facilitate candidate gene selection. Finally, the affected and unaffected strains are sequenced for these genes or regions of interest, almost always resulting in a list of hundreds of polymorphisms. The next episode will comprise of genetic experiments that may proof the influence of certain polymorphisms to the complex trait. **b)** Transgenesis by pronuclear injection. A superovulated female is mated with an untreated male to yield a large amount of fertilized oocytes. A custom made DNA construct is injected into the pronucleus of fertilized oocytes that are subsequently placed back into the oviduct of a pseudopregnant foster mother. The resulting F1 pups are screened by Southern Blot analysis for presence of the transgene. **c)** RNA interference-mediated gene silencing. This is a tool to interfere with endogenous mRNA function. Naked small interfering RNAs (siRNAs) are injected locally or produced from a transgenic construct coding for a short hairpin RNA (shRNA) that is processed by DICER. Transgenic constructs can also be injected locally or delivered by lentivirusses. The siRNAs are unwinded and will bind the endogenous mRNA with sequence complementarity. Subsequently, the targeted mRNA is degraded by the RISC complex. **d)** Target-selected ENU mutagenesis. The male germ-line is mutagenized with ENU, a mutagenic agent that induces random point mutations in the sperm population. By mating mutagenized males with untreated females an F1 population is generated that harbors many independent random heterozygous point mutations in their genomes. DNA samples are isolated for every individual and these samples are screened in a high-throughput fashion for induced heterozygous point mutations in genes of interest. Interesting mutants are outcrossed once and bred to homozygosity. **e)** Somatic Cell Nuclear Transfer (SCNT). Cloning of an animal using SCNT can be accomplished by injecting a somatic donor nucleus into an enucleated oocyte. Clonal zygotes are placed back into oviducts of pseudopregnant foster mothers. By targeting interesting genes in the somatic donor nucleus (by homologous recombination), it is potentially possible to generate knockout rats.

Transgenesis

The most straightforward way to assay causality of a potential (partial) loss-of-function polymorphism is functional rescue of the phenotype by the introduction of a transgene expressing the wild type gene into the mutant background. Alternatively, gain-of-function polymorphisms can be addressed by overexpression in a wild type background. In exceptional cases, dominant negative constructs can be used to mimic loss-of-function in a wild type background. In all of these cases, proper temporal and spatial expression of the transgene could be an issue of concern.

Classical transgenesis via pronuclear microinjection of a DNA construct is already available from the early 1990s [39]. Once present in a fertilized oocyte, tandem copies of the DNA construct may integrate randomly at a single site in the genome. The method requires manual micromanipulations. Fertilized eggs are collected from a superovulated female. The transgenic construct is injected into the pronucleus using a glass micropipet. Injected embryos are placed back into the oviduct of a pseudopregnant surrogate mother by microsurgery (Fig. 1).

Alternatively, lentiviral vector-mediated transgenesis can be used [40], which has proven to be much more efficient and less labor-intensive, making the approach more cost effective [41]. Single-cell embryos can be infected with recombinant lentiviral vectors, carrying the transgene under control of a specific promoter. So far, perivitelline injections and coincubation with denuded embryos (without zona pellucidae) have been used as infection methods, the latter being the most favorable as it does not require any micromanipulation. Infected oocytes are placed into oviducts of pseudopregnant foster mothers. Subsequently, offspring carrying the integrated construct is selected by doing Southern Blot analysis and intercrossed to obtain F1 and F2 generation animals that stably express the transgene.

Finally, transgenic rats have also been generated by lentiviral transduction of male germline stem cells [42]. A population of cells enriched for SSCs (Spermatogonial Stem Cells) was transduced in culture on laminin matrix with lentiviral vectors and transplanted to a single testis of recipient

males. It was found that about 30% of the total amount of produced pups did contain the transgene.

The major complication of transgenesis procedures is associated with random integration. Depending on the genomic locus, a transgene can be highly transcribed, partially or fully silenced, or it may influence the expression of neighboring genes. To account for these potential effects, multiple independent lines need to be generated and functionally analyzed. A clear advantage of transgenic methods is that, on a researcher's discretion, a transgene can be expressed in a tissue-specific or time-specific manner, by using tissue-specific or stage-specific promoters, respectively, and in combination with local injections [43]. Conditional transgenics, a tool to turn on and off genes, are used in mice and may also be implemented in the rat [44]. In addition, transgenesis by microinjection is not restricted to certain strains, although higher efficiencies of reimplantation have been found in outbred rat strains [45].

Transgenesis has predominantly been used to overexpress gene constructs to generate rat models that mimic clinical features of human disease, like hypertension [46], Alzheimer's Disease (AD) [47], Huntington Disease (HD) [48], but also HIV-1-related immunodeficiency [49], and many more. In many cases, transgenic rat models were found to relate closer to human disease phenotypes than transgenic mice in which similar genetic constructs were inserted [50]. Recently, transgenic rats have been used to perform complementation tests for QTL candidate genes *Cd36*, located in an interval associated with insulin resistance in the inbred spontaneous hypertensive rat (SHR) strain [51], and *Cblb* in a QTL associated with autoimmune disease type 1 diabetes mellitus (insulin-dependent diabetes mellitus, IDDM) in the Komeda diabetes-prone (KDP) strain [52]. In both cases, the causative mutation appeared to be nonsense, suggesting strong reduction or complete loss of gene function, and could be functionally complemented with the wild type copy of the gene. However, many QTL polymorphisms are likely to be much more subtle. For example, missense mutations and promoter polymorphisms will require a more delicate approach for proving causality. In the mouse, quantitative complementation or QTL-knockout interaction tests have been used successfully for testing the candidacy of a gene at a QTL [5]. This method, however, requires an endogenous null allele of the gene of interest, which only recently became technically possible for the rat by the development of knockout technology (see below).

Knockdown by siRNA-mediated gene silencing

A second approach to prove a gene's contribution to a QTL is by knockdown in a wild type background. Gene knockdowns *in vivo* can be achieved by RNAi (RNA interference)-mediated gene silencing [53,54]. RNAi is a cell-surveillance mechanism that destroys all cytoplasmic mRNAs with sequence complementation to the dsRNA (double-stranded RNA) trigger [55]. The technology is complementary to the traditional approaches of permanent gene knockouts and transgenics, as it can provide a temporal and/or tissue specific gene knockdown. The major advantage is that it is relatively time- and cost-effective. A drawback of the method is that it may not work equally efficient for all mRNAs in all tissues and, depending on its administration, its effect can be transient.

Antisense technology has already been used for several decades to silence endogenous genes. For instance, in a brain-specific angiotensin knockdown study in rats it has been reported that more than 90% of silencing was obtained using antisense RNA of 200 bp [56]. However, since dsRNAs longer than 30 nucleotides can induce an undesirable IFN (Interferon) response resulting in cell-death [57,58], either short synthetic RNA molecules of 21-23 nucleotides or a fold-back, stem-loop structure of approximately 19 perfectly matched nucleotides connected by various spacer regions and ending in a 2-nucleotide 3'-overhang (shRNAs; short-hairpin RNAs) can be used to efficiently induce RNA interference [54]. Nowadays, dsRNAs or siRNAs can be supplied directly to the target tissue or shRNA molecules can be produced from a transgenic DNA expression

construct to induce RNAi in a tissue- or time-specific fashion. Shortly, an enzyme called DICER processes long dsRNA or shRNAs into the previously described siRNA duplexes (Fig 1). These small molecules unwind and bind the endogenous mRNA, which results in assembly of the RISC complex. These complexes can direct RNA cleavage, mediate translational repression or induce chromatin modification, all resulting in gene silencing [54].

There are two major points of concern using this technology *in vivo*, namely avoiding off-target effects and efficient delivery. First, the risk of side effects is reduced by the use of a cell's own machinery for directing sequence-specific silencing, in contrast to the mode of action of other antisense technologies. In addition, a proper control for specificity would be the use of several siRNAs or shRNAs (directed to the same mRNA) as dsRNA triggers. If these result in the same phenotypic outcome, one can be almost sure that the effect is specific, since it is unlikely that completely different siRNAs/shRNAs will give similar side effects.

Second, efficient delivery of the dsRNA trigger to target tissue *in vivo* remains a serious obstacle. Various delivery systems for transient silencing have been tested in mice, such as infusion into the brain [59], electroporation into the retina [60], intravenous or local injection [61,62], and more. Recent studies, also in mice, have shown that lentiviral vectors can be used to transfect adult brains and accomplish stable RNAi, although only specific brain regions were efficiently targeted by this approach [63-66].

Long-term gene silencing has been demonstrated *in vivo* in mice using genetic mosaics, constructed by stem cell engineering with shRNA constructs and subsequent repopulation of the target tissue with these stem cells [67,68]. Germline-modified animals also exhibit long-term RNAi. Strains of mice have been constructed by standard pronuclear injection, engineered ES-cell chimaeras, or by subzonal injection of fertilized eggs with recombinant lentivirusses, to heritably suppress a target gene by shRNA expression [69-72]. These long-term silencing approaches may work the best, however, due to preceding engineering steps, the advantage of rapidity of the RNAi-mediated knockdown approach in general may be limited.

Although there are no technical limitations for the implementation of the above-mentioned techniques for the rat, considerably less reports on *in vivo* RNAi in the rat have been published so far. Transient delivery by local injection [73-75] or infusion [76,77] of naked siRNA, sometimes followed by electroporation [78] have been used successfully. In addition, viral-mediated long-term siRNA production has been achieved by local injection [79,80] and for generating transgenic animals from transfected preimplantation embryos [70]. All this work, as well as the simultaneous production of lentiviral transgenic mice and rats [40], illustrates that ES-cell-independent technology is easily transferable from the mouse into the rat model system.

Genetic knockouts by ENU-mutagenesis

The generation of stable knockouts by homologous recombination in embryonic stem (ES) cells, as is common for the mouse [81], is not available for most organisms, including the rat, due to the lack of pluripotent ES cells. An alternative approach to make genetic knockouts in such cases, has been developed for various model organisms, like *C. elegans* [82], *Drosophila* [83,84], zebrafish [85], *Arabidopsis* [86], maize [87], and *Lotus* [88]. This method, known as target-selected mutagenesis or TILLING (Targeting Induced Local Lesions in Genomes), was recently also successfully established for the rat [89,90]. In the rat, the approach starts with random mutagenesis of the male germline by intraperitoneal injection of the mutagenic agent ENU (*N*-ethyl-*N*-nitrosourea). ENU introduces random point mutations, primarily in SSCs. These mutations become fixed in the sperm cells during spermatogenesis. Mutagenized males are mated with untreated females to generate a large population of F1 animals that harbor many random heterozygous point mutations in their genome (Fig 1). Next, DNA samples are extracted from each F1 individual,

which are subsequently screened for induced mutations in exons of interest. These exonic mutations could be silent, change an amino acid, or induce a premature stop codon.

The efficacy of the procedure is largely dependent on two factors: the induced mutation frequency and the efficiency and throughput of the mutation screening methodology. The mutagenesis efficiency for the rat was found to be strain and dose dependent [89-91]. Outbred strains such as Sprague Dawley and Wistar tend to give the best mutation frequency and have superior reproduction characteristics as compared to inbred strains, making them the most suited strains for ENU mutagenesis-driven knockout technology. However, in many cases, the knockout allele is needed in a different genetic background, since QTL characteristics are associated with specific strains. This would require extensive backcrossing. Regardless, outcrossing is necessary, as mutant animals resulting from this approach, harbor additional induced mutations randomly across their genome.

Although these additional mutations seem a disadvantage, the advantage is that mutant collections can be used to screen for mutations and knockouts in multiple genes in parallel. The most commonly used methods for mutation discovery are Denaturing High-Performance Liquid Chromatography [86,92,93], CEL I-mediated heteroduplex cleavage [88,89,94,95], and high-throughput resequencing [85,91,96,97]. The first rat knockouts, however, were produced using a yeast-based screening assay that only identifies truncated mRNAs, thereby focusing on highly likely deleterious mutations, but neglecting potentially interesting amino acid substitutions [90]. We have used a cost-efficient dideoxy resequencing approach and identified more than 120 ENU-induced mutations in the rat, including 56 missense and six nonsense mutations [91]. The molecular mutation frequency deduced from these experiments turns out to be about 1 per 1.2 x 10^6 bp, which is about 4-fold lower than for zebrafish [95], but similar as obtained for the mouse [92,93,96,97]. In total, the efforts in rats resulted in eight knockouts, 68 missense mutations, and a number of uninteresting silent and non-coding mutations.

Once mutations are identified, interesting mutant animals are selected and outcrossed on wild-type background to cross out other induced mutations. Outcrossing may not eliminate all other induced mutations, in particular, closely linked mutations. However, with the current mutation frequency of 1 in 1.2 million bp, the chance that an adjacent mutation ends up in and is deleterious to a nearby gene is close to zero. There are several procedures though, to proof the causality of a mutation to a phenotype. First, genotyping in backcross progeny should result in full linkage between the mutation and the phenotype. Second, a second loss-of-function or reduction-of-function of the gene should result in a similar phenotype, and can be used to create heteroallelic animals that also must have the same phenotype. However, identification of two independent null alleles in a single gene will be problematic in the rat due to the current mutation rates and limited mutant animal library sizes. Finally, rescue of the phenotype by introduction of a wild-type copy of the gene using transgenesis or phenocopying by gene-specific RNAi will firmly establish the phenotype-genotype relationship.

Although already proven successful for the rat, the target-selected mutagenesis approach needs further optimization to become a routine method for knockout production. Either mutation rates and/or throughput of the mutation discovery step should be increased. Both the CEL I and resequencing approaches are very well suited for automation and high throughput processing, but emerging mutation detection and discovery technologies, like microarray-based [98] and massively parallel sequencing approaches [99,100] are promising alternatives.

As considerable amounts of mutagenized rats, as well as expensive infrastructure, are needed, the target-selected ENU mutagenesis screens are performed in only a few labs worldwide [89,90] (http://pga.mcw.edu; Ingenium: http://www.ingenium-ag.com), and knockouts are still produced at a relatively low rate. An additional advantage of the ENU-based approaches is that the resulting animals can also be used for systematic forward genetic screens, similar to what has been done for the mouse [101,102]. Proof of principle for this approach in the rat was demonstrated by the identification of phenotypic mutants in ENU-mutagenized progeny [89,90]

and the recent positional cloning of a nonsense mutation in *myosin7a* in the ENU-induced tornado mutant [103]. This resulted in a unique rat model for the Usher Syndrome Type 1b. Phenotype-based screens in the rat may have good perspectives as years of phenotypic characterization of rat models has not only resulted in the identification of large numbers of clinically relevant phenotypes but also in sensitive phenotyping methods and assays. When combined with large libraries of mutant animals, systematic phenotypic screening may result in the identification of novel genetic factors that contribute to human disease and are difficult to recognize in other model organisms.

Gene targeting by homologous recombination

Traditional knockouts using homologous recombination in pluripotent embryonic stem cells, analogous to the mouse, is not possible in the rat. Despite the quest that is continuously ongoing, one has not succeeded to isolate such cells over the last decades. Cultured rat blastocysts and derivative cell lines rapidly lose pluripotency due to loss of Oct-4 transcription factor [104]. Cloning, the asexual reproduction of individuals via somatic cell nuclear transfer (SCNT), has already been accomplished for many species, like sheep, cow, pig, mouse, and recently also for the rat [105]. Practically, the process of nuclear transfer consists of four steps (Fig. 1). First, the oocyte's own haploid nucleus is removed physically or chemically. The oocyte has to endure *in vitro* culturing. Second, a diploid nucleus from a somatic donor cell is collected, which is usually a nucleus from a fetal fibroblast-derived cell line. Third, this donor nucleus is microinjected in or electro-fused with the enucleated oocyte and finally, the oocyte is placed back in a pseudopregnant foster mother [106]. All of these procedures demand a lot of endurance from the cell, which makes the method very inefficient. In theory, this procedure should allow the generation of mutants by homologous recombination in somatic cells followed by nuclear transfer in oocytes, similar to what has been done for pig [107,108]. This approach may have major advantages, since the strain of the donor nucleus essentially determines the strain of the cloned animal. Unlike potential ES-cell technology and ENU mutagenesis, which are strongly strain-dependent, this technology may produce knockouts, knockins, conditional knockouts, etc. for any strain of interest with similar efforts, without having to incross the mutation or transgene into the desired genetic background. However, homologous recombination in rat somatic cells and the use of these manipulated cells as the somatic donor nucleus needs to be demonstrated before this technique can be implemented for the production of gene knockouts [106]. In addition, vast optimization and improvements on the efficiency of the SCNT procedure are required, as the current efficiency of only two fertile progeny after implantation of 129 cloned embryos is far too low for routine application.

Recently, an alternative approach, circumventing the need for the laborious and inefficient SCNT, has come in sight by using rat SSCs. These cells can be cultured for up to seven months and after transplantation to recipient rats, donor stem cell-derived progeny can be obtained [109]. When homologous recombination can be achieved in these cultured SSCs, this approach would allow for the generation of knockout rats. However, these stem cells are relatively slow growing (doubling every five to six days) and selection for the occurrence of homologous recombination may have a devastating effect on the survival rate of the SSCs.

Future developments

Rat biomedical research has moved over the years from phenotypic characterization to the identification of genetic traits by means of QTL mapping, and is currently arriving at the next stage of coupling functional genetic components to the observed pathological states. Further

expansion of the genetic toolbox for the rat and improvement of newly developed methods will be important to succeed at this step.

The establishment of knockout technology for the rat now allows for novel genetic approaches, like quantitative complementation or QTL-knockout interaction for testing the candidacy of a gene at a QTL [5], conditional transgene expression in a null background [44], or sensitized genetic screens in knockout background [110]. This in combination with lentiviral delivery of transgenes and siRNAs/shRNAs for gene knockdowns, and the prospective availability of homologous recombination in the near future, we can state that the genetic toolbox for interference with endogenous gene expression in a targeted fashion is rapidly filling up.

Over the next years, we expect to witness the identification of novel genes and polymorphisms underlying complex human disease, such as hypertension and depression. Once causally verified, these genes become potential drug targets providing novel prospects for treatment. The elegant advantage of the rat model organism approach is that a vertebrate animal model is already available for initial drug screening, testing, and validation.

References

1. Kolb, B. (2005) Neurological models. In *The behavior of the laboratory rat.* (Kolb, B. and Whishaw, I.Q., eds), pp. 449-461, Oxford University Press.
2. Gill, T.J., 3rd *et al.* (1989) The rat as an experimental animal. *Science*, 245, 269-276.
3. Greenhouse, D.D. *et al.* (1990) Catalogue of Inbred Strains of Rats. In *Genetic monitoring of inbred strains of rats* (Hedrich, H.J., ed), pp. 410-480, Gustav Fisher.
4. Lindsey, J.R. (1979) Historical Foundations. In *The laboratory rat* (1st edn) (Baker, H.J., Lindsey, J.R., and Weisbroth, S.H., eds), pp. 1-36, Academic Press.
5. Flint, J. *et al.* (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet*, 6, 271-286.
6. Gibbs, R.A. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428, 493-521.
7. Lindblad-Toh, K. (2004) Genome sequencing: three's company. *Nature*, 428, 475-476.
8. Twigger, S. *et al.* (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res*, 30, 125-128.
9. Jacob, H.J. *et al.* (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell*, 67, 213-224.
10. Paterson, A.H. *et al.* (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, 335, 721-726.
11. Cowley, A.W., Jr. *et al.* (2004) Application of chromosomal substitution techniques in gene-function discovery. *J Physiol*, 554, 46-55.
12. Jacob, H.J. and Kwitek, A.E. (2002) Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet*, 3, 33-42.
13. Lande, R. and Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124, 743-756.
14. Pravenec, M. *et al.* (1995) Mapping of quantitative trait loci for blood pressure and cardiac mass in the rat by genome scanning of recombinant inbred strains. *J Clin Invest*, 96, 1973-1978.
15. Kwitek, A.E. *et al.* (2004) High-density rat radiation hybrid maps containing over 24,000 SSLPs, genes, and ESTs provide a direct link to the rat genome sequence. *Genome Res*, 14, 750-757.
16. Steen, R.G. *et al.* (1999) A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res*, 9, AP1-8, insert.

17. Watanabe, T.K. *et al.* (2000) Characterization of newly developed SSLP markers for the rat. *Mamm Genome*, 11, 300-305.
18. Moreno, C. *et al.* (2004) Genome-Wide Scanning With SSLPs in the Rat. *Methods Mol Med*, 108, 131-138.
19. Glazier, A.M. *et al.* (2002) Finding genes that underlie complex traits. *Science*, 298, 2345-2349.
20. Nadeau, J.H. *et al.* (2000) Analysing complex genetic traits with chromosome substitution strains. *Nat Genet*, 24, 221-225.
21. Darvasi, A. and Soller, M. (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, 141, 1199-1207.
22. Mott, R. *et al.* (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci U S A*, 97, 12649-12654.
23. Churchill, G.A. *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36, 1133-1137.
24. Darvasi, A. (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet*, 18, 19-24.
25. Shifman, S. and Darvasi, A. (2005) Mouse inbred strain sequence information and yin-yang crosses for quantitative trait locus fine mapping. *Genetics*, 169, 849-854.
26. Talbot, C.J. *et al.* (1999) High-resolution mapping of quantitative trait loci in outbred mice. *Nat Genet*, 21, 305-308.
27. Vignal, A. *et al.* (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol*, 34, 275-305.
28. Guryev, V. *et al.* (2004) Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res*, 14, 1438-1443.
29. Smits, B.M. *et al.* (2004) Genetic variation in coding regions between and within commonly used inbred rat strains. *Genome Res*, 14, 1285-1290.
30. Zimdahl, H. *et al.* (2004) A SNP map of the rat genome generated from cDNA sequences. *Science*, 303, 807.
31. Cuppen, E. (2005) Haplotype-based genetics in mice and rats. *Trends Genet*, 21, 318-322.
32. Grupe, A. *et al.* (2001) In silico mapping of complex disease-related traits in mice. *Science*, 292, 1915-1918.
33. Liao, G. *et al.* (2004) In silico genetics: identification of a functional element regulating H2-Ealpha gene expression. *Science*, 306, 690-695.
34. Pletcher, M.T. *et al.* (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol*, 2, e393.
35. Wang, X. *et al.* (2004) Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res*, 14, 1767-1772.
36. Manenti, G. *et al.* (2004) Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. *Oncogene*, 23, 4495-4504.
37. Park, Y.G. *et al.* (2003) Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res*, 13, 118-121.
38. Hubner, N. *et al.* (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*, 37, 243-253.
39. Mullins, L.J. *et al.* (2002) Transgenesis in the rat. *Methods Mol Biol*, 180, 255-270.
40. Lois, C. *et al.* (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science*, 295, 868-872.
41. Pfeifer, A. (2004) Lentiviral transgenesis. *Transgenic Res*, 13, 513-522.
42. Hamra, F.K. *et al.* (2002) Production of transgenic rats by lentiviral transduction of male germ-line stem cells. *Proc Natl Acad Sci U S A*, 99, 14931-14936.

43. Dittgen, T. *et al.* (2004) Lentivirus-based genetic manipulations of cortical neurons and their optical and electrophysiological monitoring in vivo. *Proc Natl Acad Sci U S A*, 101, 18206-18211.
44. Lewandoski, M. (2001) Conditional control of gene expression in the mouse. *Nat Rev Genet*, 2, 743-755.
45. Bolon, B. *et al.* (2000) *The laboratory Rat*. Academic Press, London.
46. Popova, E. *et al.* (2005) Production of transgenic models in hypertension. *Methods Mol Med*, 108, 33-50.
47. Shaughnessy, L. *et al.* (2004) Novel approaches to models of Alzheimer's disease pathology for drug screening and development. *J Mol Neurosci*, 24, 23-32.
48. von Horsten, S. *et al.* (2003) Transgenic rat model of Huntington's disease. *Hum Mol Genet*, 12, 617-624.
49. Reid, W. *et al.* (2001) An HIV-1 transgenic rat that develops HIV-related pathology and immunologic dysfunction. *Proc Natl Acad Sci U S A*, 98, 9271-9276.
50. Charreau, B. *et al.* (1996) Transgenesis in rats: technical aspects and models. *Transgenic Res*, 5, 223-234.
51. Pravenec, M. *et al.* (2001) Transgenic rescue of defective Cd36 ameliorates insulin resistance in spontaneously hypertensive rats. *Nat Genet*, 27, 156-158.
52. Yokoi, N. *et al.* (2002) Cblb is a major susceptibility gene for rat type 1 diabetes mellitus. *Nat Genet*, 31, 391-394.
53. Dorsett, Y. and Tuschl, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov*, 3, 318-329.
54. Hannon, G.J. and Rossi, J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, 431, 371-378.
55. Fire, A. *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391, 806-811.
56. Schinke, M. *et al.* (1999) Blood pressure reduction and diabetes insipidus in transgenic rats deficient in brain angiotensinogen. *Proc Natl Acad Sci U S A*, 96, 3975-3980.
57. Caplen, N.J. *et al.* (2001) Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc Natl Acad Sci U S A*, 98, 9742-9747.
58. Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411, 494-498.
59. Thakker, D.R. *et al.* (2004) Neurochemical and behavioral consequences of widespread gene knockdown in the adult mouse brain by using nonviral RNA interference. *Proc Natl Acad Sci U S A*, 101, 17270-17275.
60. Matsuda, T. and Cepko, C.L. (2004) Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci U S A*, 101, 16-22.
61. McCaffrey, A.P. *et al.* (2002) RNA interference in adult mice. *Nature*, 418, 38-39.
62. Layzer, J.M. *et al.* (2004) In vivo activity of nuclease-resistant siRNAs. *Rna*, 10, 766-771.
63. Hommel, J.D. *et al.* (2003) Local gene knockdown in the brain using viral-mediated RNA interference. *Nat Med*, 9, 1539-1544.
64. Van den Haute, C. *et al.* (2003) Lentiviral vector-mediated delivery of short hairpin RNA results in persistent knockdown of gene expression in mouse brain. *Hum Gene Ther*, 14, 1799-1807.
65. Xia, H. *et al.* (2004) RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nat Med*, 10, 816-820.
66. Xia, H. *et al.* (2002) siRNA-mediated gene silencing in vitro and in vivo. *Nat Biotechnol*, 20, 1006-1010.
67. Hemann, M.T. *et al.* (2003) An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nat Genet*, 33, 396-400.

68. Rubinson, D.A. *et al.* (2003) A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet*, 33, 401-406.
69. Carmell, M.A. *et al.* (2003) Germline transmission of RNAi in mice. *Nat Struct Biol*, 10, 91-92.
70. Hasuwa, H. *et al.* (2002) Small interfering RNA and gene silencing in transgenic mice and rats. *FEBS Lett*, 532, 227-230.
71. Kunath, T. *et al.* (2003) Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat Biotechnol*, 21, 559-561.
72. Tiscornia, G. *et al.* (2003) A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. *Proc Natl Acad Sci U S A*, 100, 1844-1848.
73. Bahi, A. *et al.* (2005) Silencing dopamine D3-receptors in the nucleus accumbens shell in vivo induces changes in cocaine-induced hyperlocomotion. *Eur J Neurosci*, 21, 3415-3426.
74. Lai, S.K. *et al.* (2004) Selective knockdown of gene expression of N-methyl-D-aspartate receptor one ameliorates parkinsonian motor symptom in 6-hydroxydopamine-lesioned rats. *Neurochem Int*, 45, 11-22.
75. Tan, P.H. *et al.* (2005) Gene knockdown with intrathecal siRNA of NMDA receptor NR2B subunit reduces formalin-induced nociception in the rat. *Gene Ther*, 12, 59-66.
76. Dorn, G. *et al.* (2004) siRNA relieves chronic neuropathic pain. *Nucleic Acids Res*, 32, e49.
77. Isacson, R. *et al.* (2003) Lack of efficacy of 'naked' small interfering RNA applied directly to rat brain. *Acta Physiol Scand*, 179, 173-177.
78. Takabatake, Y. *et al.* (2005) Exploring RNA interference as a therapeutic strategy for renal disease. *Gene Ther*, 12, 965-973.
79. Michel, U. *et al.* (2005) Long-term in vivo and in vitro AAV-2-mediated RNA interference in rat retinal ganglion cells and cultured primary neurons. *Biochem Biophys Res Commun*, 326, 307-312.
80. Zhang, Y. *et al.* (2003) In vivo knockdown of gene expression in brain cancer with intravenous RNAi in adult rats. *J Gene Med*, 5, 1039-1045.
81. Capecchi, M.R. (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet*, 6, 507-512.
82. Jansen, G. *et al.* (1997) Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat Genet*, 17, 119-121.
83. Bentley, A. *et al.* (2000) Targeted recovery of mutations in Drosophila. *Genetics*, 156, 1169-1173.
84. Winkler, S. *et al.* (2005) Target-selected mutant screen by TILLING in Drosophila. *Genome Res*, 15, 718-723.
85. Wienholds, E. *et al.* (2002) Target-selected inactivation of the zebrafish rag1 gene. *Science*, 297, 99-102.
86. McCallum, C.M. *et al.* (2000) Targeted screening for induced mutations. *Nat Biotechnol*, 18, 455-457.
87. Till, B.J. *et al.* (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol*, 4, 12.
88. Perry, J.A. *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.
89. Smits, B.M. *et al.* (2004) Target-selected mutagenesis of the rat. *Genomics*, 83, 332-334.
90. Zan, Y. *et al.* (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol*, 21, 645-651.
91. Smits, B.M. *et al.* (2005) Routine Generation of Gene Knockouts and Mutant Models in the Laboratory Rat. *Pharmacogenetics and genomics*.
92. Coghill, E.L. *et al.* (2002) A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet*, 30, 255-256.

93. Quwailid, M.M. *et al.* (2004) A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm Genome*, 15, 585-591.
94. Till, B.J. *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res*, 13, 524-530.
95. Wienholds, E. *et al.* (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res*, 13, 2700-2707.
96. Beier, D.R. (2000) Sequence-based analysis of mutagenized mice. *Mamm Genome*, 11, 594-597.
97. Concepcion, D. *et al.* (2004) Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. *Genetics*, 168, 953-959.
98. Syvanen, A.C. (2005) Toward genome-wide SNP genotyping. *Nat Genet*, 37 Suppl, S5-10.
99. Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376-380.
100. Shendure, J. *et al.* (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309, 1728-1732.
101. Hrabe de Angelis, M.H. *et al.* (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet*, 25, 444-447.
102. Nolan, P.M. *et al.* (2000) Implementation of a large-scale ENU mutagenesis program: towards increasing the mouse mutant resource. *Mamm Genome*, 11, 500-506.
103. Smits, B.M. *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.
104. Buehr, M. *et al.* (2003) Rapid loss of Oct-4 and pluripotency in cultured rodent blastocysts and derivative cell lines. *Biol Reprod*, 68, 222-229.
105. Zhou, Q. *et al.* (2003) Generation of fertile cloned rats by regulating oocyte activation. *Science*, 302, 1179.
106. Mullins, L.J. *et al.* (2004) Nuclear transfer in rodents. *J Physiol*, 554, 4-12.
107. Dai, Y. *et al.* (2002) Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. *Nat Biotechnol*, 20, 251-255.
108. Lai, L. *et al.* (2002) Production of alpha-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science*, 295, 1089-1092.
109. Ryu, B.Y. *et al.* (2005) Conservation of spermatogonial stem cell self-renewal signaling between mouse and rat. *Proc Natl Acad Sci U S A*.
110. Kile, B.T. and Hilton, D.J. (2005) The art and design of genetic screens: mouse. *Nat Rev Genet*, 6, 557-567.

CHAPTER 2

Target-selected Mutagenesis of the Rat

Adapted from Genomics (2004) ***83:*** *332–334*

Target-selected Mutagenesis of the Rat

Bart M. G. Smits, Josine B. Mudde, Ronald H. A. Plasterk, and Edwin Cuppen*

Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

*Corresponding Author

The rat is one of the most extensively studied model organisms and with its genome being sequenced, tools to manipulate gene-function *in vivo* become increasingly important. We here report proof of principle for target-selected mutagenesis as a reverse genetic or knockout approach for the rat.

For many decades, chemical mutagenesis has been used to induce novel phenotypes in forward genetic studies. However, chemical mutagenesis has also been proven effective for reverse genetic purposes: in *Caenorhabditis elegans* [1], *Arabidopsis* [2], mouse [3,4], and zebrafish [5]. We now have successfully applied this approach to the rat: chemical mutagenesis followed by gene-specific detection of mutations in the offspring and subsequent breeding of selected mutations to homozygosity (Fig. 1).

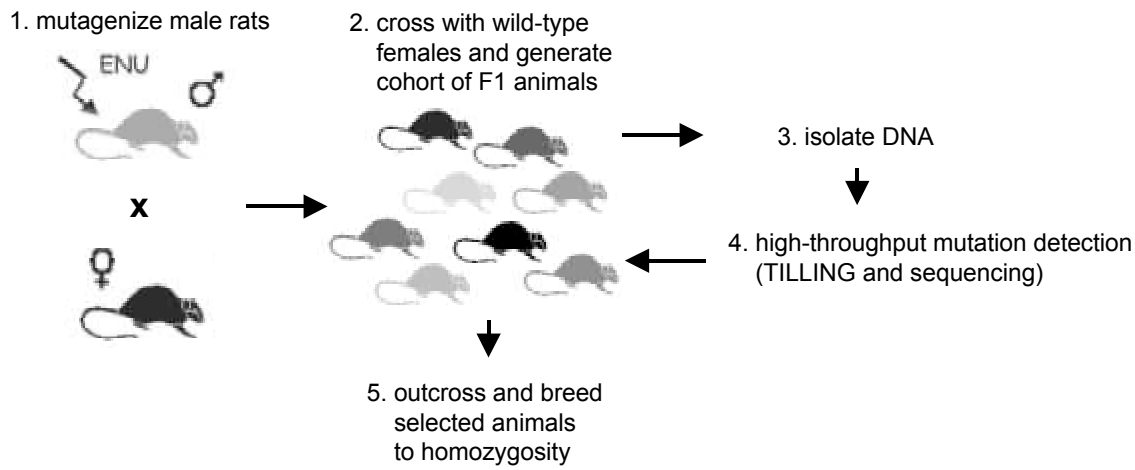


Figure 1: Schematic outline of the target-selected mutagenesis approach for reverse genetics in the rat. Male founder animals are mutagenized using ENU and crossed with wild-type females to generate F1 animals that are now heterozygous for a large number of random mutations. Mutations in genes of interest are detected using a high-throughput setup. Animals carrying mutations of interest are subsequently outcrossed and the mutation is crossed to homozygosity.

We first needed to establish an efficient mutagenesis protocol. We mutagenized male founder rats (Wistar) with different doses of the germ-line mutagen *N*-ethyl-*N*-nitrosourea (ENU) [6], which has been proven effective in diverse organisms [7]. Three weekly doses of 20, 40, and 60 mg/kg initially resulted in a decreased brood size. However, after a full cycle of spermatogenesis (about 9-10 weeks), the doses resulted in complete regain of fertility, partial sterility, and full sterility,

respectively (Table 1). Histological analysis of the testis of sterile animals showed massive degeneration of testicular structures and no signs of spermatogenesis (not shown). Fertile animals were crossed with wild-type females to generate a cohort of F1 animals that are expected to be heterozygous for many independent random point mutations. We set up a high-throughput mutation discovery pipeline, based on enzyme-mediated cleavage of heteroduplex DNA [8], similar to what is being used successfully in Arabidopsis for similar purposes [9], which allows the analysis of more than 2.5 x 10^6 base pairs per day (for experimental details see http://rat.niob.knaw.nl). Per F1 animal we scanned 100 genomic regions of about 600 bp that mainly contain protein-coding sequences, for the occurrence of induced mutations. In total, we analyzed 40 x 10^6 base pairs in 768 F1 animals and found 17 induced mutations, resulting in 3 silent, 1 intronic, 1 UTR, and 12 missense mutations. Mutation frequency per founder animal was found to be variable within a range from 0 to 1.7 x 10^{-6} (Table 1). Although these differences maybe within the statistical variation, we cannot exclude an effect of genetic variation in the outbred Wistar rat strain or differences in the effective ENU dose as a result of weight differences in the male founder rats caused by an increased percentage of fat tissue that does not take up ENU. The observed frequencies are in the same order of magnitude as found in large scale analyses for mouse (5.3 x 10^{-7}) [4] and zebrafish (2,2 x 10^{-6}) [5]. Eight of the animals carrying mutations were outcrossed and all mutations were found to be inherited to the next generation in a Mendelian way, showing that these are germ-line mutations. Crossing of the mutations to homozygosity and subsequent detailed analysis are needed to analyze the phenotype of these mutants. Currently, we have outcrossed and bred to homozygosity a mutation (A to T at position 975 in GenBank entry L02842) in FSHR (follicle-stimulating hormone receptor). Interestingly, although this missense mutation (D to V) does not obviously result in a complete loss of function, the phenotype of the mutant resembles the FSHR null mutation in mice [10]. Heterozygous females show a reduced fertility (one animal is completely sterile and three females have an average litter size of 4.3 pups (n = 7, standard deviation 1.7) compared to 10.3 pups (n = 6, standard deviation 2.9) for control breedings). Although more detailed analysis will be necessary to define fully the phenotype pf this mutant, these preliminary observations suggest that our approach has resulted in the identification of a rat knockout.

Table 1: ENU-induced mutations in the rat

founder animal*	ENU dose (mg/kg)	number of F1 animals screened	number of basepairs screened	number of mutations found	mutation frequency
1	3 x 40	Sterile	-	-	-
2	3 x 40	Sterile	-	-	-
3	3 x 40	Sterile	-	-	-
4	3 x 40	Sterile	-	-	-
5	3 x 40	Sterile	-	-	-
6	3 x 40	116	5.974.161	4	6,7 x 10^{-07}
7	3 x 40	Sterile	-	-	-
8	3 x 20	83	4.274.615	1	2,3 x 10^{-07}
9	3 x 20	89	4.583.623	0	-
10	3 x 20	68	3.502.094	1	2,9 x 10^{-07}
11	3 x 20	54	2.781.074	0	-
12	3 x 20	60	3.090.083	1	3,2 x 10^{-07}
13	3 x 20	79	4.068.609	0	-
14	3 x 20	77	3.965.606	1	2,5 x 10^{-07}
15	3 x 20	58	2.987.080	5	1,7 x 10^{-06}
16	3 x 20	84	4.326.116	4	9,2 x 10^{-07}

* Results for injection with 3 x 60 mg/kg are not included due to complete sterility

Previous studies [3,9,11] have shown that in the order of 1 in 10 ENU-induced nucleotide changes result in reduction of gene function. Obviously, for reverse genetic approaches stop mutations are preferred. Screening of a larger number of F1 animals, improvement of mutagenesis by testing other (inbred) strains should result in a robust method for generating knockouts and mutants in the rat. Although mutant F1 rats can be kept as a living resource, ultimately the establishment of a reliable method for cryopreservation of rat sperm and rederivation of live offspring using *in vitro* fertilization or alternative technology will be essential for creating a permanent resource. Although such technology is now becoming standard for mouse, no well-established protocol is currently available for the rat.

Taken together these results, we show that it is possible to generate rats containing mutations in genes of interest, providing an approach to produce 'knockouts' in the rat. Recently, Gould and co-workers independently established an ENU mutagenesis protocol using three other rat strains and, using a yeast-based screening assay, isolated two different knockout rats [12]. Compared to their screening approach, the TILLING method has the advantage that not only premature stop mutants can be retrieved, but also loss-of-function mutants as a result of missense mutations (see above) and potentially also hypomorphic, antimorphic and neomorphic alleles. Furthermore, our screening method can be fully automated to very high throughput. This allows the screening of relatively large numbers of animals within a short time for many genes/amplicons at much lower costs. Currently, we can screen about 5,000 animals for mutations in a specific amplicon (minimal size 150 bp, maximum size 1000 bp) per day, costing about 500 Euro for consumables.

Finally, in line with the observations of Gould and colleagues [12], our results indicate that it is also possible to use ENU as a mutagen for (large-scale) forward genetic screens in the rat, similar to what has been done for the mouse [13,14] to study for example complex behavioral traits. Interestingly, we have found two obvious phenotypic mutants in the F1 population (total 768 animals), a hip dysplasia and a diphallus.

Acknowledgements

This work was financially supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics and NV Organon. We thank Steve Henikoff for help setting up the TILLING technology.

References

1. Jansen, G. *et al.* (1997) Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat Genet*, 17, 119-121.
2. McCallum, C.M. *et al.* (2000) Targeted screening for induced mutations. *Nat Biotechnol*, 18, 455-457.
3. Beier, D.R. (2000) Sequence-based analysis of mutagenized mice. *Mamm Genome*, 11, 594-597.
4. Coghill, E.L. *et al.* (2002) A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet*, 30, 255-256.
5. Wienholds, E. *et al.* (2002) Target-selected inactivation of the zebrafish rag1 gene. *Science*, 297, 99-102.
6. Russell, W.L. *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.
7. Noveroske, J.K. *et al.* (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome*, 11, 478-483.

8. Oleykowski, C.A. *et al.* (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, 26, 4597-4602.
9. Till, B.J. *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res*, 13, 524-530.
10. Danilovich, N. and Sairam, M.R. (2002) Haploinsufficiency of the follicle-stimulating hormone receptor accelerates oocyte loss inducing early reproductive senescence and biological aging in mice. *Biol Reprod*, 67, 361-369.
11. Wienholds, E. *et al.* (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res*, 13, 2700-2707.
12. Zan, Y. *et al.* (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol*, 21, 645-651.
13. Hrabe de Angelis, M.H. *et al.* (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet*, 25, 444-447.
14. Nolan, P.M. *et al.* (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet*, 25, 440-443.

CHAPTER 3

Generation of Gene Knockouts and Mutant Models in the Laboratory Rat by ENU-driven Target-selected Mutagenesis

Adapted from Pharmacogenetics and Genomics (2005)

Generation of Gene Knockouts and Mutant Models in the Laboratory Rat by ENU-driven Target-selected Mutagenesis

Bart M. G. Smits[1], Josine B. Mudde[1], Jose van de Belt[1], Mark Verheul[1], Jocelien Olivier[2], Judith Homberg[1], Victor Guryev[1], Alexander R. Cools[2], Bart A. Ellenbroek[2], Ronald H. A. Plasterk[1], and Edwin Cuppen[1]*

[1]Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands
[2]Department of Psychoneuropharmacology, Radboud University, P.O. Box 9101, 6500 HC Nijmegen, The Netherlands

*Corresponding Author

The rat is one of the most important model organisms for biomedical and pharmacological research. However, the generation of novel models for studying specific aspects of human diseases largely depends on selection for specific traits using existing rat strains, thereby solely depending on natural occurring variation. This study aims to provide the tools to manipulate the rat genome in a more directed way. We developed robust, automated, and scaleable reverse genetic methodology based on ENU (*N*-ethyl-*N*-nitrosourea)-based target-selected mutagenesis. Optimal mutagenesis conditions have been determined in three different rat strains and a universal, rapid, and cost-effective dideoxy resequencing-based screening setup was established for discovery of induced mutations in genes of interest. The effectiveness of the approach was illustrated by the identification of 120 induced mutations in a set of genes of interest, including six that result in unique rat knockout models due to the introduction of premature stopcodons. In addition, 56 mutations were found that change amino acids, including critical residues in transmembrane domains of receptors and channels. The approach described here allows for the systematic generation of knockout and protein function altering alleles in the rat. The resulting induced rat models will be powerful novel tools for studying many aspects of a wide variety of human diseases.

Introduction

The laboratory rat is an important experimental animal model in human medical research, including neural regeneration, psychiatric disorders, behavioral intervention, and addiction, as well as in drug development [1,2]. To mimic neurological disorders in rats, researchers primarily depend on induction by pharmacological agents or surgery [3]. Current genetic rat models for specific biomedical relevant traits originate from selective inbreeding programs. Although hundreds of useful rat models for many aspects of a variety of human disease have been identified, for most of them the underlying gene defect or polymorphism is still unknown. In addition, the existence of a model for a specific trait depends strongly on the degree of natural occurring variation in the strains

that are used. In the mouse, this limitation was addressed by chemical mutagenesis using the supermutagen *N*-ethyl-*N*-nitrosourea (ENU) [4]. Many years of exploration on efficiency of ENU as a mutagen have preceded the initiation of systematic large-scale phenotype-driven mouse ENU-mutagenesis screens [5,6] that have now resulted in a wide variety of valuable novel mouse models [7].

An alternative, gene-driven, approach to study gene function and generate disease models is by modifying the function of a gene in a targeted way, i.e. by overexpression or inactivation (knockout). Although current gene-modification and knockout tools for the mouse are extremely versatile and flexible [7], they do depend on homologous recombination in special, pluripotent embryonic stem cells, which, despite many efforts, have not been identified for most of the commonly used model organisms, including the rat. Interestingly, chemical mutagenesis has been used successfully to overcome this limitation in various organisms. ENU or EMS (ethylmethane sulfonate) mutagenesis is used to introduce random mutations, followed by a gene-based strategy for the identification of inactivating mutations, e.g. using CEL I-based enzymatic heteroduplex cleavage [8], DHPLC (Denaturing High-Performance Liquid Chromatography)-based heteroduplex detection [9,10], or by dideoxy resequencing [11,12]. These approaches have been proven successful for a variety of species, including *Caenorhabditis elegans* [13], *Arabidopsis* [14], zebrafish [15], mouse [10], maize [16], *Lotus* [17], and most recently for the rat [18,19]. However, the efficiency of this approach primarily depends on the effectiveness of the mutagenesis. For mouse, the *N*-ethyl-*N*-nitrosourea (ENU) mutation frequency was found to be strongly dose and strain dependent [20] and estimated to be roughly one mutation per one million basepairs, based on the analysis of 38 x 10^6 bp by DHPLC [10,21], and 10 x 10^6 bp by resequencing [12,22]. For the rat, only limited information on the molecular mutation frequency is available. Gould and coworkers reported the first ENU-induced rat knockouts [18], but an estimation of the molecular mutation frequency and the associated effectiveness of target-selected mutagenesis is difficult to deduce since the mutation discovery strategy, a yeast-based truncation assay, only detects nonsense mutations. However, the rate of inherited phenotypic mutants among the F1 animals was found to be strain and dose dependent and suggested a similar mutation frequency as for the mouse. Indeed, in a small-scale study, we independently provided a proof of principle for ENU mutagenesis-based reverse genetics in the rat and showed that the molecular mutation frequency varied between 1 mutation per 0.6 and 4 x 10^6 bp, for a single rat strain tested [19].

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

Methods

Animals and ENU mutagenesis protocol

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

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

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

Tissue sampling and genomic DNA isolation

From 2639 F1 progeny, a tail clip was fetched under isoflurane anesthetics at three to four weeks of age. Tail clips were sampled in a 96 deep well block (2.5 ml Riplate, Ritter) and lysed overnight in 400 µl of lysisbuffer, containing 100 mM Tris (pH 8.5), 200 mM of NaCl, 0.2% of sodium dodecyl sulfate (SDS), 5 mM of ethylene diamine tetraacetic acid (EDTA), and 100 µg/ml of freshly added Proteinase K at 55°C. Tissue debris was spun down for 20 min at 6,000 *g* and supernatant was transferred to a fresh block. DNA was precipitated by adding an equal volume of isopropanol, mixing and centrifugation for 20 min, 6,000 *g* at room temperature. The supernatant was removed by gently inverting the block and the pellets were washed with 70% ethanol and dissolved in 400 µl water. For PCR, 5 µl of a 50 times dilution of the DNA stock was used as template.

Nested PCR conditions

The first PCR was carried out using a touchdown thermocycling program (92°C for 60 sec; 12 cycles of 92°C for 20 sec, 65°C for 20 sec with a decrement of 0.6°C per cycle, 72°C for 30 sec; followed by 20 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 30 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems). The multiplex PCR reaction using 8 primer combinations, contained 5 µl genomic DNA, 0.2 µM of each forward primer and 0.2 µM of each reverse primer, 400 µM of each dNTP, 25 mM tricine, 7.0% glycerol (w/v), 1.6% dimethyl sulfoxide (DMSO) (w/v), 2 mM $MgCl_2$, 85 mM ammonium acetate pH 8.7 and 0.2 U Taq Polymerase in a total volume of 10 µl. After thermocycling, the PCR1 reactions were diluted with 25 µl water, mixed by pipetting, and 1 µl was used as template for the second round of PCR. The second PCR was done using a standard thermocycling program (92°C for 60 sec; 30 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 30 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems). PCR2 mixes contained 1 µl diluted PCR1 template, 0.1 µM forward primer, 0.1 µM reverse primer, 100 µM of each dNTP, 25 mM tricine, 7.0% glycerol (w/v), 1.6% (DMSO) (w/v), 2 mM $MgCl_2$, 85 mM ammonium acetate pH 8.7 and 0.1 U Taq Polymerase in a total volume of 5 µl. Several samples of each amplicon were tested on a 1% agarose gel containing ethidium bromide for the presence of the proper PCR fragment. The sequence of the universal tails and M13 oligonucleotides are, M13F: TGTAAAACGACGGCCAGT, M13R: AGGAAACAGCTATGACCAT.

Sequencing reactions, purification, and analysis

PCR2 products were diluted with 20 µl water and 1 µl was directly used as template for the sequencing reactions. Sequencing reactions, containing 0.12 µl BigDYE (v3.1; Applied Biosystems,), 1.88 µl 2.5 times dilution buffer (Applied Biosystems) and 0.4 µM universal M13 primer in a total volume of 5 µl, were performed using cycling conditions recommended by the manufacturer (40 cycles of 92°C for 10 sec, 50°C for 5 sec and 60°C for 120 sec). Sequencing products were purified by ethanol precipitation in the presence of 40 mM sodium-acetate and analyzed on a 96-capillary 3730XL DNA analyzer (Applied Biosystems), using the

standard RapidSeq protocol on 36 cm array. Sequences were analyzed for presence of heterozygous mutations using PolyPhred [23] and manual inspection of the mutated positions.

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

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

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

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

Results and Discussion

Rat ENU mutagenesis

Efficiency of the target-selected mutagenesis approach strongly depends on the mutation rate. In mouse, the ENU-induced acute toxicity, long-term sterility, and mutation frequency was found to be strain and dose dependent. In addition, a regime of three weekly low-dose injections turned out to be more effective compared to a single high-dose injection [20]. Our previous experiments showed that doses of 3 x 60 mg/kg resulted in complete sterility for all strains tested (not shown) and a small-scale study indicated that the optimal dose for the Wistar strain would be between 3 x 30 and 40 mg/kg [19]. In line with these observations, we now mutagenized males from three commonly used inbred strains, Brown Norway (BN), Lewis (LEW), and Fisher (F344) with 3 x 20, 30, and 40 mg ENU/kg, as well as one outbred strain, Wistar (WI) with 3 x 30, 35, and 40 mg ENU/kg. Remarkably, all three doses caused permanent sterility in LEW, and this strain was therefore excluded from further analysis. For the other strains, fertility was measured by up to four matings during 12 weeks (WI) or 13 weeks (BN, F344) following the last ENU injections (see

Supporting Table S3). For BN, all animals in the 3 x 40 mg/kg group (n = 26) were sterile and 13 and 7 animals from the 3 x 20 and 3 x 30 mg/kg group (n = 24 for each group), respectively, were found to be fertile (Table 1 and Supporting Table S1). For F344 and WI all experimental groups contained fertile animals: 16, 12, and 9 fertile males for F344 groups 3 x 20 (n = 20), 30 (n = 18), and 40 (n = 18), respectively, and 8, 10, and 6 for WI groups 3 x 30, 35, and 40 (n = 10 for each group), respectively (Tables 1 and Supporting Table S1). Males from the two highest dosed fertile groups for each strain were selected for further breeding. All early progeny was discarded, because these animals could be chimaeric resulting from ENU-induced DNA adducts in spermatozoa that are repaired in (part of) the fertilized oocyte. Since we are only interested in the genetically fixed mutations resulting from mutagenized spermatogonial stem cells, we started collecting progeny for mutation analysis at least a full round of spermatogenesis (60 days) after the last ENU treatment. A continuous breeding program, where females were only removed when they were pregnant, was set up for up to one year after ENU mutagenesis for F344 and BN, and eight months for WI, to produce F1 offspring carrying random heterozygous mutations.

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

Table 1: Statistics of ENU mutagenesis and subsequent progeny production

Strain/Dose*	BN 3x20	BN 3x30	BN 3x40	F344 3x20	F344 3x30	F344 3x40	F344 3x60**	LEW 3x20	WI 3x30	WI 3x35	WI 3x40	WI 3x60**	Total
# Males injected	24	24	26	20	18	18	5	18	10	10	10	10	193
# Fertile males	13	7	0	16	12	9	0	0	8	10	6	0	81
# Pups for screening†	187	119	0	n.d.	742	472	0	0	n.d	757‡	362‡	0	2639
Average Pups/male	14 ± 7	17 ± 6			74 ± 18	43 ± 21				76 ± 23	60 ± 23		
# Bases screened	9.01×10^6	5.83×10^6			41.06×10^6	24.64×10^6				76.41×10^6	37.27×10^6		194.3×10^6
# Mutations													
- nonsense	1	0			0	0				5	0		6
- missense	3	0			5	10				21	17		56
- silent	2	1			3	1				11	7		24
- non-coding	2	1			7	3				14	6		34
- total	8	2			15	14				51	30		120
Mol. Mut. Freq.	8.81×10^{-7}	3.43×10^{-7}			3.65×10^{-7}	5.68×10^{-7}				6.67×10^{-7}	8.05×10^{-7}		
Mutation Rates	1 in 1.13×10^6	1 in 2.91×10^6			1 in 2.74×10^6	1 in 1.76×10^6				1 in 1.50×10^6	1 in 1.24×10^6		

* BN = Brown Norway, F344 = Fisher, WI = Wistar; Dose is indicated as mg of ENU per kg bodyweight (three weekly injections).

** F344 3 x 60 and WI 3 x 60 were adapted from our previous study [19].

† For BN, all fertile males were selected for F1 pup production for screening. For F344 and WI, only the males from the highest-dosed fertile groups were selected F1 pup production.

‡ The WI animals were used only eight months for F1 pup production, the F344 and BN males were used for one year.

Mutation discovery by high-throughout resequencing

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

on a capillary sequencer, but excluding costs for personnel and equipment, is currently less than 20 eurocents per sample (sequencing read of 500 bp).

The procedure starts with rapid isolation of genomic DNA from tail clips, followed by a nested PCR for amplification of the region of interest (Fig. 1a). The first PCR is performed in multiplex, followed by dilution and non-multiplexed second PCR reactions. We found that most amplicons (~90%) work in arbitrary combinations of up to 8 amplicons. The second PCR reaction is modified in two ways. First, gene-specific primers with universal tail sequences (M13) are used to allow a subsequent standardized sequencing reaction with a universal primer (Fig. 1b). Second, both primer and dNTP concentrations in this reaction are decreased such that only small amounts remain after thermocycling. As a result, reproducible amounts of clean PCR products with only very little sample to sample variation are obtained, allowing immediate cycle sequencing after dilution of the template without the need for additional purification. Furthermore, we found that the combination of the high quality PCR template and the universal sequencing reaction using the M13 adapter allowed us to titrate down the costly sequencing chemicals significantly. The sequencing reactions, which are performed in a volume of only 5 μl, are purified by ethanol precipitation and analyzed on ABI 3730XL DNA analyzers, resulting in 96 sequencing reads (500-600 bp) every 30 minutes. We used PolyPhred [23] to automatically score potential heterozygous peaks in the sequencing reads, followed by manual inspection of these positions. Every candidate mutation is repeated in an independent PCR and bidirectional sequencing reaction for verification. Initial candidate mutation scoring was deliberately very tolerant to prevent missing rare mutations, resulting in relatively low reconfirmation rates (~7.5%, data not shown).

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

ENU-induced mutation frequency

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

The F344 strain displayed the lowest mutation rate: 1 mutation per 2.7 and 1.8 x 10^6 bp for the 3 x 30, and 40 mg/kg groups, respectively (Table 1). The mutation rate in the BN strain is much higher for the 3 x 20 mg/kg group (1 mutation per 1.1 x 10^6 bp), but unexpectedly lower for the 3 x 30 mg/kg group (1 per 2.9 x 10^6 bp). However, the numbers of animals and bases screened for these groups are relatively small due to the poor breeding characteristics of the mutagenized BN animals. In addition, analysis of the mutation rate per individual founder (Supporting Table S2) revealed that the apparent high rate for the 3 x 20 group of BN animals results mainly from the contribution of a single mutagenized male. Finally, the WI strain gives the best overall mutation rates of 1 per 1.5 and 1.2 x 10^6 bp for groups 3 x 35 and 40, respectively (Table 1).

Figure 1: Rat target-selected mutagenesis.

a) Overview of the procedure. Male rats are injected with stem cell mutagen ENU (I), mated with untreated females to generate an F1 population that harbors random heterozygous point mutations in their genomes. From the F1 animals tissue samples are collected (II), genomic DNA is extracted in 96-well blocks (III), gridded out in 384-well plates and PCR is used to specifically amplify regions of interest. The first PCR is done in multiplex, generating eight amplicons in a single reaction (IV). In the second PCR, the multiplex PCR1 products are used as template for eight separated nested PCRs (IV). PCR samples are checked on an agarose gel for the presence of the correct product (V). Cost-effective dideoxy sequencing is performed using universal M13 primer (VI) and subsequent purification of products by precipitation. Samples are analyzed on a 96-capillary 3730XL DNA analyzer (VII) producing sequence files that are inspected using the PolyPhred

software package for the presence of heterozygous positions (VIII). Finally, candidate mutations, which are tagged by PolyPhred, are reconfirmed in an independent PCR and sequencing reaction to verify the mutation. **b)** Graphical representation of the PCR and sequencing assay design. Gene-specific oligonucleotides 1 and 4 are used in the first multiplexed PCR. The PCR product is diluted and used as template in the second PCR using oligonucleotides 2 and 3 that contain universal M13 tails. The second PCR contains limited amounts of dNTPs and oligonucleotides, eliminating the need for purification of the PCR product before sequencing. Sequencing is performed with diluted PCR product as template in a universal reaction mixture using universal M13 oligonucleotide and minimized amounts of sequencing chemicals.

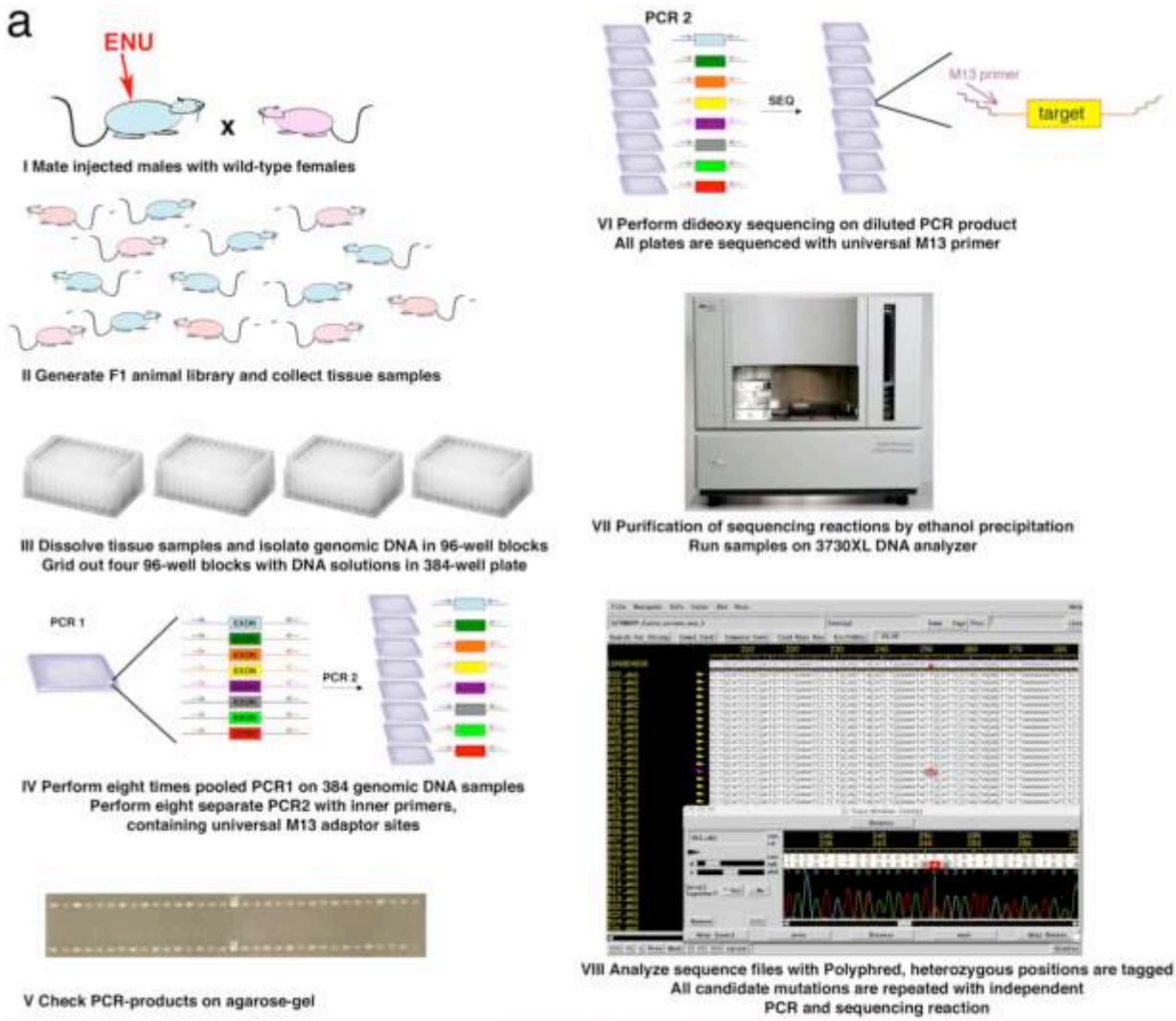


Table 2: Overview of mutations in target genes

Gene	Description	ENSEMBL_ID	Strain	Base change	AA change	Pred. cat.*	Alt. Splic.‡	Domain information†
nonsense mutations: 6 (~7% of coding mutations)								
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T10185A	C413X		YES	
MCH	Melanin Conc. Hormone	ENSRNOG00000004632	WI	A1205T	K50X		NO	
MSH6	Mismatch Repair Protein 6	ENSRNOG00000016134	WI	T12645A	L373X		NO	
NNOS	Neuronal Nitric-Oxide Synthase	ENSRNOG00000001130	WI	A28708T	K407X		NO	
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	BN	C3657G	Y62X		YES	
SERT	Serotonin Transporter	ENSRNOG00000003476	WI	C3924A	C209X		NO	
missense mutations: 56 (~64% of coding mutations)								
ADRA2B	Alpha2B Adrenergic Receptor	ENSRNOG00000013887	WI	T1116C	I39T	1	NO	In 1st tmd
AVP1A	Vasopressin V1A Receptor	ENSRNOG00000004400	WI	T1674G	V143G	1	NO	In 3rd tmd
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T8553C	I75T	1	NO	In 1st tmd
CHRND	Acetylcholine Receptor Delta	ENSRNOG00000019527	WI	T6147A	V308E**	1	NO	1st residue of 2nd tmd
DBH	Dopamine beta-Hydroxylase	ENSRNOG00000006641	WI	A15984G	Y518C	1	NO	Not annotated
DRD1	Dopamine Receptor 1	ENSRNOG00000023688	F344	T1215A	V72E**	1	NO	In 2nd tmd
FSHR	Follicle Stimulating Horm. Rec.	ENSRNOG00000016783	F344	A235725G	D566G	1	NO	Between 3rd and 4th tmd
GPR54	G Protein-coupled Receptor 54	ENSRNOG00000011954	F344	T1801G	V96G	1	NO	In 2nd tmd
IEDA	Orphan Receptor IEDA	ENSRNOG00000014793	BN	A81288T	D679V	1	NO	C-terminal part
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	T90087C	L42P	1	NO	Not annotated
LHR	Luteinizing Hormone Receptor	ENSRNOG00000016712	BN	T37228C	L269P**	1	NO	In 1st tmd
MSH6	Mismatch Repair Protein 6	ENSRNOG00000016134	WI	A15933G	N1113D	1	NO	MutS-C
MYOC	Myocilin	ENSRNOG00000003221	WI	T6634A	L124Q	1	NO	Coiled coil
NLG3	Neuroligin 3 Precursor	ENSRNOG00000003812	WI	T8023A	V230D	1	NO	Not annotated
NNOS	Neuronal Nitric-Oxide Synthase	ENSRNOG00000001130	WI	A38127T	I616F	1	NO	NO-synthase
NPY2R	Neuropeptide Y Receptor 2	ENSRNOG00000022004	F344	T1830A	V277E**	1	NO	In 6th tmd
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	F344	A3856T	I129F	1	YES	In 3rd tmd
RELN	Reelin Precursor	ENSRNOG00000006665	WI	A20691G	D2814G	1	YES	Not annotated
SLC6A8	Sodium-dep. choline transporter	ENSRNOG00000018857	WI	T6922C	F407L	1	NO	In 8th tmd
SYPH	Synaptophysin	ENSRNOG00000010223	WI	T7519A	C187S	1	NO	Between 3rd and 4th tmd
BDNF	Brain-Derived Neurotrophic Factor	ENSRNOG00000005051	F344	A46980T	H131L	2	NO	Not annotated
CCKBR	Cholecystokinin Receptor 2	ENSRNOG00000017679	WI	T8592C	V88A	2	NO	Between 1st and 2nd tmd
CHRNA6	Acetylcholine Receptor 6 alpha	ENSRNOG00000012283	WI	T6119G	I432S	2	NO	Between 3rd and 4th tmd
CYSLT2	Cysteinyl Leukotriene Receptor	ENSRNOG00000015042	F344	A740T	Y247F	2	NO	Between 3rd and 4th tmd
GLUR5	Glutamate Receptor 5	ENSRNOG00000001575	WI	A108310G	Y451C	2	NO	N-terminal part
GPR115	G Protein-coupled Receptor	ENSRNOG00000012535	WI	C4780G	T84R	2	NO	N-terminal part
GRPR	Gastrin Releasing Peptide Receptor	ENSRNOG00000004124	F344	T40617C	S289P	2	NO	3 residues behind 6th tmd
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	A299209C	Q511P	2	NO	Not annotated
LARGE	LARGE Glucosyltransferase	ENSRNOG00000013742	WI	T299400C	Y575H	2	NO	Not annotated
NMN	Neurotensin/Neuromedin N Precur.	ENSRNOG00000004179	WI	T6729A	M77K	2	NO	Not annotated
NMU2R	Neuromedin U Receptor 2	ENSRNOG00000014081	F344	T1512C	V129A	2	NO	In 3rd tmd
NNOS	Neuronal Nitric-Oxyde Synthase	ENSRNOG00000001130	WI	A38049T	S590C	2	NO	NO-synthase
NR0B2	Nuclear Receptor Subfamily 0 B2	ENSRNOG00000007229	F344	T3843A	L202M	2	NO	Hormone_rec_lig
NR1D1	Orphan Nuclear Receptor	ENSRNOG00000009329	F344	T4477A	V178D	2	NO	Znf_C4steroid
NTRK2	TRKB Tyrosine Kinase	ENSRNOG00000018839	WI	C57323A	T340K	2	NO	Between 2 tmd
OPRL	Orphanin FQ Receptor	ENSRNOG00000016768	WI	T1095C	S23P	2	YES	N-terminal part
PDYN	Prodynorphin Precursor	ENSRNOG00000016036	WI	A2671T	Q188L	2	NO	Not annotated
PINK	Serine/Threonine Kinase PINK1	ENSRNOG00000015385	WI	G12449A	V549M	2	NO	Not annotated
POMT1	Protein O-Mannosyl-Transferase 1	ENSRNOG00000010477	WI	T1825A	L47Q**	2	NO	In 1st tmd
ABCG5	ATP-Binding Cassette, G5	ENSRNOG00000005250	WI	C25706A	T319K	3	YES	N-terminal part
ATRX	Transcriptional Regulator ATRX	ENSRNOG00000002550	WI	C23842T	R271C	3	NO	Not annotated
CHRM5	Muscarinic Acetylcholine Rec. 5	ENSRNOG00000006397	F344	G2300A	V343I	3	NO	Between 5th and 6th tmd
DOPTA	Dopamine Transporter	ENSRNOG00000017302	F344	G1070T	V24L	3	NO	N-terminal part
ESDN	Endoth-/SMC-der. neuropilin-like	ENSRNOG00000001651	WI	A119852G	M458V	3	NO	1 Residue before tmd
GPR115	G Protein-coupled Receptor 115	ENSRNOG00000012535	BN	A8746T	L284F	3	NO	N-terminal part
HSP90	Heat-Shock Protein 90	ENSRNOG00000019834	WI	G2412T	V253L	3	NO	Low complexity
HD	Huntington Disease Protein	ENSRNOG00000011073	F344	T116025A	S1264T	3	NO	Not annotated
MSH2	Mismatch Repair Protein 2	ENSRNOG00000015796	F344	A5135G	N105S	3	NO	MutS_N
MYOC	Myocilin	ENSRNOG00000003221	WI	T15966A	Y458F	3	NO	Olfac_like
NLG3	Neuroligin 3 Precursor	ENSRNOG00000003812	WI	A1073T	I25F	3	NO	Carboxylesterase
NPY2R	Neuropeptide Y Receptor 2	ENSRNOG00000022004	WI	T1161C	V54A	3	NO	In 1st tmd
NPY4R	Neuropeptide Y Receptor 4	ENSRNOG00000020282	WI	A1587C	E196A	3	NO	Between 3rd and 4th tmd
OPRM1	MU-Type Opioid Receptor 1	ENSRNOG00000018191	WI	A34267G	I256V	3	NO	Last residue of 5th tmd
P53	p53	ENSRNOG00000010756	WI	C1039G	I10M	3	NO	P53
TH	Tyrosine 2-Monooxygenase	ENSRNOG00000020410	WI	C5502A	D275E	3	NO	Aaa-hydroxylase
GRM8	Metabotrophic Glutamate Rec. 8	NM_022202	WI	C71547T	T341I	n.d.	NO	Between 2nd and 3rd tmd

silent mutations: 24 (~28% of coding mutations)
non-coding mutations: 34

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

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

Mutation spectrum and characteristics

The amplicons used for determining the molecular mutation frequency were designed to contain mostly exonic sequence. 86 out of the 120 mutations are located in the protein coding sequences of target genes; six result in the introduction of a premature stop codon in the open reading frame (nonsense, ~7%) and are most likely to result in a functional knockout of the gene, 56 change an amino acid (missense, ~65%), and 24 have no effect on the protein coding capacity (silent, ~28%) (Table 2). According to the rat genome annotation (Ensembl v29.3f) two of the nonsense mutations reside in an exon that may be spliced out, resulting in a splice form-specific knockout (Table 2).
We used two programs, SIFT [24] and PolyPhen [25] that use phylogenetic conservation and structural information to predict the potential effect of a mutation on protein function, to analyze the 56 missense mutations. Based on the output of these programs, we defined three categories (Table 2). Sixteen mutations were predicted not to affect protein function by both programs (category 3), 19 are predicted to affect protein function by one of the two programs (category 2): and 20 mutations are predicted to affect protein function by both programs (category 1). These predictions may help selecting primary candidates for further analysis, as only experimental work will be able to give an answer to the question, which missense mutations have a biological effect.
Interestingly, 38 out of the 56 missense mutations occur in transmembrane domain-containing proteins and 15 mutations do alter an amino acid in a transmembrane domain (Table 2). Residue changes from a hydrophobic to a charged amino acid and the introduction of helix-distorting prolines may affect proper insertion of the protein in the membrane, resulting in inefficient or deficient trafficking to its proper location, or may affect channel properties of specific proteins. Five mutations in our set meet these criteria and four of them were indeed predicted by both SIFT and PolyPhen to affect protein function (Table 2). Together with the six knockouts, these mutations are most likely to have functional consequences for the protein. Currently, we are crossing out these animals and are breeding the mutations to homozygosity. Three of the potentially functional amino acid changes have been crossed to the next generation, including acetylcholine receptor delta subunit (CHRND), protein *O*-mannosyl transferase (POMT1), and neuropeptide Y receptor (NPY2R). Finally, four knockouts have currently been passed on to the next generation, namely the nociceptin receptor OPRL1 (opioid-like receptor-1 or orphanin FQ Receptor), the DNA mismatch repair protein MSH6 that is involved in hereditary non-polyposis colorectal cancer (HNPCC), the melanin concentrating hormone (MCH), and the serotonin transporter (SERT/SLC6A4). The latter is homozygous viable and has been confirmed to be a functional knockout by biochemical and pharmacological analysis (unpublished results).

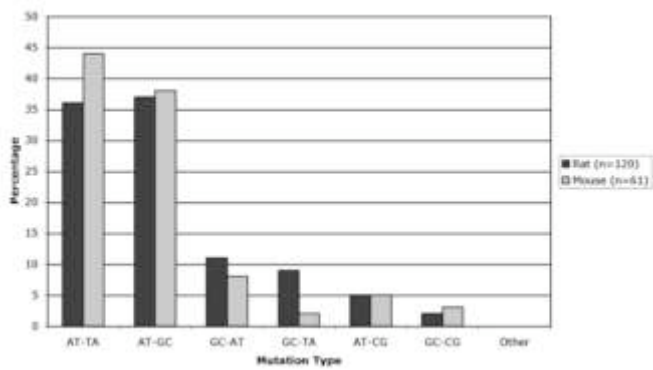
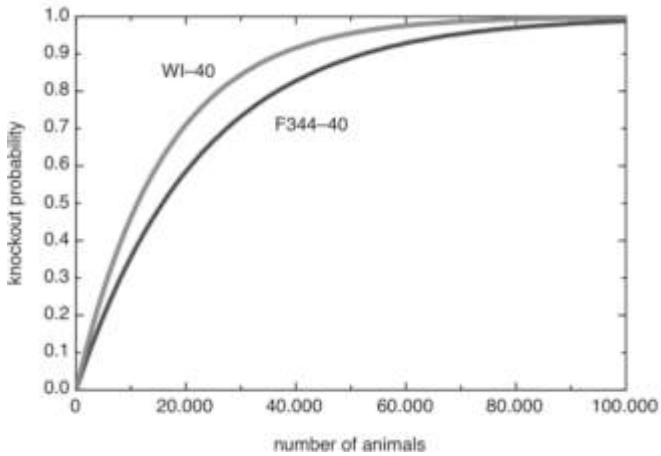


Figure 2: ENU-induced mutation spectrum of the rat (this study) and mouse [27].
Only point mutations have been identified in these studies.

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



Conclusions

We have established robust and scaleable technology to generate knockout and mutant rats using target-selected mutagenesis and have illustrated the effectiveness of the methodology by the identification of 120 induced gene mutations, of which six result in exclusive rat knockout models, due to induced nonsense mutations.
Optimal ENU mutagenesis conditions using four rat strains were determined, resulting in similar mutation spectra [27] and rates (about 1 mutation every 1 - 1.5 x 10^6 bp) as obtained for specific mouse strains [21,22]. Furthermore, we have set up an efficient resequencing approach for the identification of rare induced point mutations. The major advantages of this mutation discovery technology are the robustness and low costs per reaction and the highly informative output. The resulting raw data is very well-suited for automatic bioinformatic analysis, providing detailed sequence information for every mutation per individual animal. The platform is very flexible and scaleable, as it can be implemented for both large-scale projects using automation by robotics, as well as for moderate-scaled experiments using multi-channel pipettes. In contrast to previously used technology for the generation of rat knockouts, employing a yeast-based truncation assay [18], resequencing not only detects nonsense mutations, but also potentially interesting missense mutations. In addition, resequencing provides direct information on the actual ENU-induced molecular mutation frequency and spectrum. From the four strains tested, the WI strain was found to be the most robust strain, based on its molecular mutation frequency as well as its good breeding properties and it may therefore be the strain of choice for routine production of knockout rats.
We calculated the probability to identify a specific knockout based on the obtained molecular mutation frequencies (Fig. 3). The total number of coding bases in the current rat genome release is 28.4 x 10^6 bp, of which 1.7 x 10^6 bp (6%) could be mutated into a stop codon by the most frequent ENU mutations. Assuming random distribution of these potential ENU-inducible nonsense positions over the 22,155 annotated genes and a mutation rate of 1 per 1.2 x 10^6 bp for the Wistar strain, the chance to find a knockout for any average-sized (approximately 1,300 bp) gene in a library of 5,000, 10,000, 20,000 and 50,000 mutant F1 rats would be 27, 46, 71, and 96%, respectively.

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

	Exon of interest[a]		Gene of interest[b]	
	PCRs[c]	SEQs[c]	PCRs[c]	SEQs[c]
1 mutation 70%[d]	6,514	5,790	9,135	8,120
1 mutation 95%[d]	16,200	14,400	22,603	20,090
1 stopcodon 70%[e]	149,963	133,300	181,916	161,700
1 stopcodon 95%[e]	324,900	288,800	452,816	402,500

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

In Table 3, we illustrated how many PCRs and sequencing reactions would be required to obtain a mutation or a knockout in an exon or gene of interest with 70% and 95% probability. Based on the Wistar mutation frequency about 6,500 or 16,000 PCRs and 6000 or 14,000 sequencing reactions are needed to find a mutation in a single amplicon (250 bp coding) with 70% or 95% probability, respectively. For finding a premature stopcodon in an average sized gene of 1.25 bp (assuming seven protein-coding exons) with 95% probability, up to 450,000 PCRs and 400,000 sequencing reactions are needed.

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

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

Acknowledgements

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

References

1. Gibbs, R.A. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428, 493-521.
2. Jacob, H.J. and Kwitek, A.E. (2002) Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet*, 3, 33-42.
3. Kolb, B. (2005) Neurological models. In *The behavior of the laboratory rat.* (Kolb, B. and Whishaw, I.Q., eds), pp. 449-461, Oxford University Press.

4. Russell, W.L. *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.
5. Nolan, P.M. *et al.* (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet*, 25, 440-443.
6. Hrabe de Angelis, M.H. *et al.* (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet*, 25, 444-447.
7. Cox, R.D. and Brown, S.D. (2003) Rodent models of genetic disease. *Curr Opin Genet Dev*, 13, 278-283.
8. Colbert, T. *et al.* (2001) High-throughput screening for induced point mutations. *Plant Physiol*, 126, 480-484.
9. McCallum, C.M. *et al.* (2000) Targeted screening for induced mutations. *Nat Biotechnol*, 18, 455-457.
10. Coghill, E.L. *et al.* (2002) A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet*, 30, 255-256.
11. Wienholds, E. *et al.* (2002) Target-selected inactivation of the zebrafish rag1 gene. *Science*, 297, 99-102.
12. Beier, D.R. (2000) Sequence-based analysis of mutagenized mice. *Mamm Genome*, 11, 594-597.
13. Jansen, G. *et al.* (1997) Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat Genet*, 17, 119-121.
14. Till, B.J. *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res*, 13, 524-530.
15. Wienholds, E. *et al.* (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res*, 13, 2700-2707.
16. Till, B.J. *et al.* (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol*, 4, 12.
17. Perry, J.A. *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.
18. Zan, Y. *et al.* (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol*, 21, 645-651.
19. Smits, B.M. *et al.* (2004) Target-selected mutagenesis of the rat. *Genomics*, 83, 332-334.
20. Justice, M.J. *et al.* (2000) Effects of ENU dosage on mouse strains. *Mamm Genome*, 11, 484-488.
21. Quwailid, M.M. *et al.* (2004) A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm Genome*, 15, 585-591.
22. Concepcion, D. *et al.* (2004) Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. *Genetics*, 168, 953-959.
23. Nickerson, D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res*, 25, 2745-2751.
24. Ng, P.C. and Henikoff, S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 31, 3812-3814.
25. Ramensky, V. *et al.* (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 30, 3894-3900.
26. Larsen, L.A. *et al.* (2001) Recent developments in high-throughput mutation screening. *Pharmacogenomics*, 2, 387-399.
27. Noveroske, J.K. *et al.* (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome*, 11, 478-483.
28. Nakatsukasa, E. *et al.* (2001) Generation of live rat offspring by intrauterine insemination with epididymal spermatozoa cryopreserved at -196 degrees C. *Reproduction*, 122, 463-467.
29. Lipshutz, R.J. *et al.* (1999) High density synthetic oligonucleotide arrays. *Nat Genet*, 21, 20-24.
30. Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics*, 5, 433-438.

Supporting Material

Supporting Table S1: Fertility measurements after ENU mutagenesis

Brown Norway				Fisher				Wistar			
20 #1	8	**30 #15**	0	**20 #1**	18	**40 #1**	19	**30 #1**	32	**40 #1**	24
20 #2	14	**30 #16**	0	**20 #2**	3	**40 #2**	0	**30 #2**	0	**40 #2**	0
20 #3	0	**30 #17**	0	**20 #3**	16	**40 #3**	5	**30 #3**	0	**40 #3**	0
20 #4	0	**30 #18**	0	**20 #4**	10	**40 #4**	16	**30 #4**	43	**40 #4**	0
20 #5	16	**30 #19**	3	**20 #5**	24	**40 #5**	18	**30 #5**	42	**40 #5**	1
20 #6	7	**30 #20**	0	**20 #6**	31	**40 #6**	0	**30 #6**	40	**40 #6**	10
20 #7	9	**30 #21**	0	**20 #7**	19	**40 #7**	0	**30 #7**	30	**40 #7**	20
20 #8	9	**30 #22**	4	**20 #8**	13	**40 #8**	10	**30 #8**	39	**40 #8**	34
20 #9	6	**30 #23**	1	**20 #9**	5	**40 #9**	5	**30 #9**	42	**40 #9**	19
20 #10	1	**30 #24**	0	**20 #10**	0	**40 #10**	0	**30 #10**	22	**40 #10**	0
20 #11	8			**20 #11**	0	**40 #11**	28				
20 #12	7	**40 #1**	0	**20 #12**	13	**40 #12**	0	**35 #1**	32		
20 #13	0	**40 #2**	0	**20 #13**	0	**40 #13**	0	**35 #2**	37		
20 #14	2	**40 #3**	0	**20 #14**	5	**40 #14**	6	**35 #3**	31		
20 #15	4	**40 #4**	0	**20 #15**	9	**40 #15**	0	**35 #4**	17		
20 #16	12	**40 #5**	0	**20 #16**	11	**40 #16**	12	**35 #5**	26		
20 #17	12	**40 #6**	0	**20 #17**	13	**40 #17**	3	**35 #6**	35		
20 #18	13	**40 #7**	0	**20 #18**	10	**40 #18**	3	**35 #7**	0		
20 #19	12	**40 #8**	0	**20 #19**	23			**35 #8**	18		
20 #20	8	**40 #9**	0	**20 #20**	0			**35 #9**	45		
20 #21	19	**40 #10**	0					**35 #10**	22		
20 #22	7	**40 #11**	0	**30 #1**	0						
20 #23	0	**40 #12**	0	**30 #2**	6						
20 #24	2	**40 #13**	0	**30 #3**	0						
		40 #14	0	**30 #4**	22						
30 #1	0	**40 #15**	0	**30 #5**	8						
30 #2	0	**40 #16**	0	**30 #6**	9						
30 #3	0	**40 #17**	0	**30 #7**	27						
30 #4	0	**40 #18**	0	**30 #8**	27						
30 #5	0	**40 #19**	0	**30 #9**	16						
30 #6	0	**40 #20**	0	**30 #10**	10						
30 #7	3	**40 #21**	0	**30 #11**	14						
30 #8	0	**40 #22**	0	**30 #12**	15						
30 #9	0	**40 #23**	0	**30 #13**	16						
30 #10	4	**40 #24**	0	**30 #14**	0						
30 #11	2	**40 #25**	0	**30 #15**	23						
30 #12	0	**40 #26**	0	**30 #16**	0						
30 #13	4			**30 #17**	12						
30 #14	0			**30 #18**	0						

Fertility of all individual mutagenized founders was measured by up to four matings. Dose is indicated as mg of ENU/kg bodyweight (three weekly injections), followed by the animal number.

Supporting Table S2: Molecular mutation frequencies for individual fertile founders

Strain / Dose*	Bases Screened	Mutations	Mol. Mut. Freq.	Strain / Dose	Bases Screened	Mutations	Mol. Mut. Freq.
BN 20 1	9.29×10^{5}	0		**F344 40 1**	8.92×10^{5}	3	3.36×10^{-6}
BN 20 2	1.31×10^{5}	0		**F344 40 2**	2.43×10^{6}	1	4.12×10^{-7}
BN 20 3	4.47×10^{5}	0		**F344 40 3**	2.10×10^{6}	1	4.76×10^{-7}
BN 20 4	4.61×10^{5}	0		**F344 40 4**	2.52×10^{6}	1	3.97×10^{-7}
BN 20 5	1.11×10^{6}	0		**F344 40 5**	3.42×10^{6}	2	5.85×10^{-7}
BN 20 6	1.54×10^{6}	2	1.30×10^{-6}	**F344 40 6**	9.59×10^{5}	1	1.04×10^{-6}
BN 20 7	1.28×10^{6}	0		**F344 40 7**	4.34×10^{6}	1	2.30×10^{-7}
BN 20 8	1.25×10^{6}	4	3.20×10^{-6}	**F344 40 8**	3.32×10^{5}	3	9.04×10^{-6}
BN 20 9	6.63×10^{5}	0		**F344 40 9**	6.15×10^{5}	1	1.63×10^{-6}
BN 20 10	4.27×10^{4}	0					
BN 20 11	1.71×10^{5}	0		**WI 35 1**	7.32×10^{6}	4	5.46×10^{-7}
BN 20 12	4.20×10^{5}	2	4.76×10^{-6}	**WI 35 2**	7.50×10^{6}	6	8.00×10^{-7}
BN 20 13	8.81×10^{5}	0		**WI 35 3**	1.23×10^{7}	6	4.88×10^{-7}
				WI 35 4	8.22×10^{6}	6	7.30×10^{-7}
BN 30 1	1.56×10^{6}	1	6.41×10^{-7}	**WI 35 5**	1.42×10^{6}	1	7.04×10^{-7}
BN 30 2	6.28×10^{5}	0		**WI 35 6**	1.08×10^{7}	9	8.33×10^{-7}
BN 30 3	8.31×10^{5}	0		**WI 35 7**	3.47×10^{6}	4	1.15×10^{-6}
BN 30 4	1.09×10^{6}	1	9.17×10^{-7}	**WI 35 8**	8.87×10^{6}	7	7.89×10^{-7}
BN 30 5	2.13×10^{5}	0		**WI 35 9**	9.63×10^{6}	5	5.19×10^{-7}
BN 30 6	6.27×10^{5}	0		**WI 35 10**	6.91×10^{6}	3	4.34×10^{-7}
BN 30 7	8.81×10^{5}	0					
				WI 40 1	9.41×10^{6}	5	5.31×10^{-7}
F344 30 1	2.36×10^{6}	4	1.69×10^{-6}	**WI 40 2**	3.79×10^{6}	3	7.92×10^{-7}
F344 30 2	1.99×10^{6}	0		**WI 40 3**	1.76×10^{6}	2	1.13×10^{-6}
F344 30 3	4.77×10^{6}	2	4.19×10^{-7}	**WI 40 4**	9.07×10^{6}	7	7.17×10^{-7}
F344 30 4	2.74×10^{6}	0		**WI 40 5**	7.42×10^{6}	5	6.74×10^{-7}
F344 30 5	6.20×10^{6}	3	4.83×10^{-7}	**WI 40 6**	5.81×10^{6}	8	1.38×10^{-6}
F344 30 6	4.77×10^{6}	0					
F344 30 7	5.69×10^{6}	2	3.51×10^{-7}				
F344 30 8	4.45×10^{6}	1	2.25×10^{-7}				
F344 30 9	4.27×10^{6}	1	2.34×10^{-7}				
F344 30 10	3.81×10^{6}	2	5.25×10^{-7}				
F344 30 11	3.10×10^{6}	0					
F344 30 12	3.93×10^{6}	0					

*) BN = Brown Norway, F344 = Fisher, WI = Wistar; Dose is indicated as mg per kg bodyweight (3 weekly injections). Sterile males are excluded from this table.

CHAPTER 4

Genetic Variation in Coding Regions between and within Commonly Used Inbred Rat Strains

Adapted from Genome Research (2004) ***14:*** *1285 - 1290*

Genetic Variation in Coding Regions between and within Commonly Used Inbred Rat Strains

Bart M. G. Smits[1], Bert F. M. van Zutphen[2], Ronald H. A. Plasterk[1], and Edwin Cuppen[1*]

[1]Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[2]Department of Laboratory Animal Science, University of Utrecht, Yalelaan 2, 3584 CM Utrecht, The Netherlands

[*]Corresponding Author

Single Nucleotide Polymorphisms (SNPs) are the most common genetic variation in mammalian populations. Their significance is illustrated by their potential contribution to common disease but also by their potential for use in genetic association and mapping experiments. We have examined the genetic variation between commonly used inbred rat strains by using an efficient SNP discovery and typing assay based on enzyme-based (CEL I) heteroduplex cleavage. Screening of a panel of 96 different rat (sub-)strains for 100 genomic loci in 55 genes, whose human homologs are implicated in clinically relevant diseases like neurological disorder, cancer, schizophrenia, and obesity, resulted in the identification of 103 novel polymorphisms. As all strains are simultaneously genotyped in this setup, this allowed us to make an estimate of the genetic variation between and within commonly used rat inbred strains. Interestingly, we observed substantial genetic variation between colonies of the same inbred strain, maintained at different locations. Furthermore, we identified 17 nonsynonymous SNPs that may have an effect on protein function and contribute to phenotypic differences between different laboratory strains.

SNPs identified in this study can be found in the National Center of Biotechnology Information (NCBI) SNP database under accession numbers ss12588106 - ss12588203.

Introduction

Rat inbred strains are widely used as laboratory models in understanding basic biology and human health and disease. Currently, over 200 different rat inbred strains are used in laboratory studies, where most strains are specially selected as a model for specific human diseases [1,2]; http://rgd.mcw.edu). Knowledge of genetic variation between strains will be useful to obtain insight in the relationship between different strains, but also for the design of genetic mapping panels for association studies. Furthermore, as genetic variation within inbred strains does affect experimental design and interpretation, knowledge of the extent of this variation is essential. Although a strain is assumed to be inbred after at least 20 generations of subsequent brother-sister matings [3] and from that point on is considered to be genetically homogeneous, a small but uncharacterized degree of genetic variation will always remain in the population.

Currently, information on genetic variation in rat laboratory strains is limited to a set of microsatellite markers [4,5] that are generally located in noncoding regions. In this study, we focused on single nucleotide polymorphisms (SNPs) in gene coding regions of rat inbred strains. SNPs are the most abundant form of sequence variability in the vertebrate genome [6] and polymorphisms in gene coding regions are likely to contribute to the phenotypic differences between strains. Obviously, SNPs in noncoding regions may also have a functional effect as suggested by evolutionary conservation of parts of such sequences [7-9] and the mapping of a human disease susceptibility to a noncoding region [10]. However, the vast majority of human disease alleles that are identified up to now, have been found to be caused by single nucleotide polymorphisms resulting in nonsense or missense mutations at the gene coding level (OMIM, Online Mendelian Inheritance in Men database; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM).

Different experimental methods to detect genetic variation have been established [11]. For large-scale genome-wide detection of SNPs, DNA microarray technology is one of the most efficient methods and it has been successfully applied to the human [6,12,13] as well as to the mouse genome [14]. To discover novel SNPs, two different approaches can be followed. First, new polymorphic positions can be predicted by an *in silico* approach [15], where all available information in databases is mined and analyzed for the occurrence of polymorphisms. This results in candidate SNPs that require experimental verification. Second, experimental methods can be used directly, where direct sequencing is currently still the most accurate and informative approach, but other methods that are more cost-effective and/or allow higher throughput are continuously under development.

In this study, we have adapted high-throughput mutation discovery technology that is based on enzymatic cleavage of heteroduplexes using the CEL I nuclease isolated [16] for SNP discovery. This technology has been successfully used for reverse genetics approaches in *Arabidopsis* [17], zebrafish [18], and the rat [19], and is now commonly referred to as TILLING. The major advantage of the approach described here, is that it is well suited for simultaneous SNP discovery and genotyping using a large panel of individuals or samples. In addition, the accuracy of the method is high, the costs are relatively low, and the technology is easily scalable. We have assayed 100 coding regions in 96 rat inbred strains and substrains, scanning almost 6 x 10^6 base pairs, resulting in the identification and genotyping of 103 novel polymorphisms. Interestingly, we observed a considerable amount of genetic variation at the single-base level within inbred strains.

Methods

Genomic DNA isolation and origin of genomic DNA

Genomic DNA from reference strain BN/Crl was isolated using QIAGEN DNeasy 96® Tissue Kit (4). Genomic DNA from 96 inbred strains and substrains was made available through the Department of Laboratory Animal Science, Faculty of Veterinary Medicine, Utrecht University, The Netherlands. More detailed information on the origin of these strains can be found in Supporting Table S3. Five microliters of genomic DNA (concentration of approx. 2 ng/μl per sample), was gridded-out in 384 wells plates. In total, two sets of 96 testing strains and two sets of 96 reference strain DNA (BN/Crl), were transferred into a single 384-wells plate using a 96-channel pipettor (Hydra-96).

Primer design

Genomic organization and exon-flanking intron sequences of *Rattus norvegicus* target genes were determined using GENOTRACE. The published cDNA sequence was used as input. Nested sets of oligonucleotides for amplification of exon sequences were automatically designed using a PRIMER3-based web-application (http://primers.niob.knaw.nl) with optimal melting temperatures of 58°C. For this study we have designed amplicons for only the large exons (> 300 bp) in genes of our specific interests. As a result, coding regions for some of the genes are fully covered, whereas others are only partially screened, resulting in an average

coverage percentage of about 56%. Detailed primer information can be obtained from the authors upon request.

CEL I-based polymorphism detection

All PCR, pooling, and CEL I digestion pipetting steps were done on a Genesis Workstation 200 (Tecan). Target sequences were amplified by nested PCR in 384-wells plates. The first PCR was done with gene specific primers and carried out using a touchdown cycling program (92°C for 60 sec; 30 cycles of 92°C for 20 sec, 65°C for 20 sec with a decrement of 0.5°C per cycle, 72°C for 60 sec; followed by 10 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 60 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems). PCR reaction mixes contained 5 µl genomic DNA, 0.2 µM forward primer and 0.2 µM reverse primer, 200 µM of each dNTP, 25mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate (pH 8.7), and 0.2 U Taq Polymerase in a total volume of 10 µl.

Of the first PCR reaction, 1 µl was used as template for the second PCR reaction. The second PCR reaction contained gene specific primers, at their 5' end elongated with universal M13 adaptor sequences (forward 0.08 µM and reverse 0.04 µM). Additionally, the reaction mixture contained corresponding universal M13-forward primer (5'-TGTAAAACGACGGCCAGT, 0.12 µM) and universal M13-reverse primer (5'-AGGAAACAGCTATGACCAT, 0.16 µM). The universal M13 primers were labeled with fluorescent dyes, IR Dye 700 and IR Dye 800, respectively. Furthermore, the nested PCR mixture contained 200 µM of each dNTP, 25 mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate (pH 8.7), and 0.2 U Taq Polymerase in a total volume of 5 µl. Standard cycling conditions were used for the nested PCR reactions (30 cycles of 92°C for 20 sec, 58°C for 40 sec and 72°C for 60 sec).

After the second PCR reaction, 2.5 µl of each sample of the 96 testing strains was mixed in a fresh 384-well plate with 2.5 µl of the reference strain PCR, followed by heteroduplex formation (99°C for 10 min and 70 cycles of 70°C for 20 sec with a decrement of 0.3°C per cycle). Specific heteroduplex cleavage was performed by adding to each sample 10 µl of CEL I mixture, containing 10 mM HEPES pH 7.0, 10 mM $MgSO_4$, 10 mM KCl, 0.002% Triton X-100, 0.2 µg BSA and 0.01 µl CEL I enzyme solution (isolated from celery according to Oleykowski et al. [16] with minor modifications; protocol available at http://cuppen.niob.knaw.nl) and incubation at 45°C for 15 min. The CEL I reactions were stopped by addition of 5 µl 75 mM EDTA pH 8.0.

Fragments were purified using Sephadex G50 (medium coarse, Sigma) mini-columns in 96-wells filter plates (Multiscreen HV, Millipore) and eluted into plates prefilled with 5 µl formamide loading buffer (37% (v/v) de-ionized formamide, 4 mM EDTA pH8.0, 90 µg/ml bromophenol blue) per well. Samples were concentrated to about 1.5 µl by heating at 85°C for 45 to 60 min without cover. 0.3 µl was applied to a 96-lane membrane comb (The Gel-company) and loaded on 25 cm denaturing 6% polyacrylamide gels on LI-COR 4200 DNA analyzers. Raw TIFF-images produced by the analyzers were modified and visualized using Adobe Photoshop and potential polymorphisms were detected and scored manually.

Verification and identification of polymorphisms by sequencing

PCR reactions for polymorphism identification were done by using the same conditions as for CEL I mediated polymorphism detection. In the nested PCR, universal M13 oligo's were omitted and only gene specific primers were used, both in the concentration of 0.2 µM. Nested PCR products were diluted with 30 µl water and 1 µl was used as template for the sequencing reactions. Sequencing reactions, containing 0.5 µl DYEnamic ET Terminator (Amersham Pharmacia Biotech), 3.5 µl ET Terminator dilution buffer (Amersham Pharmacia Biotech) and 0.5 µM nested gene specific forward primer or reverse primer in a total volume of 10 µl, were performed using cycling conditions recommended by the manufacturer. Sequencing products were purified using Sephadex G50 (superfine coarse, Sigma) mini-columns and analyzed on a 96-capillary 3700 DNA analyzer (Applied Biosystems). Sequences were analyzed for polymorphisms using PolyPhred [20].

SIFT and PolyPhen

To predict the impact of the nonsynonymous SNPs we found, we used two different applications, SIFT (v2.0) [21] and PolyPhen (command-line version) [22], both standalone versions. To be able to compare the data, the same databases (Swiss-Prot+TrEMBL+TrEMBL_NEW, downloaded from ftp://ftp.expasy.org) and BLAST parameters (expectation cut-off = 1E-04) were applied for the analysis.

Results and Discussion

SNP discovery

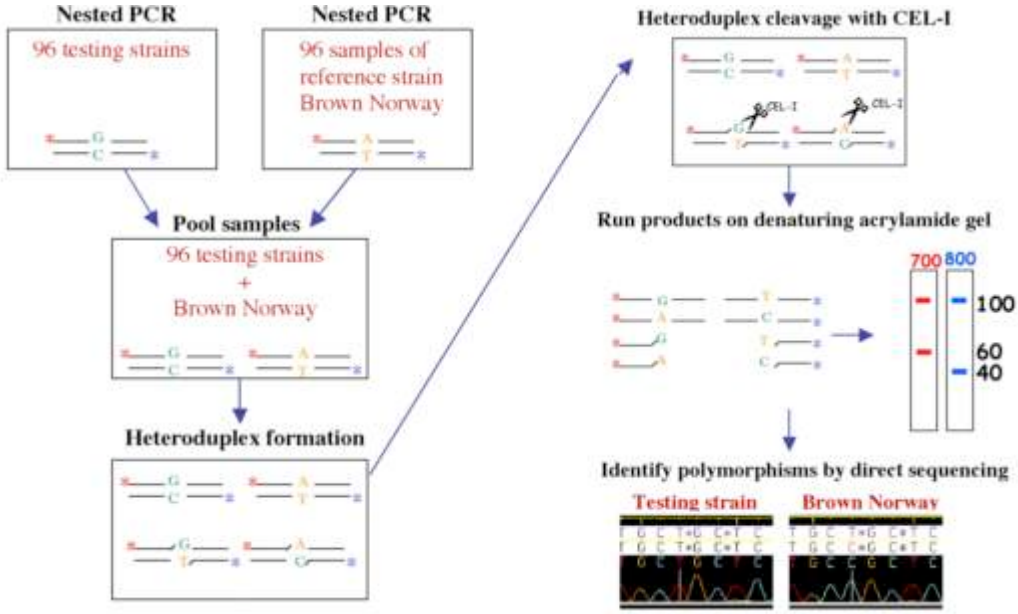
We have selected a set of 55 genes that are known to play a role in neurological and endocrine processes, cholesterol metabolism, and cancer (Supporting Table S1). Sequence variation in the human homologues of these genes may play a role in clinically relevant diseases like schizophrenia, obesity and cancer. Primers for amplification of selected exons (total 100 loci) were designed using local genomic assemblies generated by the GENOTRACE program [23], as no rat genome assembly was available at this point, in combination with a Primer3-based [24] primer design program (http://primers.niob.knaw.nl) for universal design of oligo's, allowing automated simultaneous PCR in a robotic set-up.

The resulting amplicons were screened for polymorphisms in a set of 96 inbred rat strains and substrains using a modified CEL I nuclease-based [16] SNP detection assay (Fig. 1). Briefly, fluorescently labeled PCR products from the reference strain Brown Norway, which is being sequenced as part of the rat genome sequencing project (http://www.hgsc.bcm.tmc.edu/projects/rat/), and testing strain are mixed and allowed to form heteroduplexes that are the substrate for the CEL I nuclease. Digested products are analyzed on denaturing acrylamide gels and polymorphisms will show up as additional bands (Supporting Fig. S1). Samples showing additional fragments are subsequently sequenced to reveal the molecular nature of the polymorphism. Although one has to sequence only one representative sample for every SNP, in this study we verified all polymorphisms in all amplicons by direct sequencing of the PCR product. No false positives were encountered and no novel polymorphisms were detected in the sequencing phase, indicating that also the rate of false negatives is very low.

In total, we screened more than 5.8×10^6 base pairs, with 40,977 base pairs of coding and 19,198 base pairs of noncoding sequence per (sub)strain and identified 103 novel polymorphisms (Supporting Table S1), that were simultaneously genotyped in 96 rat strains and substrains (Fig. 2). SNPs in a single gene that were found to cosegregate in our data set are represented as a single polymorphic locus in Figure 2.

The average failure rate per amplicon was found to be only 1.3%, and is most likely due to pipetting errors and inaccuracies introduced by the automated robotic setup of the PCR and CEL I reactions. No inconsistencies were observed between the CEL I-based and the resequencing results. Taken together, this strategy is an efficient and reliable approach for high throughput SNP discovery and simultaneous genotyping in medium to large sample sets.

Figure 1: Schematic overview of the polymorphism discovery strategy.
PCR on genomic DNA of 96 testing strains and on a set of 96 samples of the reference strain Brown Norway (BN/Crl) produces two sets of 96 PCR products that contain different fluorescent labels at either end (IRD700 and IRD800), indicated by red and blue stars). Pooling of tester samples with reference samples, followed by denaturing and reannealing results in the formation of heteroduplex DNA in the case of polymorphisms between the strains used. Heteroduplex DNA is specifically cleaved by the CEL I nuclease at the site of the mismatch and the resulting fragments are separated on a denaturing polyacrylamide gel with fluorescence detection. Subsequently, representative samples for a specific polymorphism are sequenced to reveal the molecular nature of the polymorphism.



SNP characteristics and frequency

Although the focus of this study is on polymorphisms in coding regions, we did identify 38 SNPs in noncoding regions as a result of the PCR-based strategy with primers designed in exon-flanking sequences. The calculated SNP frequency in exon, intron and untranslated (UTR) regions is 1 in 630, 555, and 345 base pairs, respectively (Supporting Table S1). Generally, SNP frequencies are found to be higher in noncoding sequences compared to coding sequences. The difference in our set might be due to the method of screening, for which it has been reported that it does not efficiently detect polymorphisms close to the ends of the amplicon [17]. Rather crude correction of this effect by subtraction of 30 base pairs from either end of the amplicon, would result in SNP frequencies in exon-flanking intronic regions of 1 in 367 nucleotides (29 SNPs in 10638 bp) and 1 in 269 nucleotides for UTR sequences (9 SNPs in 2419 bp), which are more in line with expectations.

Completely random mutations will be divided into transitions and transversions in a 1:2 ratio. However, biological datasets tend to have a strong bias towards transitions due to DNA methylation, chemical differences between bases, and differences in DNA repair efficiency for different types of nucleotide mismatches. As for the mouse (roughly 66.7% transitions; [14]) and for human (64% transitions; [13]), our dataset shows a similar bias towards transitions (78.4%). It is not likely that this bias is a result of a decreased sensitivity of CEL I for certain mismatches, since results obtained by others [16] and by us ([18]; E.C., unpubl.) show that all possible heteroduplex mismatches are recognized efficiently by CEL I.

SNPs between inbred strains

Previous analysis on microsatellites suggested that Brown Norway was genetically most similar to wild rats and furthest away from laboratory rat strains [4,5]. Although the current genotyping data is not sufficient to build a reliable phylogenetic tree, our results confirm that Brown Norway is clearly most distant from the other inbred strains tested. Furthermore, some clear relationships

between strains can be observed. For example, DA/Han and LEW/Cub have 96% (77 of 80) polymorphic loci identical (Fig. 2), which indicates that these strains are closely related. In rat strain databases, the DA strain is suggested to be closely related to the COP strain (http://rgd.mcw.edu/strains/; http://www.mh-hannover.de/institute/tierlabor/da.html), but in our set DA and COP share only 85% of the alleles. Of course, many more polymorphic loci should be explored to reliably trace the history of rat inbred strains, as illustrated by the results of Canzian [5] and Thomas et al. [4], showing close relationship between DA and COP using 995 and 4800 microsatellite markers, respectively.

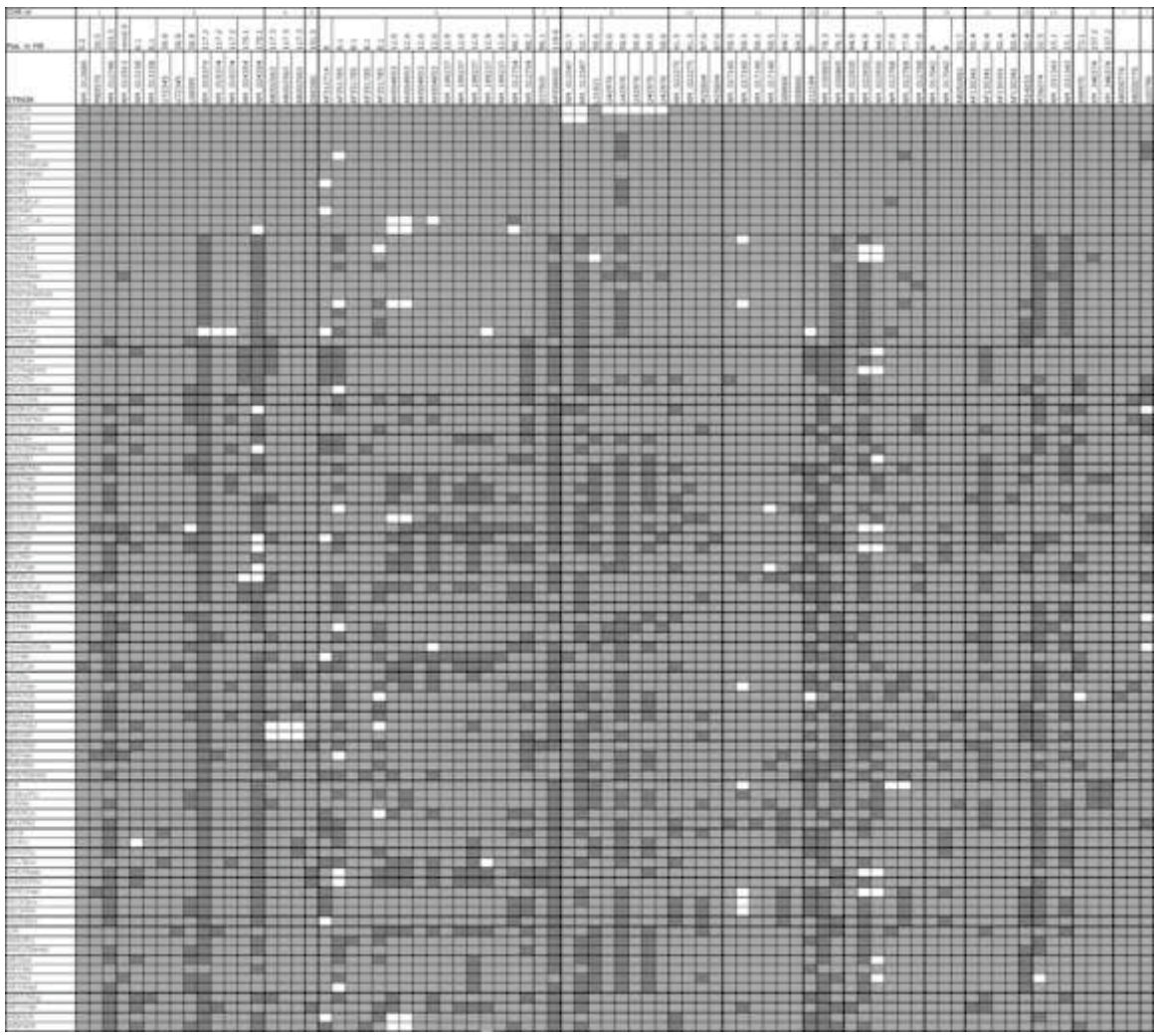


Figure 2: Genotype of the identified polymorphisms for 96 rat inbred strains and substrains.
Polymorphisms are typed in reference to the Brown Norway strain (BN/Crl). The bright grey squares represent the allele in accordance with BN/Crl. The dark grey squares represent the presence of the alternative allele. The white squares represent genotyping failures. Different isolates of the same inbred strain are grouped together, separated by thin lines. Thick lines separate different inbred strains. SNPs that are present in the same gene and cosegregate in our data set are represented as a single locus, reducing the number of polymorphic sites to 80. The polymorphic loci are ordered on the horizontal axis in line with the map position based on the June 2003 rat genome assembly (UCSC version rn3). The map position of the last 2 genes is unknown.

SNPs within inbred strains

Individual animals from inbred strains are commonly considered to be genetically homogeneous and are called isogenic. However, our genotyping results using independent isolates of inbred strains, revealed significant heterogeneity (Figs. 2 and 3). Statistically, at least 98.6% of the loci in each animal from an inbred population should be homozygous [25,26]. We have genotyped 14 Brown Norway substrains and found that five of the 80 loci were variable, representing 6% of genetic variability on single-base level. In three substrains of ACI, we found variation in nine of the 80 loci, representing 11% of genetic variation. Remarkably, in the widely used Lewis (LEW) strain was found 16 out of the 80 loci to be polymorphic, representing genetic variation of 20%. In Figure 3, variation percentages were plotted for 9 inbred strains, for which multiple individuals were typed.

The observed genetic variation within inbred strains may be the result of spontaneous mutations that occurred during breeding, inaccuracies in breeding programs, or a reflection of residual heterogeneity at the time of splitting breeding populations. The first possibility would result in novel SNPs, unique to the (sub)strain, whereas the latter two possibilities would result in nonunique SNPs shared with other strains. We only identified SNPs shared with other inbred strains, excluding spontaneous mutations as a source.

It should be noted that from these results no conclusions can be drawn about the degree of isogenicity within specific breeding colonies of inbred strains, but it illustrates the potential variation between different isolates worldwide that should be taken into account when comparing experimental results from different labs using the same inbred lines.

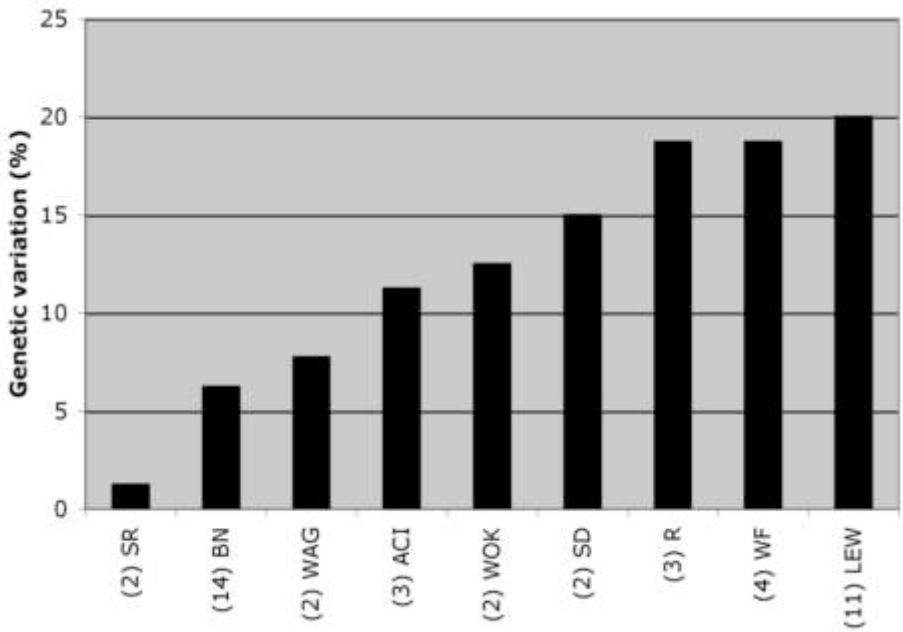


Figure 3: Genetic variation between substrains of nine frequently used rat inbred strains.
Percentages were calculated as percentage of variation of the total amount of loci for which the substrains were typed. The number below the name of the strain represents the number of substrains that were typed.

Effect of SNPs

For the 65 SNPs in coding regions, we find that about 26% of the polymorphisms (17) result in the replacement of an amino acid in the encoded protein product (nonsynonymous; Supporting Table S1). A recent study showed that functional polymorphisms (nonsynonymous) are overrepresented in a set of human SNPs that have a minor allele frequency lower than 6% [27]. Because a selection of inbred strains and substrains as used in this study, does not resemble a random (outbred) population and is limited in the number of individual samples, we used another approach to investigate if there is correlation between SNPs being nonsynonymous and their minor allele frequency. We defined three bins with minor allele frequencies between 15% and 50%, between 5% and 15%, and below 5%, similar as used by Cargill et al. [12]. In line with human functional SNP data, nonsynonymous SNPs appear more frequently in the latter class (47.1%) when compared to synonymous SNPs (27.1%) and noncoding SNPs (37.1%) (see Supporting Table S2 for details).

For estimating the direct effect of SNPs, we checked the impact of all nonsynonymous SNPs on the subsequent protein product by SIFT [21] and PolyPhen [22]. These programs use phylogenetic conservation of protein (domains), chemical properties of the polymorphic amino acids, and predictions on secondary protein structure to estimate the potential effect of a polymorphism on protein function. Although results produced by these prediction programs should be used with care, SIFT correctly distinguishes between deleterious and neutral alleles for 18 out of 22 SNPs in 5 known human disease genes [28]. One of the 17 nonsynonymous SNPs in our study (C3177G of XM_237485 in NCBI database, changing His to Gln) was predicted to be damaging by both methods and 4 others were found to be probably/possibly damaging by PolyPhen (CT1157/1158AC of XM_219840, C1277A of NM_017140, A671G of M36074, G646A of AF130341), but were called tolerant by SIFT, of which two were diagnosed at low confidence by SIFT. The other 11 SNPs were not predicted to be dramatically deleterious for the protein. The five polymorphisms that are predicted to potentially affect protein function were found in the FSH receptor (FSHR), G protein-coupled receptor 50 (Gpr50), dopamine receptor 3 (Drd3), mineralocorticoid receptor (Mr), and melatonine receptor (Mt1), respectively. Interestingly, 4 of these SNPs have minor allele frequencies lower than 6%, but the SNP in Mr is abundantly present with a minor allele frequency of 43.5%. Since these SNPs may impair the function of the protein, the strains harboring them might have aberrant characteristics in processes in which the corresponding genes are involved. Although it is currently difficult to prove that these polymorphisms account for phenotypic differences between the strains, it is clear that (combinations of) genetic polymorphisms in general do influence experimental results. For example, in the human population, the interindividual variability in drug response is a major problem for effective drug treatment and it is believed that this is caused by heritability of certain uncharacterized functional polymorphisms [29]. Pharmacogenomic approaches in model organisms are expected to contribute considerably to the understanding of the function of specific genes in relation to drug response [30]. Characterization of the natural occurring genetic variation at such loci, for example in the many rat inbred strains that are now known, could contribute equally to such understanding.

Acknowledgements

This work was financially supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics (IOP Genomics). We thank Steve Henikoff for help setting up the TILLING technology, R. Malik and V. Guryev for bioinformatic support, and J. Mudde for technical support.

References

1. Steen, R.G. *et al.* (1999) A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res*, 9, AP1-8, insert.
2. Kwitek-Black, A.E. and Jacob, H.J. (2001) The use of designer rats in the genetic dissection of hypertension. *Curr Hypertens Rep*, 3, 12-18.
3. Kacew, S. and Festing, M.F. (1996) Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances. *J Toxicol Environ Health*, 47, 1-30.
4. Thomas, M.A. *et al.* (2003) Phylogenetics of rat inbred strains. *Mamm Genome*, 14, 61-64.
5. Canzian, F. (1997) Phylogenetics of the laboratory rat Rattus norvegicus. *Genome Res*, 7, 262-267.
6. Wang, D.G. *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 280, 1077-1082.
7. Loots, G.G. *et al.* (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, 288, 136-140.
8. Hare, M.P. and Palumbi, S.R. (2003) High intron sequence conservation across three Mammalian orders suggests functional constraints. *Mol Biol Evol*, 20, 969-978.
9. Dubchak, I. *et al.* (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res*, 10, 1304-1306.
10. Ueda, H. *et al.* (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*, 423, 506-511.
11. Kristensen, V.N. *et al.* (2001) High-throughput methods for detection of genetic variation. *Biotechniques*, 30, 318-322, 324, 326 passim.
12. Cargill, M. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*, 22, 231-238.
13. Halushka, M.K. *et al.* (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet*, 22, 239-247.
14. Lindblad-Toh, K. *et al.* (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet*, 24, 381-386.
15. Vignal, A. *et al.* (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol*, 34, 275-305.
16. Oleykowski, C.A. *et al.* (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, 26, 4597-4602.
17. Till, B.J. *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res*, 13, 524-530.
18. Wienholds, E. *et al.* (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res*, 13, 2700-2707.
19. Smits, B.M. *et al.* (2004) Target-selected mutagenesis of the rat. *Genomics*, 83, 332-334.
20. Nickerson, D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res*, 25, 2745-2751.
21. Ng, P.C. and Henikoff, S. (2001) Predicting deleterious amino acid substitutions. *Genome Res*, 11, 863-874.
22. Ramensky, V. *et al.* (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 30, 3894-3900.
23. Berezikov, E. *et al.* (2002) GENOTRACE: cDNA-based local GENOme assembly from TRACE archives. *Bioinformatics*, 18, 1396-1397.
24. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386.
25. Beck, J.A. *et al.* (2000) Genealogies of mouse inbred strains. *Nat Genet*, 24, 23-25.

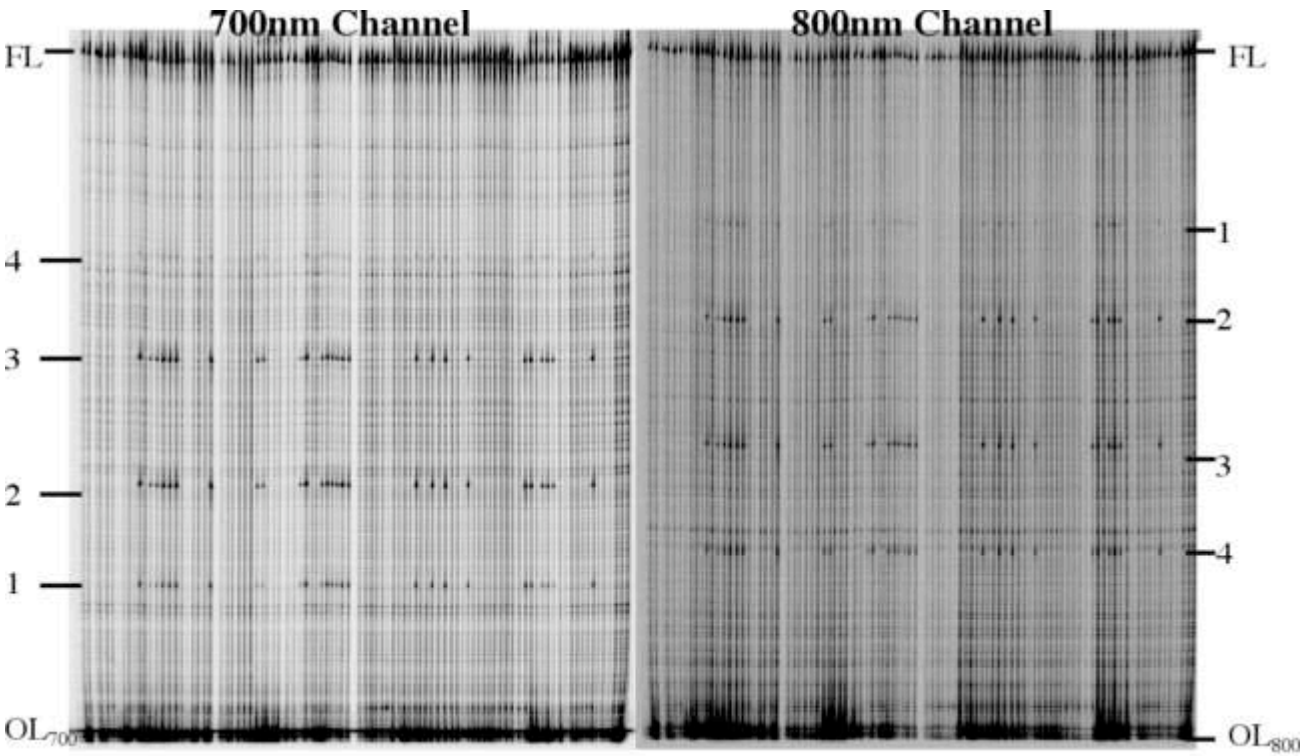
26. Davisson, M.T. (1997) Rules and guidelines for genetic nomenclature in mice: excerpted version. Committee on Standardized Genetic Nomenclature for Mice. *Transgenic Res*, 6, 309-319.
27. Wong, G.K. *et al.* (2003) A population threshold for functional polymorphisms. *Genome Res*, 13, 1873-1879.
28. Ng, P.C. and Henikoff, S. (2002) Accounting for human polymorphisms predicted to affect protein function. *Genome Res*, 12, 436-446.
29. Nebert, D.W. (1999) Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet*, 56, 247-258.
30. Watters, J.W. and McLeod, H.L. (2002) Murine pharmacogenomics: using the mouse to understand the genetics of drug therapy. *Pharmacogenomics*, 3, 781-790.

Website References

http://www.hgsc.bcm.tmc.edu/projects/rat/; Rat whole-genome sequencing project, Baylor College of Medicine.
http://rgd.mcw.edu/strains/; Rat Genome Database, information on rat inbred strains.
http://www.mh-hannover.de/institute/tierlabor/; Institut fur Versuchstierkunde und Zentrales Tierlabor Hannover, additional information on rat inbred strains.
http://cuppen.niob.knaw.nl; Detailed protocols for CEL I-based mutation/SNP detection.
http://primers.niob.knaw.nl; Web-based primer-design application for the design of sets of nested (tailed) oligo's.
http://www.ncbi.nlm.nih.gov/SNP/; dbSNP of NCBI.
http://www.anex.med.tokushima-u.ac.jp/rat/w433-e.html; Biochemical markers in inbred rat strains, Institute for Animal Experimentation, Tokushima, Japan

Supporting Material

Supporting Figure S1: Example of a gel produced by the CEL I based heteroduplex cleavage method. The left and right panels represent the fluorescence at 700 and 800 nm, respectively. Each of the 96 lanes represents a single (sub)strain. The undigested full length PCR product is indicated by FL and unincorporated labeled oligo's by OL. Numbered bars indicate polymorphisms in a part of the coding region of CHRNA3, as compared with the Brown Norway strain. SNP4 is rather weak in the 700 nm channel, but strong in the 800 nm channel, whereas the opposite is observed for SNP1.



Supporting Table S1: Summary of polymorphisms found in rat gene

Gene	NCBI acc.nr.	Coding bp screened	Nonsyn. SNPs	Synon. SNPS	Intronic bp screened	Intronic SNPs	UTR bp screened	UTR SNPs
ABCG5	AF312714	658	0	2	414	0	0	0
ABCG8	AF351785	267	0	0	292	3	571	1
ACHATR	AB075946	281	0	0	74	0	0	0
BDNF	NM_012513	686	0	1	2	0	24	0
CHRNA3	NM_052805	578	1	3	0	0	0	0
CHRNA4	NM_024354	1,391	1	2	14	0	0	0
CHRNA9	NM_022930	743	0	2	267	1	86	0
CHRNB4	NM_052806	992	2	3	60	0	0	0
CYSLT2	AB052661	930	0	1	52	0	0	0
DAX1	X99470	763	0	0	35	1	0	0
DBH	L12407	836	0	0	677	1	325	1
DOPT	M80570	287	0	1	33	0	44	0
DRD1	M35077	1,428	0	0	11	0	60	0
DRD2	NM_022547	286	0	0	234	2	33	1
DRD3	NM_017140	1,252	2	0	1,742	4	139	0
DRD5	NM_012768	1,428	1	2	141	1	0	0
ESR1	NM_012689	337	0	4	96	0	0	0
ESR2	NM_012754	524	0	1	365	1	231	2
FSHR	XM_237485	1,940	1	3	560	2	25	0
GDNF	NM_019139	632	0	0	317	0	29	0
GPCR5	NM_022216	300	0	0	114	0	0	0
GPR6	NM_031806	746	0	0	0	0	0	0
GPR9	NM_053415	665	0	0	0	0	0	0
GPR24	NM_031758	435	0	0	10	0	170	1
GPR50	XM_219840	1,811	1	1	367	0	0	0
GR	M14053	735	0	1	2	0	0	0
GRIN1	NM_017010	437	0	0	473	0	0	0
LHR	AH004953	826	0	2	685	2	30	0
MR	M36074	1,942	1	1	93	0	0	0
MT1	AF130341	680	2	3	23	0	0	0
MYOC	AF289235	1,585	1	3	0	0	0	0
NMU2R	NM_022275	702	1	1	273	0	0	0
NRP1	AF018957	333	0	0	66	0	0	0
NURR1	U72345	932	0	1	907	1	16	0
OR1	U20389	551	0	0	208	0	0	0
P53	NM_030989	680	0	0	360	0	0	0
PGPR	U12184	254	0	0	177	1	377	0
PDYN	NM_019374	615	3	1	21	0	0	0
PKCRX1	AF203907	637	0	0	0	0	0	0
PPAR	M88592	123	0	0	140	0	371	0
PPP3CB	NM_017042	487	0	0	578	2	0	0
PSPLA	D88666	291	0	3	612	0	0	0
REVERBA	M25804	1,252	0	2	281	0	0	0
REVERBAB	U20796	849	0	0	486	0	340	1
RNR1	L08595	863	0	1	109	0	4	0
SEMIII	X95286	457	0	0	160	0	0	0
SEMW	AB002563	1,262	0	1	914	2	0	0
SEMY	AB000817	755	0	0	994	0	0	0
SEMZ	AB000776	735	0	0	646	0	213	2
SHPH	D86580	575	0	1	88	0	0	0
SLC6A2	NM_031343	378	0	0	681	3	0	0
SPF	AF309558	417	0	0	268	1	0	0
SQLE	NM_017136	576	0	1	192	0	21	0
STAR	NM_031558	344	0	0	251	0	0	0
TH	NM_012740	508	0	0	524	1	0	0
Total		**40,977**	**17**	**48**	**16,089**	**29**	**3109**	**9**
SNP frequency		**coding: 1 in 630**		**intron: 1 in 555**		**UTR 1 in 345**		

Supporting Table S2: Minor allele frequency by polymorphism

Allele frequency	Nonsynonymous	Synonymous	Noncoding
< 5%	8 (47.1%)	13 (27.1%)	13 (37.1%)
5% - 15%	4 (23.5%)	12 (25.0%)	5 (14.3%)
> 15%	5 (29.4%)	23 (47.9%)	17 (48.6%)
Total	17 (100%)	48 (100%)	35 (100%)

Minor allele frequencies. All SNPs are grouped into three types. For every SNP type, the minor allele frequencies are classified into low (<5%), medium (5-15%) and high (>15%) [12]. The minor allele frequency is calculated by counting the amount of strains carrying the minor allele, divided by the total amount of strains. The total amount of strains is 62, counting together different substrains of the same inbred strain. If a SNP appears in one of the substrains of an inbred strain, this strain was counted positive.

Supporting Table S3: Information on rat inbred strains

	Full strain name	RGD ID	Strain name	Substrain name	References
1	BN/Cub		BN	Cub	3, 5, 7
2	BN/Gro		BN	Gro	3, 5, 7
3	BN/Gut		BN	Gut	3, 5, 7
4	BN/Han		BN	Han	2, 3, 5, 7
5	BN/Maas		BN	Maas	1, 2, 3, 5, 7
6	BN/Mol	631698	BN	Mol	3, 5, 7
7	BN/NHsdCpb		BN	NHsdCpb	5
8	BN/OlaHsd		BN	OlaHsd	3, 5, 7
9	BN/Orl		BN	Orl	2, 3, 5, 7
10	BN/Rij		BN	Rij	2, 3, 5, 7
11	BN/RijKun		BN	RijKun	5, 7
12	BN/SsN		BN	SsN	3, 5
13	BN-Lx/Cub		BN-Lx	Cub	3, 4, 5, 7
14	BN/Crl		BN	Crl	5
15	LEW/Cub		LEW	Cub	3, 5, 7
16	LEW/Gut		LEW	Gut	3, 5, 7
17	LEW/Han		LEW	Han	2, 3, 5, 7
18	LEW/Ipcv		LEW	Ipcv	3, 5, 7
19	LEW/Maas		LEW	Maas[a)]	1, 2, 3, 5, 7
20	LEW/Nhg		LEW	Nhg	3, 5, 7
21	LEW/NHsdCpb		LEW	NHsdCpb	5
22	LEW/Orl		LEW	Orl	3, 5, 7
23	LEW/SsNHsd		LEW	SsNHsd	5
24	LEW/Ztm		LEW	Ztm	3, 5, 7
25	LEW/Kuv		LEW	Kuv	3, 5, 7
26	F344/Han		F344	Han	1, 2, 3, 5, 7
27	A2/Colle	737949	A2	Colle	3, 4, 5
28	ACI/Kun		ACI	Kun	3, 4, 5, 7
29	ACI/SegHsd		ACI	SegHsd	3, 4, 5, 7
30	ACI/Ztm		ACI	Ztm	3, 4, 5, 7
31	AGUS/OlaHsd		AGUS	OlaHsd	1, 3, 4, 5, 7
32	ALC/Colle		ALC	Colle	3, 4, 5
33	AMORAT/Wsl		AMORAT	Wsl	4, 5
34	AO/OlaHsd		AO	OlaHsd	1, 3, 4, 5, 7
35	ARISTORAT/Wsl		ARISTORAT	Wsl	4, 5
36	AS/Ztm		AS	Ztm	1, 3, 4, 5, 7
37	AUG/OlaHsd		AUG	OlaHsd	1, 3, 4, 5, 7
38	AVN/Orl	10001	AVN	Orl	1, 3, 4, 7
39	BBWB/Mol		BBWB	Mol	1, 3, 4, 5, 7
40	BDE/Han		BDE	Han	1, 2, 3, 4, 7
41	BDII/Han		BDII	Han	1, 2, 3, 4, 5, 7
42	BDIV/Ifz		BDIV	Ifz	1, 3, 4, 5, 7
43	BDIX/Orl		BDIX	Orl	1, 3, 4, 7
44	BDVII/Cub	10006	BDVII	Cub	1, 3, 4, 7
45	BDX/Cub		BDX	Cub	1, 3, 4, 5, 7
46	BH/Ztm		BH	Ztm	1, 2, 3, 4, 7
47	BP/Cub		BP	Cub	1, 2, 3, 4, 5, 7
48	BS/Ztm		Bs	Ztm	1, 3, 4, 5, 7
49	BUF/Han		BUF	Han	1, 3, 4, 5, 7
50	CAP/Kuv		CAP	Kuv	1, 3, 4, 5, 7
51	CHOC/Cub		CHOC	Cub	1, 3, 4, 7
52	COP/OlaHsd	10011	COP	OlaHsd	1, 3, 4, 5, 7
53	DA/Han		DA	Han	1, 2, 3, 4, 5, 7
54	DZB/Gro		DZB	Gro	3, 4, 5, 7
55	E3/Han		E3	Han	1, 2, 4, 5, 7
56	GC/Kun		GC	Kun	1, 3, 4, 5, 7
57	Hooded/Colle		Hooded	Colle	3, 4
58	LE/Han		LE	Han	1, 2, 3, 4, 5, 7
59	LEP/Cub		LEP	Cub	1, 3, 5, 7
60	LH/Ztu		LH	Ztu [b)]	1, 3, 4, 5, 7

61	LOU/CHan	10023	LOU	CHan	1, 3, 4, 5, 7
62	MHS/Gib	10025	MHS	Gib	1, 3, 4, 7
63	MNS/Gib	10028	MNS	Gib	1, 3, 4, 5, 7
64	MW/Hsd		MW	Hsd	1, 3, 4, 5, 7
65	NAR/SaU		NAR	SaU	1, 3, 4, 5, 6, 7
66	NEDH/K	10030	NEDH	K	1, 3, 4, 7
67	OKA/Wsl	10033	OKA	Wsl	1, 3, 7
68	OM/Han	10034	OM	Han	1, 2, 3, 4, 5, 7
69	PAR/Wsl		PAR	Wsl	1, 3, 4, 5, 6, 7
70	PVG/OlaHsd		PVG	OlaHsd	1, 3, 4, 5, 7
71	R/A		R	A	1, 2, 3, 4, 5, 7
72	R/AEurRij		R	AEurRij	4, 5
73	R/AWa		R	AWa	3, 4, 5, 7
74	RHA/Kun		RHA	Kun	1, 3, 4, 7
75	RNU/Mol		RNU	Mol	4, 6
76	SD/A		SD	A	2, 3, 4, 7
77	SD/Rij	10037	SD	Rij	1, 2, 3, 4, 5, 7
78	SDH/Ztu		SDH	Ztu [b)]	1, 3, 4, 7
79	SDL/Ipcv		SDL	Ipcv	1, 3, 4, 5, 7
80	SHR/Maas		SHR	Maas	3, 4, 7
81	SHRSP/Rivm	10039	SHRSP	Rivm	1, 3, 4, 7
82	SPRD/Han		SPRD	Han	1, 2, 3, 4, 5, 7
83	SR/JrIpcv		SR	JrIpcv	1, 3, 4, 5, 7
84	SR/JrMol		SR	JrMol	3, 4, 5, 7
85	SS/JrIpcv		SS	JrIpcv	1, 3, 4, 5, 7
86	U/A		U	A	1, 2, 3, 4, 5, 7
87	WAG/Rij		WAG	Rij	1, 2, 3, 4, 7
88	WAG/OlaHsd		WAG	OlaHsd	3, 4, 7
89	WF/Gut		WF	Gut	3, 5, 7
90	WF/Han		WF	Han	1, 3, 5, 7
91	WF/Mol		WF	Mol	3, 5, 7
92	WF/NHsd		WF	NHsd	3, 5, 7
93	WIST/Nhg	10044	WIST	Nhg	1, 3, 5, 7
94	WKY/Han		WKY	Han	2, 3, 5, 7
95	WOKA/K		WOKA	K	1, 3, 4, 5, 7
96	WOKW/K		WOKW	K	1, 3, 4, 7

1 Otsen et al., 1995. Use of Simple Sequence Length Polymorphisms for Genetic Characterization of Rat Inbred Strains *Mammalian Genome* 6: 595-601

2 Bender et al., 1984. Biochemical Markers in Inbred Strains of the Rat. *Immunogenetics* 19: 257-266

3 http://www.anex.med.tokushima-u.ac.jp/rat/w433-e.html

4 Bonne et al., 2003. Sequencing and chromosomal localization of Fabp6 and an intronless Fabp6 segment in the rat. *Mol Biol Rep* 30: 173-176

5 Heeley et al., 1998. Polymorphisms of the glucocorticoid receptor gene in laboratory and wild rats: steroid binding properties of trinucleotide CAG repeat length variants. *Mammalian Genome* 9: 198-203

6 http://www.mh-hannover.de/institute/tierlabor/genearchiv.html

7 Bender et al., 1994. Genetic characterization of inbred strains of the rat (Rattus norvegicus). *Journal of Experimental Animal Science* 36:151-165

a) Maas, labcode not registered at ILAR: Dr. J.E.M. van den Bogaard, Central Animal Unit, Univeristy of Limburg, PO Box 616, 6200 NL Maastricht, The Netherlands

b) Ztu, labcode not registered at ILAR: Dr. Hp. Schnappauf, Zentrale Tierversuchsanlage Oberberghof der Universitat Ulm, D-89060, Ulm, Germany

CHAPTER 5

Efficient Single Nucleotide Polymorphism Discovery in Laboratory Rat Strains using Wild Rat-derived SNP Candidates

Adapted from BMC Genomics (2005)

Efficient Single Nucleotide Polymorphism Discovery in Laboratory Rat Strains using Wild Rat-derived SNP Candidates

Bart M. G. Smits[1], Victor Guryev[1], Dimphy Zeegers[1], Dirk Wedekind[2], Hans J. Hedrich[2], Edwin Cuppen[1*]

[1]Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[2]Institute for Laboratory Animal Science, Hannover Medical School, Carl-Neuberg-Str.1, 30625 Hannover, Germany

[*]Corresponding Author

The laboratory rat (*Rattus norvegicus*) is an important model for studying many aspects of human health and disease. Detailed knowledge on genetic variation between strains is important from a biomedical, particularly pharmacogenetic point of view and useful for marker selection for genetic cloning and association studies. We show that Single Nucleotide Polymorphisms (SNPs) in commonly used rat strains are surprisingly well represented in wild rat isolates. Shotgun sequencing of 814 Kbp in one wild rat resulted in the identification of 485 SNPs as compared with the Brown Norway genome sequence. Genotyping 36 commonly used inbred rat strains showed that 84% of wild rat alleles are also present in a representative set of laboratory rat strains. We postulate that shotgun sequencing in a wild rat sample and subsequent genotyping in multiple laboratory or domesticated strains rather than direct shotgun sequencing of multiple strains, could be the most efficient SNP discovery approach. For the rat, laboratory strains still harbor a large portion of the haplotypes present in wild isolates, suggesting a relatively recent common origin and supporting the idea that rat inbred strains, in contrast to mouse inbred strains, originate from a single species, *R. norvegicus*.

Background

Genetic variation exists between individuals (or strains) of all organisms and it makes up the genetic basis for phenotypic differences between individuals. In addition, genetic variation functions as a valuable resource for mapping phenotypic traits in model organisms. Single Nucleotide Polymorphisms (SNPs) are the most abundant form of genetic variation and therefore dominate high-resolution genetic mapping strategies. Moreover, numerous well-performing high-throughout SNP detection technologies have been developed, like oligonucleotide array-based technology, mass-spectrometry-based technology (MALDI-TOF), and sequence-based technology (pyrosequencing, DHPLC) [1], which makes automated SNP detection favored above the more labor-intensive detection of microsatellite markers [2].
Since the availability of its genome, the laboratory rat is gaining influence as a genetic model organism [3]. In addition, over 200 well-characterized inbred strains that are models for a wide

variety of human diseases are available [4,5]. However, the availability of genetic tools, like a dense genome-wide SNP marker set, is still subordinate compared to other commonly used model organisms. This is illustrated by the number of entries in dbSNP, the central SNP repository of NCBI (http://www.ncbi.nlm.nih.gov/SNP/): the amount of human (>10,000,000), chicken (>3,000,000), and mouse (>500,000) entries surpass the amount of rat entries (>43,000) spectacularly. In search for rat SNPs, experimental [6,7] and computational [8] approaches have been employed, but these efforts primarily resulted in SNPs associated with coding regions. For genetic mapping purposes, a much denser marker set, preferentially equally distributed over the genome, is required.

Laboratory rat strains are thought to be established from a limited number of founder animals originating from a domesticated wild population [9,10]. The value of inbred strains emanates from the close genetic uniformity that facilitates phenotyping and genotyping. In principle, inbred strains are selectively bred for certain traits from a genetically diverse pool, comprising diverse genetic information about the trait. However, since many of the current rat strains were derived from common ancestral stocks and simply inbred to increase genetic uniformity, inbred strains clearly share alleles [11]. Although such simplified models are essential for biomedical research, modulating effects on the clinical manifestation of a trait resulting from genetic heterogeneity in a population can only be studied to a limited extent in F1 hybrids. The use of a carefully chosen selection of inbred strains may address this issue, but the choice depends on knowledge on the relationship between the strains and hence the degree of genetic variation. Alternatively, wild-derived strains may be good alternatives to introduce sufficient genetic variation in laboratory experiments [12,13].

Based on a preliminary observation that SNPs from laboratory rat strains are frequently detected in wild-derived samples, we developed a wild rat-based SNP discovery approach. The method consists of shotgun sequencing of a wild rat-derived genomic library followed by comparison with the published rat genome (strain Brown Norway). Genotyping commonly used rat strains for newly identified SNPs revealed that in 84% of the cases, the wild rat SNP-allele (and in 87% of the cases including all genetic variation) is also represented in one or more laboratory strains. A user-friendly webtool allows exploration of the genetic variation between any arbitrary combinations of two strains that were used in this study, making all information directly available for experimental use.

Methods

Genomic DNA isolation, shotgun library construction

Wild rat 1 (*Rattus norvegicus*) was caught in the canals of Utrecht and kindly provided by the Pest Control Service of the City of Utrecht (Utrecht, The Netherlands). Wild rat 2 was trapped in Gassel, a village located approximately 100 km south-east of Utrecht and was kindly provided by Tien Derks (Gassel, The Netherlands). Wild rat 3 was caught in a basement in Amsterdam, located 50 km north of Utrecht and kindly provided Romke Koch (Amsterdam, The Netherlands). Rat strains BN/Crl and Crl:Wistar (outbred) were obtained from Charles River The Netherlands. Liver samples of commonly used rat strains ACI/Ztm, BDE/Ztm, BDII/Ztm, BDIX/Ztm, BDV/Ztm, BH/Ztm, BN/Ztm, BS/Ztm, DA/Ztm, E3/Ztm, F344/Ztm, LE/Ztm, LEW/Ztm, LOU/CZtm, MNS/Ztm, MWF/Ztm, NAR/Ztm, OM/Ztm, PAR/Ztm, R33/Ztm, WC/Ztm, WF/Ztm, WKY/Ztm were provided by D.W. (Hannover Medical School, Germany) and liver samples of strains AO/OlaHsd, AUG/OlaHsd, BUF/SimRijHsd, COP/Hsd, DA/OlaHsd, LUDW/OlaHsd, PVG/OlaHsd, RP/AEurRijHsd, SHR/NHsd, SR/JrHsd, SS/JrHsd, WAG/RijHsd and 2 individuals of Hsd:SD (outbred) were kindly provided by Harlan (Horst, The Netherlands). Samples were lysed overnight in 20 ml lysis buffer, containing 100 mM Tris (pH 8.5), 200 mM of NaCl, 0.2% of SDS, 5 mM of EDTA, and 100 μg/ml of freshly added Proteinase K at 55°C under continuous rotation. Tissue debris was spinned down for 20 min at 10,000 x g and supernatant was transferred to a fresh tube. DNA was purified by phenol-chloroform extraction and precipitated by adding

an equal volume of isopropanol, mixing and centrifugation for 20 min, 10,000 x g at 4°C. The supernatant was removed by gently inverting the tube and the pellets were washed with 70% ethanol and dissolved in 1000 µl water. The concentration was measured by Optical Densitometry at 260 nm.

Wild rat-derived genomic library construction and shotgun sequencing
Sheared wild rat-derived genomic DNA of approximately 1-2 Kbp in size was cloned into the *Sma*I-site of pUC19. Fractions of the glycerol stock of the transformed library (*E. coli* DH10B) were plated on LB-plates containing 50 µg/ml ampicilin, 200 µg/ml IPTG, and 0.01% X-gal for standard blue/white screening on inserts. White colonies were picked in 20 µl water. Lysis occurred at 95°C for 10 min. 5 µl of 5x diluted lysate was used for the PCR reaction. For PCR, universal M13 primers were used, namely M13F: TGTAAAACGACGGCCAGT, M13R: AGGAAACAGCTATGACCAT. PCR, sequencing and cycling conditions were similar as for strain genotyping, described below. Sequencing was performed using universal M13 primers.

PCR conditions for strain genotyping
PCR was carried out using a touchdown thermocycling program (92°C for 60 sec; 12 cycles of 92°C for 20 sec, 65°C for 20 sec with a decrement of 0.6°C per cycle, 72°C for 30 sec; followed by 20 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 30 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems) and contained 30-50 ng genomic DNA, 0.2 µM of each forward primer and 0.2 µM of each reverse primer, 400 µM of each dNTP, 25 mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate pH 8.7 and 0.2 U Taq Polymerase in a total volume of 10 µl.

Sequencing reactions, purification, and analysis
PCR products were diluted with 25 µl water and 1 µl was directly used as template for the sequencing reactions. Sequencing reactions, containing 0.25 µl BigDYE (v3.1; Applied Biosystems, Foster City, CA, USA), 3.75 µl 2.5x dilution buffer (Applied Biosystems) and 0.4 µM universal M13 primer in a total volume of 10 µl, were performed using cycling conditions recommended by the manufacturer (40 cycles of 92°C for 10 sec, 50°C for 5 sec and 60°C for 120 sec). Of sequencing products, 5 µl was purified by ethanol precipitation in the presence of 40 mM sodium-acetate and analyzed on 96-capillary 3730XL DNA analyzers (Applied Biosystems), using the standard RapidSeq protocol. Sequences were analyzed for presence of heterozygous mutations using PolyPhred [14], followed by manual inspection of the polymorphic positions.

Automation
All PCR and sequencing reactions were set up on a Tecan Genesis RSP200 liquid handling workstation, with a robotic and an 8-channel pipetting arm, an integrated 96-channel pipetting head (TEMO96, Tecan), and four integrated dual-384 well PCR blocks (Applied Biosystems).

Mapping of shotgun reads and SNP discovery
Shotgun reads were assigned to positions in the RGSC 3.1 rat genome assembly using BLAT search [15]. Shotgun reads that complied with our mapping criteria, namely those having at least 80 identical bp for the best hit and no more than 60 identical bp for second blat hit were retained for further analysis. BLAST nucleotide sequence alignments between shotgun read and corresponding genomic segment were used for discovery of single base variations (including single base indels). A site was treated as polymorphic only in the case when it has identical 5'- and 3'-flanks of at least 5 bp. A custom designed web-application was employed for manual chromatogram inspection and confirmation of a correct shotgun base-call for every polymorphic SNP locus. Primer design for resequencing was performed using a local web-interface (http://primers.niob.knaw.nl) to the primer3 program [16].

Simulation model for wild rat-based SNP discovery
To estimate the number of SNPs to be discovered by the wild rat resequencing approach we performed computer simulations using the observed sample-specific polymorphism frequencies and the rat genome size of 2.48 Gbp as an input. We used a Monte-Carlo method for the placement of N 400-bp shotgun reads to the genome and calculated the total size of genome covered by N shotgun reads. To obtain a conservative estimate by assuming low heterozygosity in wild-derived strain the estimate of number of SNPs is given by product of covered genome size and polymorphism rate.

Results

Wild rat-based SNP discovery

It is generally believed that commonly used rat strains originate from a wild-derived founder population of limited size [9]. To examine whether polymorphisms found in laboratory strains are still represented in individuals of the wild population, we typed two wild-derived samples for confirmed SNPs of the CASCAD database [8]. Interestingly, about 53% of SNPs (n=147), which were confirmed to exist in laboratory strains, were also identified as a polymorphic allele for wild 1, wild 2 or both as compared to the BN genome sequence (not shown). Hence, a pre-selection of highly likely candidate SNPs could potentially be made by genotyping wild individuals and comparing the sequences to the rat genome sequence (Brown Norway).
Accordingly, we performed random shotgun sequencing on a genomic library of a wild rat (wild 1). We generated shotgun traces (814 Kbp) by bidirectional sequencing of about 1,600 colonies (Table 1). 85.5% of the reads (2545/2975; Table 1) could be mapped to a unique location in the Brown Norway rat genome using BLAT [15], resulting in the automated identification of nearly 5,000 ambiguous nucleotide positions (potential polymorphisms). Manual inspection of the sequencing reads reduced this set of potential polymorphisms to a set of 746 real SNPs and 122 indels. The average SNP rate between BN (BN/SsNMcw; genome sequencing project) and this single wild rat is estimated to be about 1 per 900 bp and, hence, discovery of a novel SNP can be expected every second shotgun read. A subset of the discovered SNPs was verified and genotyped in 36 commonly used strains (including BN). To this end, we designed primers for 451 SNP-containing amplicons (about 300 bp) of which 416 (92.2%) were successfully read by unidirectional sequencing of the PCR products, resulting in roughly 119 Kbp high quality sequence per strain or individual (Table 1).

Table 1: Statistics on shotgun sequencing of the wild rat-derived genomic library

picked colonies	1632
readable sequence reads / sequenced bases	2975 / 814,440
uniquely mapped (BLAT) reads / bases	2545 / 768,683
ambiguous positions	4902
strong candidates after manual inspection	868 (746 SNPs + 122 indels)
successfully read / amplicons designed*	416 / 451 (~1.65 candidate SNP / amplicon) (92.2%)
amplified bases per strain or wild individual	118,971

* Amplicons are designed for the 746 SNP candidates.

Wild rat-derived SNP characteristics

The verification of 746 candidate SNPs by amplicon-based resequencing in 36 inbred rat strains and three wild-derived samples (wild 1, 2, and 3) revealed 960 polymorphisms, consisting of 90 indels, seven 2-bp substitutions, one 3-bp substitution, one 5-bp substitution, and 861 SNPs, of which only one was tri-allelic. The amplicons are randomly distributed over the genome (Fig. 1). We observed heterozygous positions in the outbred strains, but unexpectedly some were also found in the inbred strains (see Additional Materials BMC Genomics website (http://www.biomedcentral.com/bmcgenomics/) or dbSNP (http://www.ncbi.nlm.nih.gov) for detailed allele information). For our analysis, we considered these loci to be polymorphic as compared to the BN genome sequence.

From the 746 shotgun-based candidate SNPs, 685 were located in the 416 amplicons that worked, and 485 (71%) were reconfirmed by resequencing (shotgun-based; Table 2). Strikingly, for 408 (84%) of the confirmed SNPs, the wild rat allele is also present in one or more commonly used strains, with only 36 (7.4%) being specific to BN (Table 2). Of the remaining 77 (16%) SNPs, wild rat alleles are not present in any of the 36 selected strains and could be considered wild rat-specific. These results illustrate that shotgun sequencing one wild individual efficiently identifies shared polymorphisms among commonly used rat strains.

While genotyping by resequencing, 358 novel SNPs were discovered that were not identified in the shotgun sequencing experiment (genotyping-based; Table 2). About 39% (139) of this set can be accounted for by differences in the sequence coverage between the shotgun reads and the resequencing genotyping reads (Table 2), whereas the remaining part of this set is strongly biased towards SNPs that are not polymorphic between BN and wild rat 1 and thus could not have been discovered in the shotgun experiment. Interestingly, about 37% of the newly discovered SNPs are polymorphic between the shotgun sequenced wild rat and any of the inbred strains (Table 2). When considering all SNPs that are polymorphic in the set of 36 commonly used laboratory strains, of the majority (66%) the wild rat allele is found back in one of the strains (total; Table 2) and this percentage increases only slightly (70%) when two other wild individuals (wild 2 and 3) are included in the analysis. This indicates that wild rat-based SNP discovery is already highly efficient using a single wild sample.

Based on the genotyping results, the SNP rate between BN and the shotgun sequenced wild rat (wild 1) is 1 SNP per 190 bp (626 SNPs/119 Kbp). The SNP rate within the 36 rat strains, including BN, is 1 in 158 (Table 2; 45+204+505 SNPs/119 Kbp) and the SNP rate in the entire experiment, including the wild rat (wild 1), BN, and the other strains is 1 in 141 bp (Table 2; 843 SNPs/119 Kbp). To compare wild rat inter-individual variation with the inter-strain variation for commonly used inbred strains, we calculated the number of SNPs that are polymorphic when comparing arbitrary combinations of 3 strains. Genotyping of 861 SNP positions in the three wild rats resulted in 438 polymorphic positions, whereas the most polymorphic combination of inbred strains in this experiment (BN, BH, and SHR) yielded 427 SNPs. This indicates that three random, but potentially related, Dutch wild rats are about equally polymorphic as three carefully selected inbred strains. Inclusion of wild isolates from other locations worldwide may increase the efficiency of the SNP discovery approach.

Table 2: SNP discovery results

	shotgun-based	genotyping-based (only wild 1*)	genotyping-based (wild 1, 2, 3*)	total (only wild 1*)	total (wild 1, 2, 3*)
BN specific	36 (7.4 %)	9 (2.5 %)	7 (1.9 %)	45 (5.3 %)	43 (5.0 %)
wild specific	77 (15.9 %)	12 (3.4 %)	30 (8.0 %)	89 (10.6 %)	107 (12.4 %)
in 35 strains, not in wild	0	204 (57.0 %)	156 (41.5 %)	204 (24.2 %)	156 (18.1 %)
in 35 strains, shared with wild	372 (76.7 %)	133 (37.1 %)	183 (48.7 %)	505 (59.9 %)	555 (64.5 %)
total	485 (100 %)	358 (100 %)	376 (100 %)	843 (100 %)	861 (100 %)

* By genotyping two other wild individuals (wild 2 and 3), additional polymorphisms were identified, which could not have been found by shotgun sequencing only wild 1.

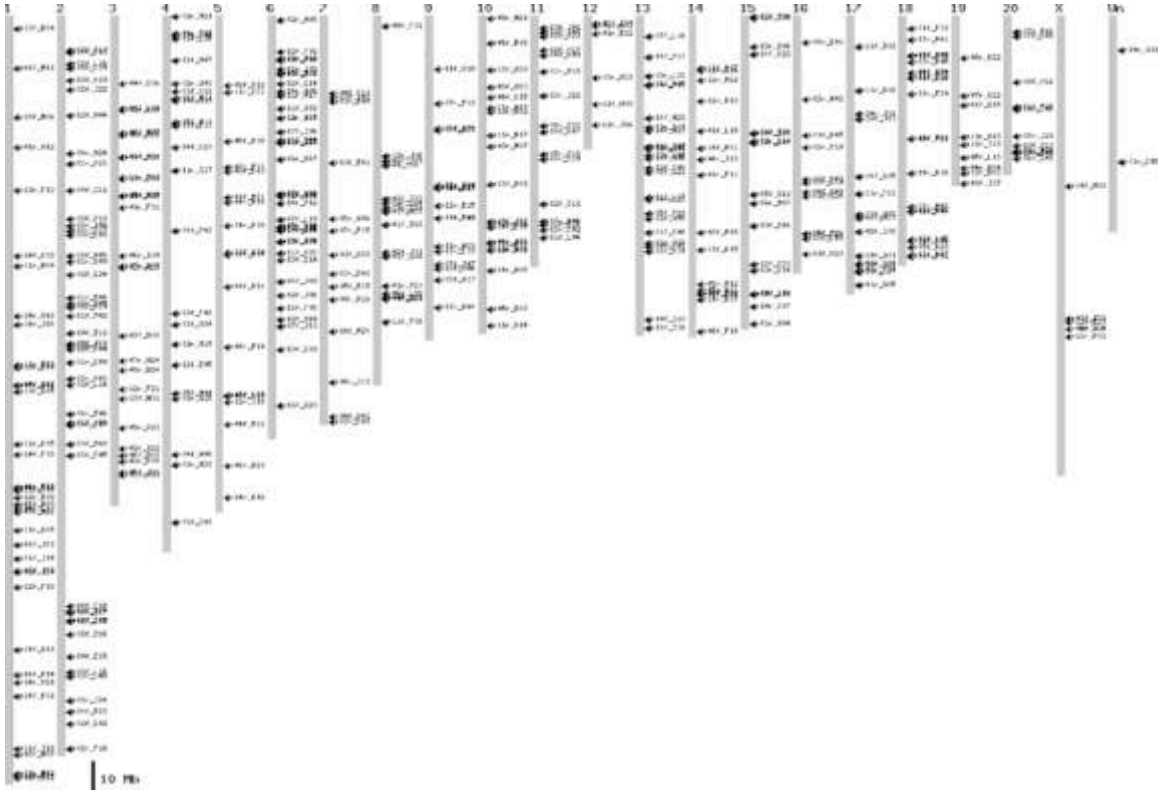


Figure 1: Distribution of amplicons (451 loci) designed for verification and subsequent genotyping of candidate shotgun-based SNPs in 36 commonly used inbred strains.

Intraspecific phylogenetic network

Relationships among different rat strains have been determined previously by phylogenetic tree reconstruction based on microsatellite markers [17,18]. However, intraspecific relationships for laboratory strains are often very challenging to determine, due to small genetic distances and complex gene flow. The resulting multitude of plausible trees is best expressed by a network, which displays alternative potential evolutionary paths in the form of cycles [19]. We used Network software (v4.111 Reduced-Joining, http://fluxus-engineering.com) to construct a spatial network, based on 861 SNP markers in 36 rat strains and three wild rat individuals (Fig. 2). The three wild individuals are grouped together, possibly due to the geographic and possibly genetic relation between the samples, but in accordance with the last paragraph of the previous section, they appear relatively unrelated as compared to the set of inbred strains.

The majority of the SNPs (485 of 861) was selected for being polymorphic between wild 1 and BN. As a result, different BN substrains (BN/Ztm, BN/Crl), depicted as a double-sized end node because of high similarity, and different wild rat individuals (wild 1, wild 2, and wild 3) are grouped together as the outliers. Several strains that are known to be closely related (source RGD-strains: http://rgd.mcw.edu/strains/) are also grouped together, like F344 and LEW or SS and SR. Interestingly, WKY is also an outlier, indicating that besides BN, this strain can be utilized as an alternative mapping strain. WKY is already commonly used as a normotensive control strain in genetic mapping of blood pressure quantitative trait loci [20]. WKY is known to be closely related to SHR and these strains are indeed grouped together (Fig. 2). Additionally, BDII and BDIX are related and BDE is an RI strain from E3. These strain combinations are also grouped together. Wistar is contributing to a large subset of these strains, like WKY, WC, BDII, MWF, LEW, and WF, which contributes to the complexity of the network structure.

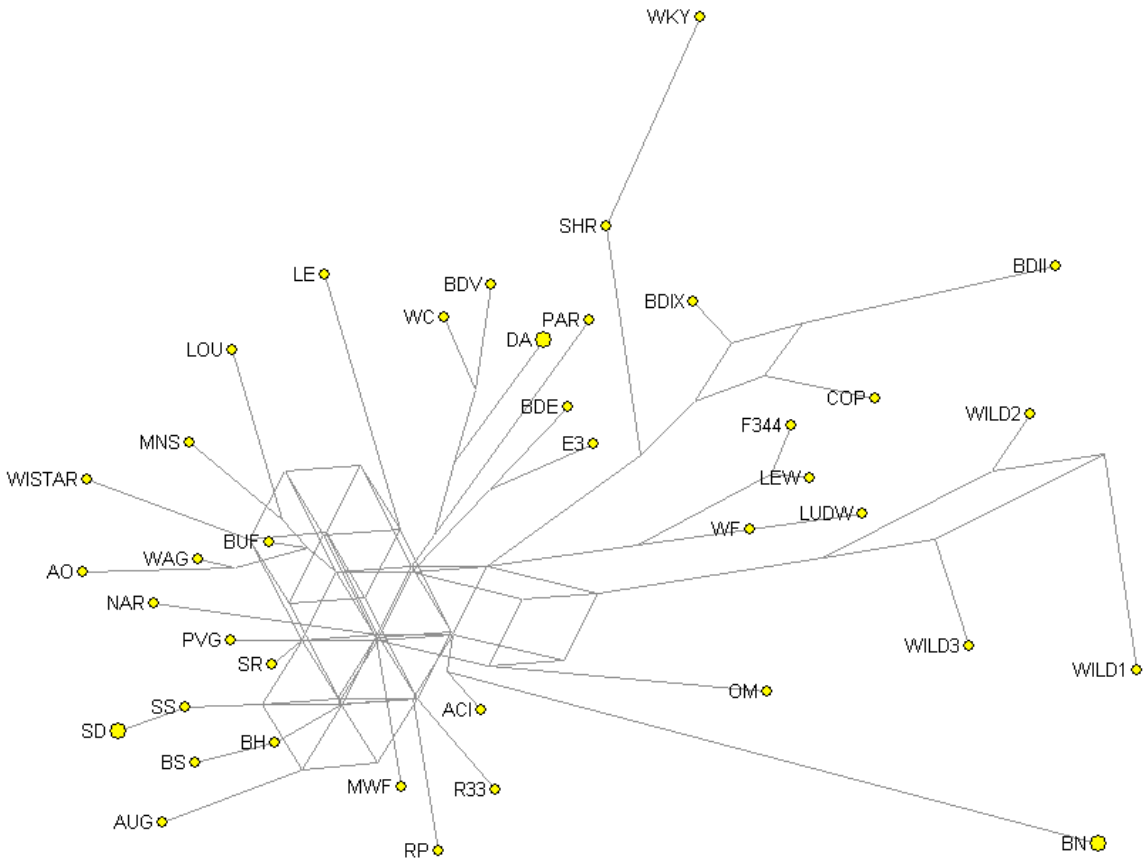


Figure 2: Strain relationships in a network structure.
End nodes (dots) represent strains. Some end nodes are double-size, meaning that they are supported by two samples. Interconnecting nodes where lines come together, represent a possible precursor.

Data availability

The use of genetic markers for mapping traits in rat strains has been exploited for long time already. Current marker sets in rats are mostly limited to microsatellites [21,22], which are not abundantly available and are commonly detected in a more laborious way than SNPs. In this study, we have determined a total of about 35,000 genotypes (about 960 loci in 36 inbred strains), out of which the vast majority are SNPs. This data is accessible via a versatile webtool (http://cascad.niob.knaw.nl/snpview). Pairs of strains of interest can be selected and explored on presence of verified genetic variation. Besides a graphical representation of the location of the SNPs on a genome map, primer sequences that were successfully used in our experiments are also provided. In a pairwise comparison matrix (Table 3), we plotted the absolute number of polymorphic positions for each of the (sub-)strains or individuals used. Interestingly, different substrains of an inbred strain also are polymorphic on some positions (e.g. BN/Crl and BN/Ztm: 4 polymorphisms), in line previous observations [7].

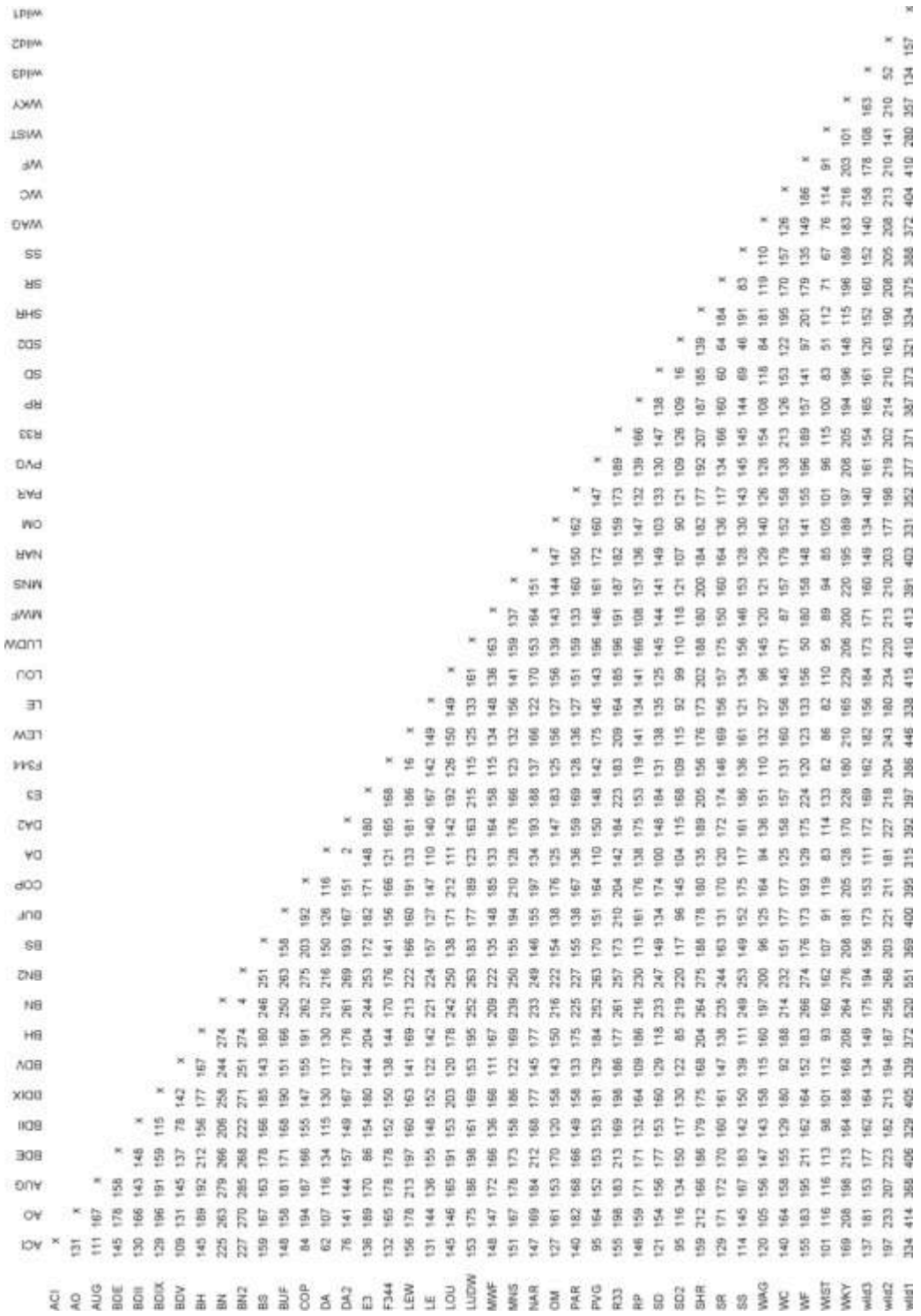


Table 3: Absolute number of polymorphic positions in a pairwise comparison matrix.
The matrix is built from genotyping data of 960 polymorphisms in 36 strains and three wild individuals. Two inbred strains are represented by two substrains (BN and DA) and outbred SD is represented by two individuals from different stocks. Sets of polymorphisms, including a graphical representation, can be retrieved from http://cascad.niob.knaw.nl/snpview.

Simulation experiment wild rat-based SNP discovery

To get insight in the benefits of using wild rats in SNP discovery studies, we simulated larger scale experiments based on the results obtained in the experiments described above. Shotgun sequencing of 814 Kbp resulted in the identification of 485 SNPs. For 408 of those, the wild rat allele was also represented in laboratory rat strains and is thus of interest for research purposes. The maximum amount of SNPs that can be discovered by fully sequencing this single rat is calculated by multiplying the SNP frequency (408/814,440) with the rat genome size (2,48 Gbp), which is 1,252,911 SNPs. Since none of our shotgun reads were overlapping, we can calculate the relation between shotgun sequencing reads of the wild rat and the amount of SNPs that will be found by scaling up this methodology, assuming random distribution of 400 bp shotgun reads over the genome (Fig. 3a). One million shotgun reads of a single wild rat would already result in the discovery of 200,000 novel SNPs that are polymorphic in commonly used rat strains. This simulation indicates that a relatively small sequencing effort could potentially result in a vast expansion of the amount of genetic variation for the rat.

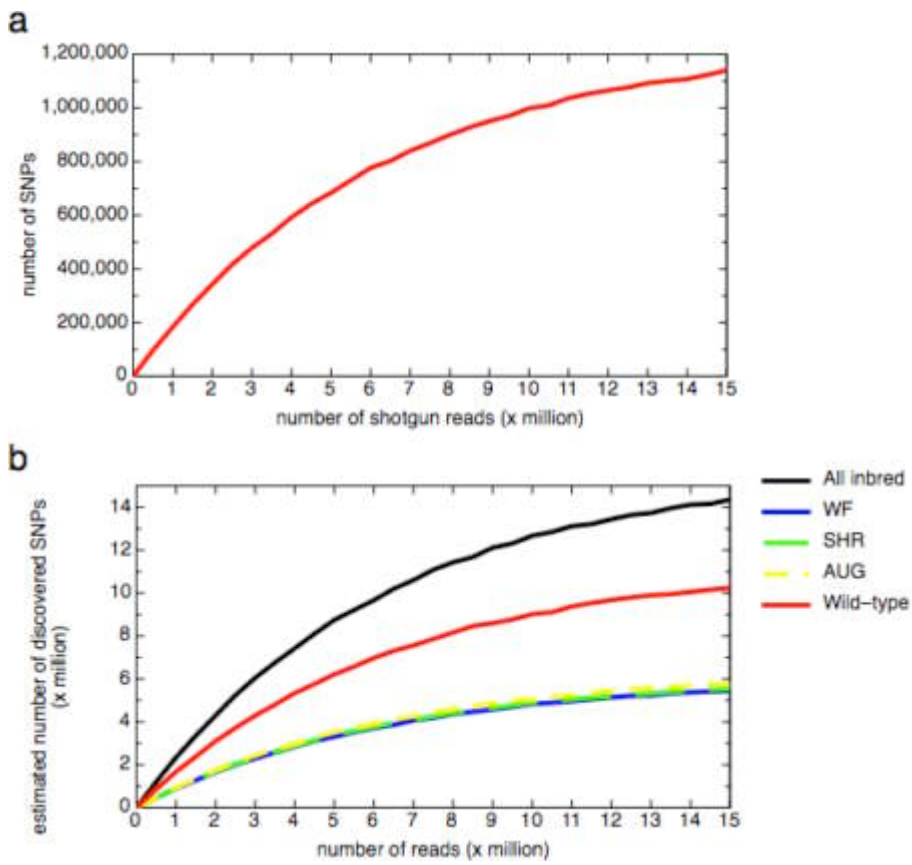


Figure 3: Simulation experiments.
a) Simulation of wild rat-based SNP discovery experiment. Simulation is based on the discovery of 485 SNPs between wild 1 and BN in 814 Kbp of shotgun sequence. For 408 of those, the wild rat alleles is found back in one or more inbred strains. The relation between generation of randomly distributed 400 bp shotgun reads and estimated number of newly discovered SNPs is plotted. **b)** Simulation of SNP discovery experiment, using carefully selected (most polymorphic compared to BN) rat strains (SHR, AUG, and WF) or all rat strains, in comparison with wild rat-based SNP discovery. Simulation is based on 539, 304, 292, 287, and 754 SNPs for wild 1, AUG, SHR, WF, and all strains respectively, in 119 Kbp of genotyped sequence.

Because shotgun sequencing was only done in the wild rat 1, we cannot make a direct comparison between wild rat-based SNP discovery and SNP discovery based on rat strains separately. However, a similar simulation experiment can be performed by treating the genotyping resequencing as shotgun reads. For wild 1, this would result in the identification of 577 SNPs as compared to the BN genome sequence. For 539 of those, the wild rat allele is found back in one of the inbred strains. For the combination of three strains most polymorphic as compared to BN in this experiment, the latter number would be 304, 292, and 287 for AUG, SHR, and WF, respectively. Simulations based on these numbers show that it requires nearly two times as much shotgun sequencing in different inbred strains separately to discover the same amount of SNPs that can be found using the wild rat shotgun sequencing approach. It should be mentioned that parallel shotgun sequencing of all 36 inbred strains until saturation has the potential to yield 1.6 times as many SNPs as compared to the wild-derived approach (Fig. 3b). An advantage of using inbred strains for SNP discovery is that the genotype of the strain is immediately known. Nevertheless, reconfirmation of the SNP or genotyping of other strains of interest may be necessary anyway, minimizing the relevance of this advantage.

Discussion

An increase in the amount of documented genetic variation for the rat will be essential to allow for high-resolution genetic mapping of the many inherited traits that have now been described for a wide variety of rat inbred strains. In addition, insight into genetic variation between rat strains provides valuable information on genetic relationships between strains, which can be instrumental to dissect the genetic basis of phenotypic differences. The wild rat-based shotgun sequencing method described here provides an efficient approach to generate such a dense map of genetic variation. To be able to benefit from haplotype-based mapping approaches [23-26] a high marker density is needed to first reliably define haplotype blocks in strains of interest [27]. For the mouse, it has recently been announced that 15 inbred strains will be fully resequenced to achieve this goal [28]. With extreme dense genotype maps, it may even become possible to clone traits by haplotype-based *in silico* mapping [23], but to achieve this, it is estimated that complete sequences of over 50 strains are needed [27]. Although densities needed for these approaches are not reached, we do show here that wild rat-based SNP discovery is potentially much more effective than shotgun sequencing different inbred strains. We propose that the most effective SNP discovery strategy for the rat would be one based on shotgun sequencing of a single wild-derived sample and subsequent low-cost high-throughput genotyping of the resulting candidates in the laboratory strains of interest. Many other model organisms are currently undergoing full coverage sequencing and SNP discovery in these organisms will become increasingly important, especially for those organisms that are selectively bred for specific traits, such as cow and pig. Pilot experiments using for example wild-derived swine samples could be performed to test whether it is eligible to efficiently transfer the wild isolate-based SNP discovery strategy to other organisms.

Our results do provide insight in the genetic descent of the laboratory rat. It is generally accepted that current rat strains underwent two major genetic bottlenecks. First, they originate from a small founder population of domesticated wild rats and second, they were selectively inbred to obtain homogeneity [10]. The three Dutch wild rats used in this study are potentially relatively closely related as compared to wild rats from different parts of the world, but the genetic variation between them is mostly larger than or sporadically equal to any combination of three inbred strains, indeed suggesting the existence a common genetic bottleneck for laboratory strains. In addition, the laboratory rat does not show an extensive polymorphism rate in the MHC (major histocompatibilty complex) as compared to other species [29], like human, cattle etc. Cramer et

al. has analyzed the MHC of wild rats and compared the data with those from inbred strains [30]. In line with our observation, there were not many new haplotypes.

We observed that wild rat genetic variation is to a large extent represented in the inbred strains, which is in sharp contrast to genetic variation in wild-derived mouse strains that is mostly unique [31]. Contrary to classical mouse inbred strains, where multiple subspecies contribute to the genetic make-up [12,32] and recent mouse strains, derived from different *Mus* species [33], laboratory rat strains are most likely descending from a single rat species, *Rattus norvegicus* [9].

An independent study using 42 microsatellites in German and Japanese wild-derived samples showed that the genetic profiles were quite divergent, partially owing to different geographic locations [34]. Our study involved only Dutch wild rats, suggesting that the inclusion of wild rats from different parts of the world could result in even more efficient SNP discovery, although it also remains to be demonstrated what proportion of the additional discovered alleles is present in the inbred strains and if a geographic bias for this exists.

When multiple SNPs are present per locus/amplicon, independent haplotypes can be discerned. The genetic variation identified here is mostly organized in a limited amount of haplotypes per locus (Table 4). Theoretically, an amplicon containing two or three SNPs can be represented by four and eight haplotypes, respectively, but in our dataset the vast majority of amplicons harboring multiple SNPs is represented by only two or three haplotypes (Table 4). Again, these observations suggest the existence of a common and small founding population with very limited haplotype diversity and/or a very narrow genetic bottleneck before inbred strain selection. The observed small genetic basis in a wide selection of laboratory rat strains does not mimic genetic variation in the human population and as a result, studies and pharmacological tests in rat models neglect potential modulatory effects caused by genetic variation. Although the use of F1 crosses and mosaic populations [35] could address this issue, our data suggests that wild-derived rats may be very useful to this end, since a large amount of all genetic variation present in a large selection of inbred strains, is already represented in a limited number of individuals. Therefore, it would be very interesting to investigate genetic variation in recently domesticated inbred [36] and outbred rats such as wild-type Groningen rats (WTG) [37]. Alternatively, careful selection of inbred strains based on genotyping data and subsequent random breeding may also expose the wild side of laboratory rats.

Table 4: Haplotype analysis in 36 strains for all SNP-containing amplicons

	number of haplotypes											
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
2 SNPs	46	57	8									
3 SNPs	11	26	8	3	0	0	0					
4 SNPs	4	11	5	3	1	1	0	0	0	0	0	
5 SNPs	1	3	3	2	0	0	0	0	0	0	0	
6 SNPs	1	1	1	1	1	0	0	0	0	0	0	
7 SNPs	1	0	1	0	2	0	0	0	0	0	0	
8 SNPs	0	0	0	0	0	0	0	0	0	0	0	
9 SNPs	0	0	0	0	0	0	0	0	0	0	1	
10 SNPs	0	0	0	0	0	0	0	0	0	0	0	
11 SNPs	0	0	0	0	1	0	0	0	0	0	0	
total	64	97	27	9	5	1	0	0	0	0	1	204*

*) Total number of amplicons that contains at least two SNPs. Amplicons containing no SNPs or only indels were excluded from this analysis. Amplicons containing 1 SNP are also excluded, since two-state SNPs always give rise to two haplotypes.

Conclusions

We describe a SNP discovery platform for the rat that is based on two steps. First, candidate SNPs are discovered by shotgun sequencing a wild rat, followed by genotyping laboratory strains of interest. We show that 84% of alleles in wild rats as compared to the sequenced Brown Norway rat genome are also represented in a set of 36 laboratory strains. Hence, the approach described here would be an efficient strategy for the discovery of novel informative SNPs in the laboratory rat. Inclusion of other wild samples, preferably from different locations in the world could result in an even more effective SNP discovery platform, as the three wild rats in our study, caught in relative close vicinity to each other, were already more polymorphic than the most polymorphic combination of carefully selected inbred strains. Based on the more than 34,000 genotyping datapoints obtained in this study, we postulate two things. First, laboratory rats originate from a single rat species, and inbred stains are relatively closely related with a limited number of haplotypes, reflecting known genetic bottlenecks in strain establishment. Second, wild rats have the potential to represent the degrees of genetic variation as present in the human population much more efficiently than a random selection of inbred strains. This makes them or wild-derived strains potentially well-suited for studying modulatory effects of genetic background variation on specific phenotypes, such as behavior or responses to drug treatment.

Author's Contributions

BMGS contributed to the production of the results, supervised the ongoing of the study, and drafted the manuscript. VG contributed to the computational support of the results, and contributed to the writing of the manuscript. DZ contributed to the production of sequencing reads and initial analysis of the results. DW contributed to the preparation of samples for the study and revised the manuscript. HJH participated in the interpretation of the results and revision of the manuscript. EC outlined and supervised the study, and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Harlan (Horst - Netherlands), the Pest Control Service of the City of Utrecht (Utrecht, The Netherlands), Romke Koch (Amsterdam, The Netherlands), and Tien Derks (Gassel, The Netherlands) for kindly providing rat tissue samples. This work was supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics.

References

1. Kristensen, V.N. *et al.* (2001) High-throughput methods for detection of genetic variation. *Biotechniques*, 30, 318-322, 324, 326 passim.
2. Sellick, G.S. *et al.* (2004) Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res*, 32, e164.
3. Gibbs, R.A. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428, 493-521.
4. Greenhouse, D.D., Festing, M.F.W., Hasan, S., Cohen, A.L. (1990) Catalogue of Inbred Strains of Rats. In *Genetic Monitoring of Inbred Strains of Rats* (Hedrich, H.J., ed.) Gustav Fisher.

5. Hedrich, H.J. (2005) Taxonomy and Stocks and Strains. In *The Laboratory Rat*, 2nd edition (Krinke G.J., ed.) Academic Press.
6. Zimdahl, H. *et al.* (2004) A SNP map of the rat genome generated from cDNA sequences. *Science*, 303, 807.
7. Smits, B.M. *et al.* (2004) Genetic variation in coding regions between and within commonly used inbred rat strains. *Genome Res*, 14, 1285-1290.
8. Guryev, V. *et al.* (2004) Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res*, 14, 1438-1443.
9. Lindsey, J.R. (1979) Historical Foundations. In *The laboratory rat,* 1st edition. Academic Press.
10. Hedrich, H.J. (2000) History, Strains and Models. In *The Laboratory Rat* (Krinke, G.J., ed.) Academic Press.
11. Koch, L.G. and Britton, S.L. (2005) *Strains*. Oxford University Press, Oxford, UK.
12. Ideraabdullah, F.Y. *et al.* (2004) Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res*, 14, 1880-1887.
13. Yalcin, B. *et al.* (2004) Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. *Proc Natl Acad Sci U S A*, 101, 9734-9739.
14. Nickerson, D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res*, 25, 2745-2751.
15. Kent, W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res*, 12, 656-664.
16. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386.
17. Canzian, F. (1997) Phylogenetics of the laboratory rat Rattus norvegicus. *Genome Res*, 7, 262-267.
18. Thomas, M.A. *et al.* (2003) Phylogenetics of rat inbred strains. *Mamm Genome*, 14, 61-64.
19. Bandelt, H.J. *et al.* (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*, 16, 37-48.
20. Rapp, J.P. (2000) Genetic analysis of inherited hypertension in the rat. *Physiol Rev*, 80, 135-172.
21. Kwitek, A.E. *et al.* (2004) High-density rat radiation hybrid maps containing over 24,000 SSLPs, genes, and ESTs provide a direct link to the rat genome sequence. *Genome Res*, 14, 750-757.
22. Steen, R.G. *et al.* (1999) A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res*, 9, AP1-8, insert.
23. Grupe, A. *et al.* (2001) In silico mapping of complex disease-related traits in mice. *Science*, 292, 1915-1918.
24. Liao, G. *et al.* (2004) In silico genetics: identification of a functional element regulating H2-Ealpha gene expression. *Science*, 306, 690-695.
25. Pletcher, M.T. *et al.* (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol*, 2, e393.
26. Wang, X. *et al.* (2004) Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res*, 14, 1767-1772.
27. Flint, J. *et al.* (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet*, 6, 271-286.
28. Pearson, H. (2004) Mouse sequencing plan aims to boost models. *Nature*, 432, 5.
29. Gunther, E. and Walter, L. (2001) The major histocompatibility complex of the rat (Rattus norvegicus). *Immunogenetics*, 53, 520-542.
30. Cramer, D.V. *et al.* (1988) Genetic diversity within and between natural populations of Rattus norvegicus. *J Hered*, 79, 319-324.
31. Campino, S. *et al.* (2002) Unique genetic variation revealed by a microsatellite polymorphism survey in ten wild-derived inbred strains. *Genomics*, 79, 618-620.

32. Abe, K. *et al.* (2004) Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis. *Genome Res*, 14, 2439-2447.
33. Guenet, J.L. and Bonhomme, F. (2003) Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet*, 19, 24-31.
34. Voigt, B. *et al.* (2000) Genetic comparison between laboratory rats and Japanese and German wild rats. *Mamm Genome*, 11, 789-790.
35. Cholnoky, E. *et al.* (1974) Aspects of genetically defined populations in toxicity testing. II. Genotypic differences in sensitivity to a toxic dextran preparation. *Z Versuchstierkd*, 16, 43-48.
36. Ohno, K. *et al.* (1994) Establishment of new inbred strains derived from Japanese wild rats (Rattus norvegicus). *Jikken Dobutsu*, 43, 251-255.
37. de Boer, S.F. *et al.* (1999) Selective antiaggressive effects of alnespirone in resident-intruder test are mediated via 5-hydroxytryptamine1A receptors: A comparative pharmacological study with 8-hydroxy-2-dipropylaminotetralin, ipsapirone, buspirone, eltoprazine, and WAY-100635. *J Pharmacol Exp Ther*, 288, 1125-1133.

Website References

http://www.ncbi.nlm.nih.gov/SNP/; dbSNP NCBI
http://rgd.mcw.edu/strains; RGD-strains
http://cascad.niob.knaw.nl/snpview; Web interface for exploring verified SNP data between two arbitrary strains
http://primers.niob.knaw.nl; Local Primer3-based design application
http://fluxus-engineering.com; Homepage Fluxus Engineering

CHAPTER 6

Identifying Polymorphisms in the *Rattus norvegicus* D_3 Dopamine Receptor Gene and Regulatory Region

Adapted from Genes, Brain and Behavior (2004) ***3:*** *138–148*

Identifying Polymorphisms in the *Rattus norvegicus* D_3 Dopamine Receptor Gene and Regulatory Region

Bart M. G. Smits[1], Ursula M. D'Souza[2], Eugene Berezikov[1], Edwin Cuppen[1*], and Frans Sluyter[2]

[1]Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[2]MRC SGDP Centre, Institute of Psychiatry, King's College London, UK

[*]Corresponding Author

The D_3 dopamine receptor has been implicated in several neuropsychiatric disorders, including schizophrenia, Parkinson's disease and addiction. Sequence variation in the D_3 gene can lead to subtle alteration in receptor structure or gene expression and thus to a different phenotype. In this study we examined the sequence variation in the D_3 gene in 96 rat strains and substrains. Interestingly, the analyses revealed 10 polymorphisms in the 5' flanking region and four polymorphisms in intronic regions of the gene. Moreover, two single nucleotide polymorphisms (SNPs) that result in amino acid changes were found in the last exon of the D_3 gene in the RNU/Mol strain. Additionally, bioinformatic analysis of the 5' flanking region and first intron of the gene, revealed putative transcription factor binding sites that are conserved between mouse and human and are affected by the SNPs, possibly resulting in altered regulation of the subsequent transcription factor.

Introduction

Dopamine is a major neurotransmitter in the brain, where it plays an important role in controlling motor function and behavioral homeostasis. The receptors through which dopamine exerts its effects, belong to the large family of seven-transmembrane domain G protein-coupled receptors and, based on their structural and functional properties, are divided into 'D_1-like' (D_1 and D_5) and 'D_2-like' (D_2, D_3 and D_4) subgroups [1]. In this paper we focused on the D_3 dopamine receptor, which has a unique expression profile in the brain and is known to play a key role in several neuropsychiatric disorders including schizophrenia and addiction [2].
Of the five major dopamine receptor subtypes the D_3 receptor shows preferential localization in the mesocorticolimbic system [3,4]. Highest distributions are found in limbic parts of the striatum complex including the islands of Calleja, olfactory tubercles, anterior and shell parts of the nucleus accumbens and also in the bed nucleus of stria terminalis [5-8]. This restricted distribution to specific brain areas makes the D_3 receptor a promising target for the development of clinical compounds because they are not as abundantly and widely distributed as dopamine D_1 and D_2 receptors, thereby reducing the risk of undesirable side effects. Moreover, distribution of D_3 mRNA partially overlaps with, but is different from that of D_2. For instance, high levels of D_3

mRNA are detected in the islands of Calleja, where there is only weak expression of the D_2 mRNA. Similar to the distribution pattern of the D_3 mRNA, receptor-binding experiments have revealed high levels of the D_3 dopamine receptor protein in the ventral striatum [9,10].

Abnormal expression of D_3 is associated with typical 'dopamine' disorders. Accordingly, D_3 has been implicated in Parkinson's disease [11,12]. However, most attention has been paid to its role in the development of schizophrenia and substance abuse. For instance, post-mortem brain studies have shown that D_3 mRNA is reportedly down-regulated in the cerebral cortex of schizophrenics [13], suggesting that untreated patients have elevated D_3 receptor binding which is decreased by antipsychotic treatment [14].

Likewise, it has been suggested that D_3 is involved in substance abuse. Individuals who died from cocaine overdose have a two- to threefold increase in D_3 receptor density in the caudate, nucleus accumbens and substantia nigra [15] and an up-regulation of D_3 mRNA in the nucleus accumbens [16]. The importance of D_3 in addiction, in particular cocaine addiction, has been confirmed and expanded in studies using animal models. First indications came from a study of Caine and Koob [17], who showed that selective D_3 agonists such as 7-OH-DPAT and quinpirole potently decreased cocaine self-administration in the rat at doses that were not by themselves reinforcing. The role of D_3 was further accentuated by the finding that BP 897, a D_3 receptor partial agonist (high affinity for the receptor but low intrinsic activity), inhibited cocaine-seeking behavior in rats that depends upon the presentation of drug-associated cues [18], without having any intrinsic, primary rewarding effects [18,19] and without altering incentive motivation for food, morphine and/or their associated cues [19]. Additional evidence for the involvement of D_3 came from a recent study by Vorel *et al.* [20] who tested the effects of a selective D_3 receptor antagonist (SB-277011-A) and found that this antagonist blocked the rewarding effects of cocaine and cocaine-induced drug-seeking behavior, as well as attenuated cocaine-induced conditioned place preference.

Studies in mice seem to confirm the involvement of D_3 in cocaine addiction. Mice lacking functional D_3 receptors display an increased behavioral responsiveness to cocaine [21]. These knockout mice also showed an increase in cocaine cue-conditioned hyperactivity [22]. Taken together, all these rat and mouse studies support a crucial role for D_3 receptors in cocaine abuse, although the underlying mechanism still requires further research.

In the present paper we focus on the gene encoding the D_3 dopamine receptor gene, which was first cloned in rat [8], then in human [23] and last in mouse [24]. In the rat and human the D_3 dopamine receptor gene contains six exons and five introns, although an additional seventh exon was identified in this study by feeding the rat mRNA sequence to GENOTRACE [25], a tool which builds a local genome assembly from raw traces of the rat genome sequencing project. In rat, the gene encodes a 446 amino acid protein [8,23], whereas the protein in humans is only 400 amino acids long. Yet, relatively little is known about the regulatory region of this gene. Although short segments of the mouse 5' untranslated [24] and human 5' flanking regions [26,27] have been published, only the 5' flanking region of the rat D_3 dopamine receptor gene has been analyzed in detail. Using deletion constructs D'Souza et al. [28] discovered the presence of two untranslated exons upstream from the previously published coding exons. Moreover, functional analysis of the D_3 deletion constructs revealed promoter activity within 36 nucleotides upstream from the transcription start site and silencer regions located between −36 and −537 nucleotides and between −783 and −1046 nucleotides [28].

Genetic variation in these key regulatory regions can therefore severely affect the expression of the D_3 dopamine receptor gene and in this way, for instance, affect the process and extent of cocaine addiction. The aim of the present paper is to identify key polymorphisms in the regulatory, as well as in the coding regions, of the rat D_3 dopamine receptor gene. This is achieved using high-throughput SNP detection technology, which identifies variation in the D_3 dopamine receptor gene in a set of 96 commonly used rat strains and substrains. The following regions will be examined: all exons, first intron, exon-flanking intron sequences and 5' flanking

sequence of the first exon (-2752 bp). Effects of polymorphisms in the 5' regulatory sequence on potential transcription factor (TF) binding sites are further analyzed using an *in silico* phylogenetic footprinting approach. This revealed that some of the discovered polymorphisms have a deleterious effect on the binding of a TF to a putative site, which might alter the expression of D_3. This information can be used to study the effect of expression variation of the D_3 receptor on different disease processes.

Methods

Genomic DNA isolation and origin of genomic DNA of the rat strains

Genomic DNA from reference strain BN/Crl was isolated using QIAGEN DNeasy 96® Tissue Kit (4; Hilden, Germany). Genomic DNA from 96 testing strains was kindly provided by L.F.M. van Zutphen, Department of Laboratory Animal Science, Faculty of Veterinary Medicine, Utrecht University, The Netherlands [29].

Five microliters of genomic DNA (concentration of approx. 2 ng/µl per sample) was aliquoted out in 384 wells plates. In total, two sets of 96 testing strains and 2 sets of 96 reference strain DNA (BN) were transferred into a single 384-wells plate using a 96-channel pipette (Hydra-96).

Determination of the genomic structure of the D_3 receptor gene and primer design

First, we determined the genomic organization of *Rattus norvegicus* dopamine receptor D_3 gene using GENOTRACE [25]. This software uses the published cDNA sequence (GenBank accession NM_017140) as a template and builds a local genome around it from overlapping traces of the rat genome sequencing project. Based on this genomic organization (Fig. 1), we designed nested sets of oligonucleotides for amplification of exon sequences using a PRIMER3-based web-application (http://primers.niob.knaw.nl) with optimal melting temperatures of 58°C [30]. This program was also used for designing primers in the 5' flanking region of the rat dopamine receptor D_3 gene. Detailed primer information can be obtained from the authors upon request.

CEL I based polymorphism detection

This strategy is based on specific heteroduplex cleavage by CEL I nuclease, an enzyme isolated from the celery plant [31]. A schematic overview of the methodology is outlined in Figure 2.

All subsequent pipetting steps were done on a Genesis Workstation 200 (Tecan). Target sequences were amplified by nested PCR in 384-wells plates. The first PCR was done with gene specific primers and carried out using a touchdown cycling program (92°C for 60 sec; 30 cycles of 92°C for 20 sec, 65°C for 20 sec with a decrement of 0.5°C per cycle, 72°C for 60 sec; followed by 10 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 60 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems). PCR reaction mixes contained 5 µl genomic DNA, 0.2 µM forward primer and 0.2 µM reverse primer, 200 µM of each dNTP, 25mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate pH 8.7 and 0.2 U Taq Polymerase in a total volume of 10 µl.

After adding water to the first PCR to a total volume of 30 µl, 1 µl of the first PCR reaction was used as template for the second PCR reaction. The second PCR reaction contained gene specific primers, at their 5' end elongated with universal M13 adaptor sequences (forward 0.08 µM and reverse 0.04 µM). Additionally, the reaction mixture contained corresponding universal M13-forward primer (5'-TGTAAAACGACGGCCAGT, 0.12 µM) and universal M13-reverse primer (5'-AGGAAACAGCTATGACCAT, 0.16 µM). The universal M13 primers were labeled with fluorescent dyes, IR Dye 700 and IR Dye 800, respectively. Furthermore, the nested PCR mixture contained 200 µM of each dNTP, 25 mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate pH 8.7 and 0.2U Taq Polymerase in a total volume of 5 µl. Standard cycling conditions were used for the nested PCR reactions (30 cycles of 92°C for 20 sec, 58°C for 40 sec and 72°C for 60 sec).

After the second PCR reaction, 2.5 µl of each sample of the 96 testing strains was mixed in a fresh 384-well plate with 2.5 µl of the reference strain PCR, followed by heteroduplex formation (99°C for 10 min and 70 cycles of 70°C for 20 sec with a decrement of 0.3°C per cycle). Specific heteroduplex cleavage was

performed by adding to each sample 10 µl of CEL I mixture [31], containing 10 mM HEPES pH 7.0, 10 mM $MgSO_4$, 10 mM KCl, 0.002% Triton X-100, 0.2 µg BSA and 0.01 µl CEL I enzyme solution (isolated from celery according to Oleykowski *et al.* [31] with minor modifications; protocol available at http://www.niob.knaw.nl/researchpages/cuppen/protocols.html) and incubation at 45°C for 15 min. The CEL I reactions were stopped by addition of 5 µl 75 mM EDTA pH 8.0.
Fragments were purified using Sephadex G50 (medium coarse, Sigma, St. Louis, MO) minicolumns in 96-wells filter plates (Multiscreen HV, Millipore, Billerica, MA) and eluted into plates prefilled with 5 µl formamide loading buffer (37% (v/v) de-ionized formamide, 4 mM EDTA pH8.0, 90 µg/ml bromophenol blue) per well. Samples were concentrated to about 1.5 µl by heating at 85°C for 45 to 60 min without cover. 0.3 µl was applied to a 96-lane membrane comb (The Gel-Company, San Francisco, CA) and loaded on 25 cm denaturing 6% polyacrylamide gels on LI-COR 4200 DNA analyzers (LI-COR, Lincoln, NE). Raw TIFF-images produced by the analyzers were modified and visualized using Adobe Photoshop and potential polymorphisms were detected and scored manually.
All polymorphisms were verified using automated sequencing.

Bioinformatic analysis for putative transcription factor binding sites in the 5' flanking region and first intron of the D_3 gene

The last step of the project was to investigate whether the identified SNPs affected putative transcription factor binding. This was achieved using a (bioinformatic) phylogenetic footprinting method, which utilized human, rat and mouse sequences.
Since D_3 5' UTR and exons 1 and 2 (GenBank acc. nr. AF326967) had not been annotated in Ensembl database, AF326967 sequence was mapped to Ensembl contig RNOR01023028 by BLAST searches, and region of this contig from 100861 to 106740 bp was used in the analysis. In addition, the mouse (ENSMUSG00000022705) and human (ENSG00000151577) Ensembl genes were identified as orthologs to the rat D_3 gene (AF326967) by BLAST searches. Upstream sequences of these genes were extracted from Ensembl database and used in phylogenetic footprinting analysis performed by CONREAL software [32].
CONREAL utilizes all 398 mammalian position weight matrices (PMW) from TransFac database (v.7.1) [33] to find and align potential TF binding sites in a pair of orthologous promoter sequences. The analysis was performed under two conditions: (i) a stringent search, characterized by the following parameters: 80% for PWM threshold, 50% for homology threshold and 15 bp flank length for stringent search conditions (80-50-15) and (ii) a less stringent, more flexible search, characterized by 70% PWM threshold, 50% homology threshold and 10 bp flank length (70-10-10). The output of CONREAL (pairwise promoter alignment and list of putative TF binding sites) was used to associate identified SNPs with TF binding sites.
The 70-10-10 searches, the full list of putative TF binding sites and the outputs of all CONREAL alignments are not shown in the results section, but are available at http://www.niob.knaw.nl/researchpages/cuppen/publications/suppl_data_03/.

Results

Determination of the genomic structure of the D_3 receptor gene

In contrast to the previously reported genomic structure of the rat D_3 receptor gene, the GENOTRACE output revealed 7 coding exons (Fig. 1). The existence of an additional intron, splitting exon 6 in exons 6 and 7, was verified by searching the recently released draft version of the rat genome with the D_3 receptor cDNA sequence using BLAST (available at http://www.ncbi.nlm.nih.gov/genome/seq/RnBlast.html).

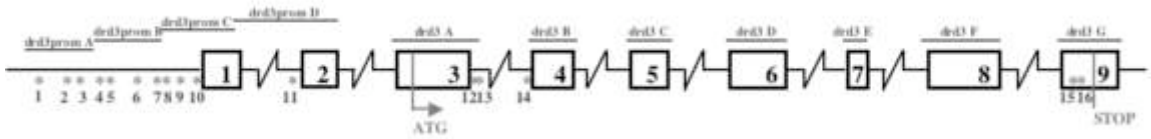


Figure 1: Schematic representation of the genomic structure of the rat D_3 dopamine receptor.
The structure is depicted as determined by GENOTRACE. Each numbered box represents an exon. The start ATG and the STOP codon are located in the third exon and ninth exon, respectively. The lines above the genomic structure indicate the relative positions of the amplicons designed for SNP discovery. Asterisks, numbered from 1 to 16, indicate the position of the polymorphisms detected in the amplicons.

Polymorphisms in the D_3 receptor gene

In total, we screened 6706 bp per strain, divided over 11 amplicons of the D_3 receptor gene, resulting in 16 polymorphisms (Table 1; Fig. 1). In 2752 bp of the 5' flanking sequence of the first exon including the promoter, 10 polymorphisms were found: one five-bp deletion, two two-bp substitutions and seven single nucleotide polymorphisms.
Screening of the first exon and the exon-flanking intron sequences resulted in the identification of 4 polymorphisms: one SNP is located in the first intron, three SNPs are located in the third intron.
Interestingly, two SNPs were found in the last exon of the D_3 receptor gene. Both SNPs occur in one strain only: RNU/Mol. The first SNP results in a threonine to asparagine change and the second SNP gives rise to a valine to alanine change, at positions 399 and 434 in the protein respectively. A snake-like representation of the dopamine receptor D_3 protein is shown in Figure 3, with the altered amino acid residues indicated by arrows (http://www.gpcr.org/seq/vis/D3DR_RAT/D3DR_RAT.html). The different colors indicate the major chemical characteristics of the amino acid residues (according to RasMol CPK). This representation reveals that Thr399 is located in the beginning of the last extracellular part of the D_3 protein and that the Thr399Asp change does modify the chemical properties of the protein at this position, implying functional consequences. Although the Val434Ala change does not seem to change the chemical characteristics of the protein at this position, this is the last residue of the seventh transmembrane domain. Multiple sequence alignment (data not shown) shows that this position in the protein is conserved between human, mouse and rat, suggesting that this residue is important for proper function of the receptor and that a change at this position might influence the structure of the transmembrane domain.

Polymorphism distribution in 96 rat strains

In many cases, SNPs are coupled, resulting in haplotype blocks [34,35]. Table 2 gives an overview of the distribution of the polymorphisms in the dopamine receptor D_3 gene over the 96 testing strains. Most of the discovered polymorphisms show physical linkage. Polymorphisms 2 and 3 are linked to each other in all cases, as well as polymorphisms 4 and 5, 12 and 13, 8 and 9 and 15 and 16. In 5 of the 8 strains two sets of polymorphisms are linked (2 & 3 with 4 & 5). In two strains polymorphism 14, located near the fourth exon, is also linked to polymorphisms 1 to 5.
Although all 96 strains used in this study are considered to be inbred, polymorphisms were found among strain samples obtained from different locations, e.g. SNP 14 in strain R/A. In another sample of this strain, there is even heterozygosity in the D_3 gene. Although numbers are low, these data suggest that different isolates of inbred strains can vary genetically.

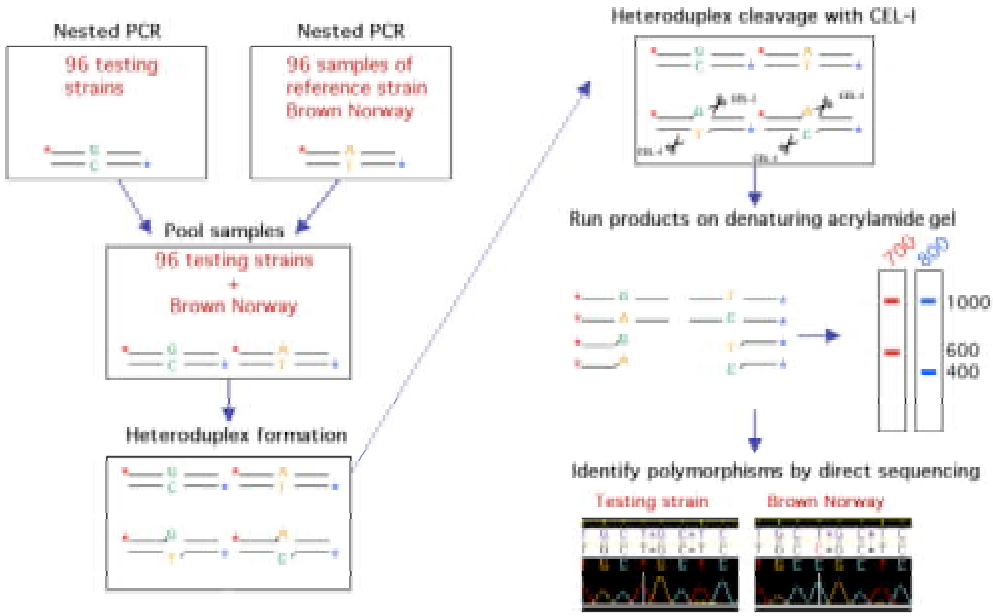


Figure 2: Schematic overview of the polymorphism discovery strategy.
Independent nested PCRs on genomic DNA of 96 testing strains and on a set of 96 samples of the reference strain Brown Norway results in two sets of 96 PCR products. The example product, which is differently labeled at both sides indicated by the red and blue stars, contains a SNP between one of the testing strains and the reference strain Brown Norway. By pooling this sample with the reference strain sample, followed by denaturing and reannealing the DNA, heteroduplexes are formed that are specifically cleaved by the CEL I nuclease. After purification on sephadex columns, full-length products and cleaved products are separated on a polyacrylamide gel and fluorescence is detected at two different wavelengths on a LI-COR machine. Subsequently, positive samples are resequenced reveal the molecular nature of the polymorphism.

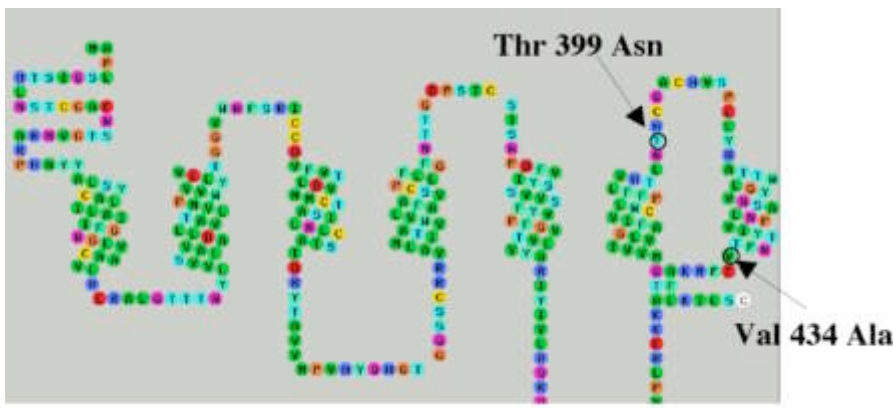


Figure 3: Viseur snake-like plot of the rat D3 dopamine receptor.
The amino acid changes are indicated by arrows (adapted from http://www.gpcr.org/7tm/seq/vis/D3DR_RAT/D3DR_RAT.html). Different colors represent different amino acid residues (corresponding to their chemical properties, RasMol CPK colors). The large intracellular loop is cut out and indicated by two vertical arrows and the word 'loop'.

Polymorphisms in putative transcription factor binding sites in the regulatory region of receptor D_3

Applying CONREAL under stringent conditions (http://conreal.niob.knaw.nl, conditions 80-50-15) to the different species' D_3 5' flanking regions and first intron, we identified a large set of putative TF binding sites harboring one of the identified polymorphic sites. A subset of the TF binding sites might be affected by the introduction of the various polymorphisms in the promotor and intronic sequences (Table 3). However, it should be noted that the majority of TF binding sites is probably not affected by genetic variation.

Table 1: Identified Polymorphisms in the Rat D_3 Dopamine Receptor Gene

Position in gene	SNP ID	Nucleotide Change	Amino Acid Change	Position in AF236967	Position in NM_017140
Promoter	1	TG → CA		946	
	2	A → G		1103	
	3	C → T		1224	
	4	ACAAAA del		1845	
	5	C → T		1876	
	6	G → A		2288	
	7	G → C		2474	
	8	T → A		2655	
	9	A → T		2676	
	10	CC → TT		3019	
Exon-flanking intron sequences	11	G → T		4207	
	12	T → A			
	13	A → G			
	14	T → C			
Coding region	15	C → A	Thr399Asn		1277
	16	T → C	Val434Ala		1382

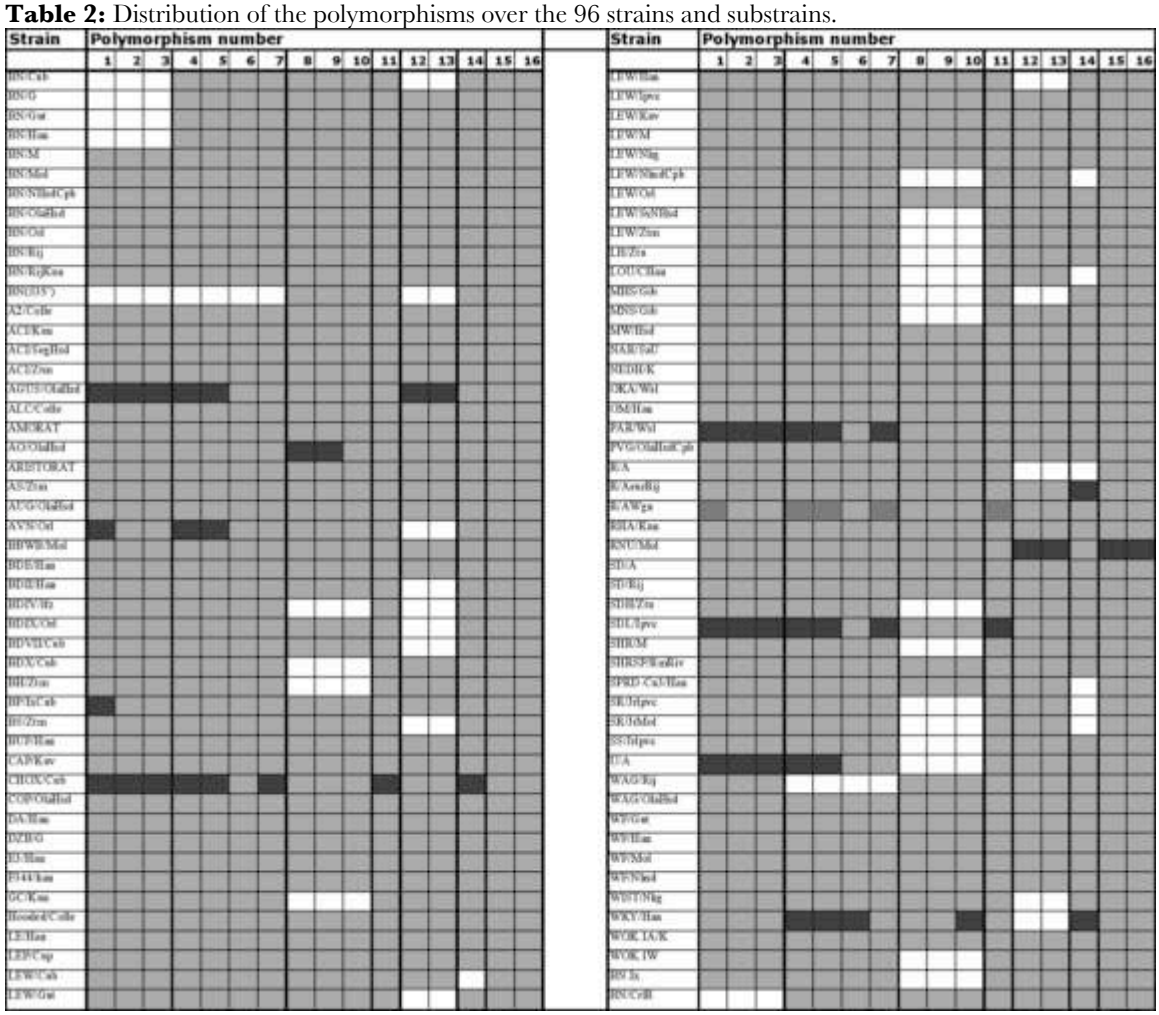
The SNP ID's correspond to the numbers schematically outlined in Figure 1. The identified polymorphisms occur at least once in the 96 strains or substrains when compared to the reference strain Brown Norway. Numbers AF326967 and NM_017140 correspond to Locus ID in the NCBI Entrez nucleotide database (available at http://www.ncbi.nlm.nih.gov).

Discussion

The purpose of the present paper is to identify polymorphisms in the regulatory, as well as in the coding regions, of the rat D_3 dopamine receptor gene. To achieve this goal, we screened a set of 96 inbred rat strains and substrains [29] for polymorphisms using a high-throughput SNP typing methodology. This resulted in an inventory of 16 polymorphisms in the coding region, exon-flanking intron sequences and 5' flanking region (Fig. 1). Of these 16 polymorphisms, only two were found in coding exons, i.e. in the last exon of the D_3 receptor gene. Both SNPs occur in only one strain (RNU/Mol) and lead to amino acid changes. Whether these amino acid changes lead to alteration of protein function, remains to be investigated. The valine to alanine change seems not dramatic as both amino acids have non-polar side chains. The threonine to asparagine change looks more promising because it alters the chemical properties of the D_3 protein (Fig. 3). This change is located in the beginning of the last extracellular part of the protein and, although the actual function of this part is not known, it is possible that the threonine to asparagine change could have an effect on the binding properties of the receptor to various drugs. These mutations

can be mimicked in an *in vitro* system to study the functional effect and ultimately animals of this particular strain can be studied in various drug response assays. In the human and mouse orthologs of D_3, seven and three mutations are found, respectively (http://www.gpcr.org/mutation/receptors_mut.html). These are coding mutations and are different from the mutations we found.

Table 2: Distribution of the polymorphisms over the 96 strains and substrains.



The light grey squares represent a genotype identical to the reference strain. The dark grey squares represent polymorphic sites compared to reference strain BN/Crl. Intermediate grey squares are heterozygous polymorphisms. White squares are not analyzed. The polymorphisms are organized by amplicon number, separated by thick black lines (see also Fig. 1).

Screening of the first (noncoding) exons and the exon-flanking intron sequences resulted in the identification of four SNPs: one in the first and three in the third intron (polymorphisms 11 to 14 in Table 1 and Fig. 1). The majority of polymorphisms (1 to 10) was, however, found in the 5' flanking sequence of the first exon (-2752 bp), including the promotor, and can be further subdivided into seven SNPs, two two-bp substitutions and one five-bp deletion.

Polymorphisms scarcely arise on their own and generally show physical linkage, i.e. on genome-wide scale they are organized in haplotype blocks [34]. Linkage is also observed for the SNPs and other polymorphisms in our set of strains (Table 2). High rates of linkage are not surprising in rodent strains as most strains originate from a mixed but limited founder population [34,36]. Another interesting, but slightly disturbing finding is that an inbred strain from one location is not

necessarily genetically identical to the same inbred strain from another location, e.g. SNP 14 in strain R/A. In another sample of this strain, there is even heterozygosity in the D_3 gene. Inbred strains, which are the product of at least 20 generations of brother-sister matings, are supposed to be genetically identical across different labs over the world [37] and homozygous at 98.6% of all loci [38,39].

We also investigated *in silico* whether the polymorphisms in the promotor region and first intron possibly affected putative transcription factor binding sites (and, therefore, possibly gene expression). This was achieved using phylogenetic footprinting analysis that was performed by a program entitled CONREAL. This method utilizes a matrix-based search to predict conserved transcription factor (TF) binding sites among different species, such as rat, mouse and human [32]. The results show that, although the majority of the putative TF binding sites are not changed dramatically by the occurring genetic variation, some polymorphisms seem to influence binding of TFs at D_3 sites. For instance, the putative binding sites of STAT6 (Signal Transducer and Activator of Transcription 6) and STAT5A (Signal Transducer and Activator of Transcription 5A) at polymorphic site 11 (Fig. 1; Table 3) are conserved in rat, mouse and human, implying a selective pressure on this site. This suggests that STAT6 and STAT5A normally influence gene expression by binding to this intronic DNA sequence. When the SNP is introduced, these binding sites are not recovered in the CONREAL analysis, neither under standard nor under less stringent search conditions. In other words, in strains with a SNP at position 11 STAT6 and STAT5A may not bind to this site. STAT transcription factors are, however, far from specific and bind to many sites, generally indicated by $TT(N)_5AA$ [40]. Moreover, since STAT binding sites arise abundantly in the D_3 promoter sequence (data not shown), it is difficult to predict the functional effect of this particular SNP. Furthermore, STATs 5A and 6 still have unknown functions in the brain (review of STATs: [41,42]).

In contrast, transcription factor LEF-1 (lymphoid enhancer factor-1) is more specific and binds to specific DNA sequences [43-45]. This transcription factor is mainly known to control embryonic patterning and cell-fate decisions in development [46]. Our analysis shows that the LEF-1 site at polymorphic position 7 is present in the 'normal' rat-mouse comparison (i.e. using the Brown Norway strain as a reference), but is lost if the SNP is introduced, even when the search is performed under less stringent conditions. This result implies that LEF-1 binds to the regulatory sequence of D_3 at polymorphic position 7 in the wild type strain, but does not bind to the regulatory region of D_3 in strains containing the SNP at that position. This could lead to slight alterations in the expression of the D_3 receptor gene in specific cells in which LEF-1 regulates the expression of D_3. It is important to realize that slight alterations might be very effective. For instance, a homozygous insertion/deletion polymorphism in the promoter region of the *presenilin 2* gene has been suggested to increase the risk for Alzheimer´s Disease. DNA-protein binding analysis experiments proposed that this site interacts with a TF [47].

There are also TF binding sites that are created by the introduction of a SNP. For instance, the binding sites for GATA-binding factors 1 and x are not found in the 'normal' rat-mouse comparison (nor under less stringent conditions (data not shown)), but are added to the set when the rat sequence containing SNP 5 is compared to the mouse sequence. The GATA-binding factors 1 and x may therefore only bind to this position in the regulatory region of the D_3 gene of the strains with SNP 5.

SNPs 7 to 10 inclusive are located in regions already characterized with regard to their effects on transcriptional activity [28]. Thus, SNP 7, which, according to the *in silico* analysis, leads to a loss of the LEF-1 binding site, is located in the silencer region between nucleotides −782 and −1046 from the start of transcription. SNPs 8 and 9 are situated in the potential activator region present between nucleotides −537 and −782. SNP 8 creates a binding site for TATA Binding Protein, while the effects of SNP 9 are multiple (Table 3). SNP 10, which does not seem to affect binding of a TF, is present in the silencer region between −288 and −537 nucleotides from the start of transcription.

Table 3: Putative Transcription Factor Binding Sites in the 5' flanking region and the first intron of the D_3 gene affected by the polymorphisms as predicted by CONREAL.

SNP	TFBS	TF description	R – M	R – H	Rmut-M
1	C/EBP	CCAAT/enhancer binding protein	X		X
	C/EBP beta	CCAAT/enhancer binding protein beta	X		
	C/EBP alpha	CCAAT/enhancer binding protein alpha	X		
	C/EBP beta	CCAAT/enhancer binding protein beta			X
	C/EBP	CCAAT/enhancer binding protein	X		X
	USF	Upstream Stimulating Factor	X		
	AP-2 gamma	Activator Protein 2 gamma	X		
	AP-2 alpha	Activator Protein 2 alpha	X		X
2	ELF1	ELF1	X		X
	C/EBP	CCAAT/enhancer binding protein	X		X
	Lyf-1	Lyf-1	X		X
	C/EBP	CCAAT/enhancer binding protein	X		X
	STATx	Signal Transducer and Activator of Transcription x	X		X
	E2F	E2F	X		
	E2F-1	E2F-1	X		
3	Freac-7	Fork head RElated ACtivator-7			X
	FOXO4	Fork head bOX O 4			X
	SOX5	SRY-related HMG bOX 5			X
5	GATA-1	GATA-binding factor 1			X
	GATA-x	GATA-binding factor x			X
	GATA	GATA-binding factor			X
	GATA-1	GATA-binding factor 1			X
	IRF1	IRF1	X		X
7	LEF-1	LEF-1	X		
	HOXA4	HomeobOX cluster protein A4	X		X
	AP-2 alpha	Activator Protein 2 alpha	X		X
	AP-2 gamma	Activator Protein 2 gamma	X		X
8	HOXA3	HomeobOX A3	X		X
	TBP	TATA Binding Protein			X
9	C/EBP	CCAAT/enhancer binding protein	X		
	NF-AT	Nuclear Factor of Activated T-cells	X		
	C/EBP beta	CCAAT/enhancer binding protein beta	X		
	C/EBP delta	CCAAT/enhancer binding protein delta			X
	C/EBP	CCAAT/enhancer binding protein	X		X
	S8	S8	X		
	Nkx6-2	NK related homeoboX factor 6-2	X		
	Oct-1	Octamer 1	X		X
	Oct-1	Octamer 1	X		
	Nkx6-2	NK related homeoboX factor 6-2	X		
	Nkx2-5	NK related homeoboX factor 2-5	X		
	Lhx-3	LIM homeobox transcription factor 3	X		
11	AP-2 alpha	Activator Protein 2 alpha			X
	E2F-1	E2F-1	X		
	STAT1	Signal Transducer and Activator of Transcription 1		X	X
	STAT6	Signal Transducer and Activator of Transcription 6	X	X	
	PAX-2	PAired boX 2	X		X
	STAT6	Signal Transducer and Activator of Transcription 6	X	X	
	STAT3	Signal Transducer and Activator of Transcription 3	X		
	STAT5A	Signal Transducer and Activator of Transcription 5A	X		
	STAT4	Signal Transducer and Activator of Transcription 4	X	X	X
	STAT1	Signal Transducer and Activator of Transcription 1		X	
	STAT5A	Signal Transducer and Activator of Transcription 5A	X	X	
	HMG IY	HMG IY		X	

TFBS are predicted based on homology with the mouse (M) and human (H) 5'flanking regions and first exons of the D_3 gene. Comparisons are made using CONREAL at stringency 80-50-15, where 80 stands for 80% PWM (Position Weight Matrix from TransFac database) threshold, 50 stands for 50% homology threshold and 15 for 15bp flanks to calculate homology.

Future experiments are to reveal whether these SNPs truly affect transcriptional activity. For instance, reporter gene assays, in combination with electrophoresis mobility shift assays, are outstanding techniques to investigate whether these SNPS affect the expression of the D3 gene *in vitro*. Such an approach has proven to be successful for the human serotonin receptor 1B gene.

Using a secreted alkaline phosphatase (SEAP) reporter gene system [48] found that specific polymorphisms in the promoter of this gene affected transcriptional activity. Electrophoretic mobility shift experiments further showed that certain SNPs either generated the binding of new TFs or modified the binding of existing TFs. If such *in vitro* experiments demonstrate that, similar to the human serotonin receptor 1B gene, transcriptional activity of the D3 gene is also affected by certain SNPs in the 5´ flanking region, the next logical step is to determine whether this change in gene expression is functional *in vivo* as well. Possible experiments include the creation of strains in which SNPs that have been demonstrated to be significant *in vitro*, are introduced on a standard genetic background by targeted mutagenesis. These strains are then to be neurobehaviorally compared with wild type strains, i.e. strains that are genetically identical to the created mutant except for the introduced SNPs. Differences between the mutant and wild type strains would indicate an *in vivo* effect of the SNP.
Finally, it is important to note that the technique described in this study is suited for high throughput polymorphism analysis in any gene of interest. In this study we have not only identified frequently occurring polymorphisms in the rat D_3 receptor gene, but we have also traced rare spontaneous mutations, such as two amino acid changing SNPs in strain RNU/Mol. There is no reason to assume why this should not be possible for other genes, as well as for other species, such as the mouse. Accordingly, the SNP discovery and genotyping approach as described here, is an efficient methodology to obtain an inventory of natural occurring polymorphisms in genes of interest in laboratory animal populations, potentially allowing a motivated choice for substrains to use in comparative studies.

Acknowledgements

We thank L.F.M. van Zutphen of the Department of Laboratory Animal Science, Veterinary Department, University Utrecht for providing us with the rat genomic DNA.

References

1. Strange, P.G. (1993) Dopamine receptors: structure and function. *Prog Brain Res*, 99, 167-179.
2. Shafer, R.A. and Levant, B. (1998) The D3 dopamine receptor in cellular and organismal function. *Psychopharmacology (Berl)*, 135, 1-16.
3. Suzuki, M. *et al.* (1998) D3 dopamine receptor mRNA is widely expressed in the human brain. *Brain Res*, 779, 58-74.
4. Levant, B. (1998) Differential distribution of D3 dopamine receptors in the brains of several mammalian species. *Brain Res*, 800, 269-274.
5. Bouthenet, M.L. *et al.* (1991) Localization of dopamine D3 receptor mRNA in the rat brain using in situ hybridization histochemistry: comparison with dopamine D2 receptor mRNA. *Brain Res*, 564, 203-219.
6. Diaz, J. *et al.* (1995) Phenotypical characterization of neurons expressing the dopamine D3 receptor in the rat brain. *Neuroscience*, 65, 731-745.
7. Landwehrmeyer, B. *et al.* (1993) Dopamine D3 receptor mRNA and binding sites in human brain. *Brain Res Mol Brain Res*, 18, 187-192.
8. Sokoloff, P. *et al.* (1990) Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. *Nature*, 347, 146-151.
9. Levesque, D. *et al.* (1992) Identification, characterization, and localization of the dopamine D3 receptor in rat brain using 7-[3H]hydroxy-N,N-di-n-propyl-2-aminotetralin. *Proc Natl Acad Sci U S A*, 89, 8155-8159.

10. Levant, B. *et al.* (1992) Characterization of [3H]quinpirole binding to D2-like dopamine receptors in rat brain. *J Pharmacol Exp Ther*, 262, 929-935.
11. Nagai, Y. *et al.* (1993) Expression of the D3 dopamine receptor gene and a novel variant transcript generated by alternative splicing in human peripheral blood lymphocytes. *Biochem Biophys Res Commun*, 194, 368-374.
12. Ryoo, H.L. *et al.* (1998) Dopamine D3 receptor is decreased and D2 receptor is elevated in the striatum of Parkinson's disease. *Mov Disord*, 13, 788-797.
13. Schmauss, C. *et al.* (1993) Selective loss of dopamine D3-type receptor mRNA expression in parietal and motor cortices of patients with chronic schizophrenia. *Proc Natl Acad Sci U S A*, 90, 8942-8946.
14. Gurevich, E.V. *et al.* (1997) Mesolimbic dopamine D3 receptors and use of antipsychotics in patients with schizophrenia. A postmortem study. *Arch Gen Psychiatry*, 54, 225-232.
15. Staley, J.K. and Mash, D.C. (1996) Adaptive increase in D3 dopamine receptors in the brain reward circuits of human cocaine fatalities. *J Neurosci*, 16, 6100-6106.
16. Segal, D.M. *et al.* (1997) Up-regulation of D3 dopamine receptor mRNA in the nucleus accumbens of human cocaine fatalities. *Brain Res Mol Brain Res*, 45, 335-339.
17. Caine, S.B. and Koob, G.F. (1993) Modulation of cocaine self-administration in the rat through D-3 dopamine receptors. *Science*, 260, 1814-1816.
18. Pilla, M. *et al.* (1999) Selective inhibition of cocaine-seeking behaviour by a partial dopamine D3 receptor agonist. *Nature*, 400, 371-375.
19. Duarte, C. *et al.* (2003) Effects of a dopamine D(3) receptor ligand, BP 897, on acquisition and expression of food-, morphine-, and cocaine-induced conditioned place preference, and food-seeking behavior in rats. *Neuropsychopharmacology*, 28, 1903-1915.
20. Vorel, S.R. *et al.* (2002) Dopamine D3 receptor antagonism inhibits cocaine-seeking and cocaine-enhanced brain reward in rats. *J Neurosci*, 22, 9595-9603.
21. Carta, A.R. *et al.* (2000) Cocaine effects on gene regulation in the striatum and behavior: increased sensitivity in D3 dopamine receptor-deficient mice. *Neuroreport*, 11, 2395-2399.
22. Le Foll, B. *et al.* (2002) Role of the dopamine D3 receptor in reactivity to cocaine-associated cues in mice. *Eur J Neurosci*, 15, 2016-2026.
23. Giros, B. *et al.* (1991) Shorter variants of the D3 dopamine receptor produced through various patterns of alternative splicing. *Biochem Biophys Res Commun*, 176, 1584-1592.
24. Park, B.H. *et al.* (1995) Structural organization of the murine D3 dopamine receptor gene. *J Neurochem*, 64, 482-486.
25. Berezikov, E. *et al.* (2002) GENOTRACE: cDNA-based local GENOme assembly from TRACE archives. *Bioinformatics*, 18, 1396-1397.
26. Sivagnanasundaram, S. *et al.* (2000) A cluster of single nucleotide polymorphisms in the 5'-leader of the human dopamine D3 receptor gene (DRD3) and its relationship to schizophrenia. *Neurosci Lett*, 279, 13-16.
27. Anney, R.J. *et al.* (2002) Characterisation, mutation detection, and association analysis of alternative promoters and 5' UTRs of the human dopamine D3 receptor gene in schizophrenia. *Mol Psychiatry*, 7, 493-502.
28. D'Souza, U.M. *et al.* (2001) Characterization of the 5' flanking region of the rat D(3) dopamine receptor gene. *J Neurochem*, 76, 1736-1744.
29. Bender, K. *et al.* (1984) Biochemical markers in inbred strains of the rat (Rattus norvegicus). *Immunogenetics*, 19, 257-266.
30. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386.
31. Oleykowski, C.A. *et al.* (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, 26, 4597-4602.

32. Berezikov, E. *et al.* (2004) CONREAL: conserved regulatory elements anchored alignment algorithm for identification of transcription factor binding sites by phylogenetic footprinting. *Genome Res*, 14, 170-178.
33. Matys, V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, 31, 374-378.
34. Wiltshire, T. *et al.* (2003) Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A*, 100, 3380-3385.
35. Gabriel, S.B. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, 296, 2225-2229.
36. Wade, C.M. *et al.* (2002) The mosaic structure of variation in the laboratory mouse genome. *Nature*, 420, 574-578.
37. Kacew, S. and Festing, M.F. (1996) Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances. *J Toxicol Environ Health*, 47, 1-30.
38. Beck, J.A. *et al.* (2000) Genealogies of mouse inbred strains. *Nat Genet*, 24, 23-25.
39. Davisson, M.T. (1997) Rules and guidelines for genetic nomenclature in mice: excerpted version. Committee on Standardized Genetic Nomenclature for Mice. *Transgenic Res*, 6, 309-319.
40. Seidel, H.M. *et al.* (1995) Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA binding and transcriptional activity. *Proc Natl Acad Sci U S A*, 92, 3041-3045.
41. Aaronson, D.S. and Horvath, C.M. (2002) A road map for those who know JAK-STAT. *Science*, 296, 1653-1655.
42. Horvath, C.M. (2000) STAT proteins and transcriptional responses to extracellular signals. *Trends Biochem Sci*, 25, 496-502.
43. Oosterwegel, M. *et al.* (1991) Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J Exp Med*, 173, 1133-1142.
44. Travis, A. *et al.* (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev*, 5, 880-894.
45. Waterman, M.L. and Jones, K.A. (1990) Purification of TCF-1 alpha, a T-cell-specific transcription factor that activates the T-cell receptor C alpha gene enhancer in a context-dependent manner. *New Biol*, 2, 621-636.
46. Eastman, Q. and Grosschedl, R. (1999) Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr Opin Cell Biol*, 11, 233-240.
47. Riazanskaia, N. *et al.* (2002) Regulatory region variability in the human presenilin-2 (PSEN2) gene: potential contribution to the gene activity and risk for AD. *Mol Psychiatry*, 7, 891-898.
48. Duan, J. *et al.* (2003) Polymorphisms in the 5'-untranslated region of the human serotonin receptor 1B (HTR1B) gene affect gene expression. *Mol Psychiatry*, 8, 901-910.

Website References

http://conreal.niob.knaw.nl; CONREAL-website
http://primers.niob.knaw.nl; Primer3-based primer design application
http://www.niob.knaw.nl/researchpages/cuppen/protocols.html; Protocols mut. detection set-up
http://www.niob.knaw.nl/researchpages/cuppen/publications/suppl_data_03/; Supporting material of this paper (data not shown)
http://www.ncbi.nlm.nih.gov/genome/seq/RnBlast.html; BLAST rat genome page NCBI
http://www.gpcr.org/seq/vis/D3DR_RAT/D3DR_RAT.html; overview of mutations found in dopamine receptors of different mammals

CHAPTER 7

Identification of a Rat Model for Usher Syndrome type 1B by ENU Mutagenesis-driven Forward Genetics

Adapted from Genetics (2005) ***170(4):*** *1887-96*

Identification of a Rat Model for Usher Syndrome type 1B by ENU Mutagenesis-driven Forward Genetics

Bart M. G. Smits[1], Theo A. Peters[2], Joram D. Mul[1], Huib J. Croes[3], Jack A. M. Fransen[3], Andy J. Beynon[2], Victor Guryev[1], Ronald H. A. Plasterk[1], and Edwin Cuppen[1]*

[1]Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
[2]Department of Otorhinolaryngology, [3]Department of Cell Biology, Radboud University, 6500 HB Nijmegen, The Netherlands.

*Corresponding Author

The rat is the most extensively studied model organism and broadly used in biomedical research. Current rat disease models are selected from existing strains and their number is thereby limited by the degree of naturally occurring variation or spontaneous mutations. We have used ENU mutagenesis to increase genetic variation in laboratory rats and identified a recessive mutant, named tornado, showing aberrant circling behavior, hyperactivity, and stereotypic head shaking. More detailed analysis revealed profound deafness due to disorganization and degeneration of the organ of Corti that manifests already at the onset of hearing. We set up a Single Nucleotide Polymorphism (SNP)-based mapping strategy to identify the affected gene, revealing strong linkage to the central region of chromosome 1. Candidate gene resequencing identified a point mutation that introduces a premature stopcodon in *Myo7a*. Mutations in human *MYO7A* result in Usher syndrome type 1B, a severe autosomal inherited recessive disease that involves deafness and vestibular dysfunction. Here, we present the first characterized rat model for this disease. In addition, we demonstrate proof of principle for the generation and cloning of human disease models in rat using ENU mutagenesis, providing good perspectives for systematic phenotypic screens in the rat.

Introduction

The rat is the most frequently used organism in a wide spectrum of biomedical studies, e.g. physiological and nutritional studies, and drug development. Recently, the rat became the third mammal for which the complete genome was sequenced [1]. Nevertheless, mammalian geneticists favor the mouse above the rat as a model organism [1], which is mainly due to the differences in the availability of tools for the manipulation of their genomes. Reverse genetic or knockout technology to disrupt the expression of a gene is well established for the mouse [2,3], but has become available only very recently for the rat by means of *N*-ethyl-*N*-nitrosourea (ENU)

mutagenesis [4,5]. In addition, forward genetic approaches have been successfully employed in mouse for several decades, primarily by employing ENU mutagenesis for the efficient generation of novel models [6].

Several large-scale mouse forward genetic screens for ENU-induced mutations have been initiated [7]. The screens are phenotype-driven, employing appropriate assays on mutant animals to identify phenotypes that resemble aspects of human disease. The resulting models are expected to contribute substantially to the understanding of genetic diseases [8].

For the rat, however, the more than 200 existing models [9,10] are primarily selected from existing laboratory strains and their number and nature is therefore limited by the degree of naturally occurring variation or the occurrence of spontaneous novel mutations. Nevertheless, important rat models exist for common human diseases [11], including hypertension, diabetes, cancer, and many others [12-14]. ENU mutagenesis would have the potential to induce a major increase in the number of rat phenotypes and models, as it has done for the mouse [15,16].

While ENU mutagenesis conditions have recently been established for the rat [4,5], additional bottlenecks for efficient forward genetics are still present. First, establishment of comprehensive and efficient high-throughput screening setups will be challenging, although the rat is being used already for decades as the primary model organism in for example behavioral neurobiology, hypertension and diabetes research. As a result, extensively validated assays are available and provide a solid basis for the development of a rat screening protocol. Second, cloning of the causal mutations underlying aberrant phenotypes depends on the availability of efficient genetic mapping and cloning tools. Currently, the availability of genetic markers to scan the rat genome for linkage to traits of interest is limited. Although a large collection of microsatellite markers has been genotyped in variety of strains and successfully used for mapping and cloning purposes [17,18], the application of the more versatile Single Nucleotide Polymorphism (SNP) markers in the rat is still in its infancy. Large repositories of rat SNPs and candidate SNPs became available recently [5,19], but information on SNP distribution in different strains is still mostly lacking.

In this study, we describe the identification and characterization of the first autosomal recessive mutant rat strain from an ENU mutagenesis-driven study. Animals from this strain display locomotory hyperactivity, circling behavior, and stereotypic head shaking. More detailed characterization reveals that these animals are deaf due to progressive degeneration of the organ of Corti. For mapping and cloning purposes, we designed and tested a genome-wide SNP-based mapping panel for Wistar versus Brown Norway (BN)-based crosses, which allowed us to pinpoint the mutation to a chromosomal subregion containing a limited number of strong candidate genes. With the subsequent identification of a premature stopcodon in one of these genes, *Myo7a*, we established a rat model for the human Usher syndrome type 1B (USH1B) and provide proof of concept for ENU-driven forward genetics in the rat.

Methods

Animals, ENU mutagenesis and crosses

Male Wistar founder rats (Crl:Wistar) were mutagenized using ENU (SIGMA, St. Louis). Mutagenesis was carried out as described [4]. Briefly, intraperitonial (i.p.) injections of ENU were given in three weekly doses. One male received 3 x 40 mg/kg body weight, three males received 3 x 20 mg/kg body weight. The injected animals were mated with untreated females (Crl:Wistar) to establish a F1 population, which was primarily used for a large scale reverse genetics screen described elsewhere [4]. Eight F1 males and eight F1 females were selected from this population and used to set up crosses, preventing brother-sister matings.

Next, F2 progeny from these crosses was used in 18 different brother-sister matings to breed induced mutations to homozygosity, resulting in a single nest with the aberrant tornado (*tnd*) phenotype.
For the mapping cross, two male tornado animals ware mated with four Brown Norway females (BN/Crl). From the F1 generation, 10 males and 10 females were intercrossed to restore the tornado phenotype.

DNA isolation, PCR, and sequencing
For genotyping, animals were tail clipped under isoflurane anesthetics. Lysis on tail clips was done overnight at 55°C in 400 µl lysis buffer, containing 100 mM Tris (pH 8.5), 200 mM of NaCl, 0.2% of SDS, 5 mM of EDTA, and 100 µg/ml of freshly added Proteinase K. Samples were centrifuged for 15 min at 6000 x *g* and the supernatant was transferred to a fresh tube or plate. Genomic DNA was isolated by adding an equal amount of isopropanol, mixing and subsequent centrifugation for 20 minutes at 6000 x *g*. Pellets were rinsed with 70% ethanol and dissolved in 400 µl H_2O. For PCR, 5 µl of a 50 times dilution in water was used.
PCR was carried out using a touchdown thermocycling program (92°C for 60 sec; 12 cycles of 92°C for 20 sec, 65°C for 20 sec with a decrement of 0.4°C per cycle, 72°C for 30 sec; followed by 20 cycles of 92°C for 20 sec, 58°C for 20 sec and 72°C for 30 sec; 72°C for 180 sec; GeneAmp9700, Applied Biosystems, Foster City, CA). PCR reaction mixes contained 5 µl genomic DNA, 0.2 µM forward primer and 0.2 µM reverse primer, 200 µM of each dNTP, 25 mM Tricine, 7.0% Glycerol (w/v), 1.6% DMSO (w/v), 2 mM $MgCl_2$, 85 mM Ammonium acetate pH 8.7 and 0.2 U Taq Polymerase in a total volume of 10 µl.
PCR products were diluted with 25 µl water and 1 µl was used as template for the sequencing reactions. Sequencing reactions, containing 0.25 µl BigDYE (v1.1; Applied Biosystems), 3.75 µl 2.5x dilution buffer (Applied Biosystems) and 0.4 µM gene specific primer in a total volume of 10 µl, were performed using cycling conditions recommended by the manufacturer. 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). Sequences were analyzed for presence of polymorphisms using PolyPhred [20]. Primers for PCR amplification and sequencing were designed using the Ensembl genome database (http://www.ensembl.org) and a customized interface to Primer3 [21].

Marker selection and testing for mapping
The panel of SNP markers, polymorphic between Brown Norway and Wistar rat strains, was extracted from rat CASCAD candidate SNP database [19] (http://cascad.niob.knaw.nl) and verified in the six parents (two tornado males and four BN females) of the mapping cross by sequencing. The rat genome was virtually divided into 90 bins of similar size and the final SNP panel contained 84 verified SNPs representing 67 of these bins. The cross scheme already elucidated the autosomal recessive manner of inheritance for the mutation, so only markers on the autosomes were included to map the mutation. In total, the SNP distribution patterns for 67 mutant F2 animals from the mapping cross were determined.

Scanning electron microscopy (SEM)
Tornado animals (*Myo7a*$^{tnd\text{-}1Hubr}$/*Myo7a*$^{tnd\text{-}1Hubr}$) (n=7) and phenotypically normal littermates (*Myo7a*$^{tnd\text{-}1Hubr}$/+ or +/+) (n=6) of 6 days, 10 days, 20 days, and 13 weeks of age were examined. Rats were anaesthetized by i.p. injection of Nembutal (60 mg/kg) and intracardially perfusion-fixed with 2.5% glutaraldehyde in 0.1 M sodium-cacodylate buffer (pH 7.4). Inner ears were dissected and postfixed overnight in the same fixative. After fixation for 1 hour in 2% osmium tetroxide in 0.1 M sodium cacodylate buffer, the specimens were dehydrated, critical point dried and sputter coated with gold. Specimens were examined in a Jeol 6310 scanning electron microscope operating at 15 kV.

Light Microscopy
Pigmented tornado animals (*Myo7a*$^{tnd\text{-}1Hubr}$/*Myo7a*$^{tnd\text{-}1Hubr}$) (n=2), derived from a cross with Brown Norway animals and having a crossing over between the albino and tornado locus, and heterozygous, pigmented littermates (*Myo7a*$^{tnd\text{-}1Hubr}$/+) (n=2) of 5 weeks of age were sacrificed. The eyes were dissected, fixed in buffered formalin overnight, and embedded in paraffin. Sections of 6 µm through the retina were prepared and stained with hematoxylin/eosin and examined using a Nikkon Eclipse 6600 light microscope.

ABR measurements

Auditory brainstem response measurements were performed in a soundproof room with low reverbation. Needle electrodes were placed on M1 and M2 (left and right mastoids) and referred to Cz (vertex) to record auditory evoked potentials. A ground electrode was placed halfway the tail of the rat. Interelectrode impedances were measured before and after each measurement (< 8 kOhm). Click stimuli were presented in a soundfield by placing the loudspeakers 5 cm in front of each ear. The loudness levels at the position of the ear were measured and calibrated with a Bruel and Kjaer 2203 sound pressure level meter (SPL). All thresholds were afterwards corrected for the soundfield setup by -7 dB (SPL). Before the measurements were done, the rats were i.p. injected with Nimatek (100 mg/kg) anaesthetics. A standard-evoked potential recording system (Synergy, Oxford Instruments) was used to present 100 μsec click stimuli with a fixed stimulation rate of 20 Hz. The analysis time was set at 15 msec from the onset of the click with 1.5 msec prestimulus time in order to assess the baseline. The recorded EEG signals were high-pass filtered at 100 Hz and low-pass filtered at 3 kHz; an automatic artifact rejection and a 60 Hz notch-filter were used to obtain auditory brainstem responses from contra- and ipsilateral stimulation sites. The EEG signals were averaged for different stimulation levels according to standard audiometrical top-down procedures, starting at 90 dB (SPL), uncorrected for the soundfield. Peaks were identified according to the Jewett and Williston nomenclature [22]. The auditory hearing threshold was defined as the level (in decibels) at which no reproducible responses were visually recognized in the responses obtained from the ipsilateral measured ear.

Results

Mutant phenotype

Rat models for human disease traditionally result from breeding of inbred strains that were specifically selected for naturally occurring aberrant traits [9]. We performed ENU mutagenesis in the rat [4] to increase the amount of genetic variation and conducted a small-scale screen for phenotypes caused by recessive mutations. Eight male and eight female F1 animals derived from four ENU-treated founders were paired. Eighteen brother-sister matings for inbreeding-induced recessive mutations were set up from the resulting F2 progeny. Animals were visually inspected for abnormalities and a single family was found in which three out of the eight F3 animals show an aberrant, circling behavior (Fig. 1a; Supporting movie 1 at http://www.niob.knaw.nl/researchpages/cuppen/publications/tornado) and were therefore named tornado. The mutant phenotype was confirmed to be inherited in an autosomal recessive manner by the second litter of that cross, producing two tornado animals amongst seven animals (Fig. 1). Although initially only male animals with the tornado phenotype were identified in the F3 generation, this should be considered a coincidence, since later matings using other animals also yielded female tornado animals, in a normal Mendelian ratio (data not shown).

More detailed analysis of the tornado rats revealed, besides circling behavior, additional abnormalities. Deafness in the mutants was observed by a negative Preyer's reflex. Moreover, the mutant rats show locomotory hyperactivity, stereotypic head flicking, and vestibular dysfunction, which became evident in a swimming test (Supporting movie 2 at http://www.niob.knaw.nl/researchpages/cuppen/publications/tornado). The complete phenotype strongly resembles the mouse shaker [23,24], waltzer [25,26], and whirler [27] phenotypes for which a variety of underlying gene defects have been identified.

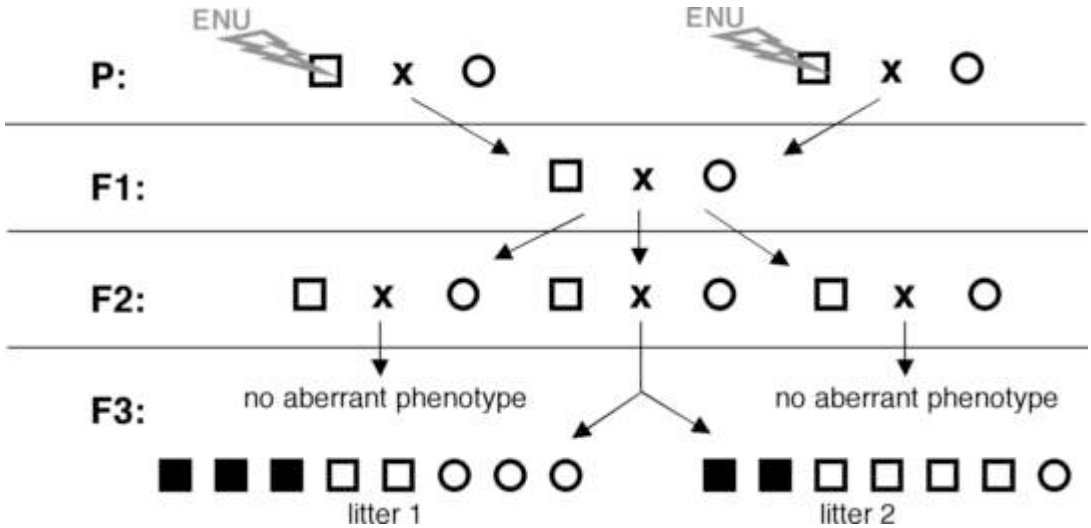


Figure 1: Overview of the crosses that resulted in the identification of the tornado mutant.
Four ENU mutagenized males were used to generate F1 animals by crossing to untreated females. Sixteen breeding pairs were selected from this progeny to produce the F2 generation. Brother-sister matings were set up for these animals to breed induced mutations to homozygosity to reveal potential recessive phenotypes. The figure represents all matings set up for a specific F1 pair (the others did not result in any visual aberrant phenotype). One out of the three F2 matings resulted in progeny with an aberrant phenotype that was named tornado. Genetic inheritance of the phenotype was confirmed in a second mating. Squares and circles represent males and females, respectively. Solid symbols indicate animals with the tornado phenotype.

Mapping cross

To identify the genetic defect underlying the rat tornado phenotype, we set up a mapping cross using a Brown Norway background. The causal mutation was introduced in a Wistar background [4] and the Brown Norway strain is genetically most distant as compared to other commonly used laboratory rat strains [28,29], making it the best strain for mapping purposes. Ten F1 intercrosses were set up to restore the phenotype. In the first mating round, we obtained 117 F2 animals, of which 34 displayed the tornado phenotype (not shown). Strikingly, all tornado animals were found to be albino, whereas none of the pigmented (brown or black) animals showed the phenotype, indicating linkage to the gene causing albinism in the Wistar strain (c), which is located on chromosome 1 (*Tyr*). Two albino animals did not show the tornado phenotype. Considering these animals to be crossing overs, the genetic distance between the mutation and *Tyr* was estimated to be ~3 cM.

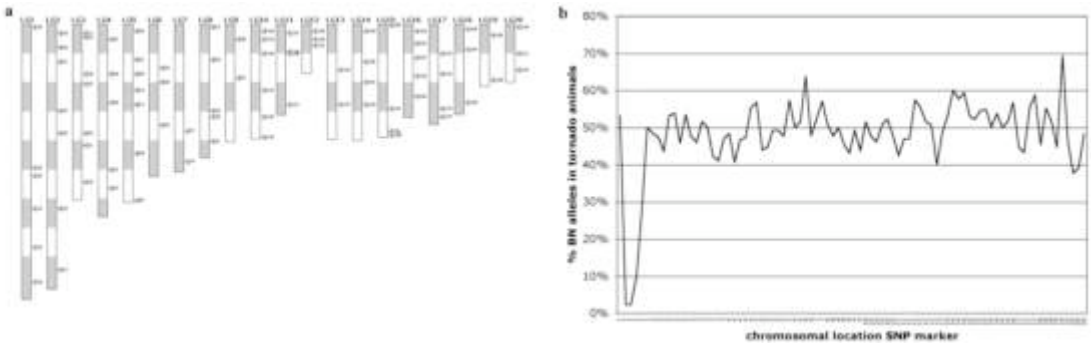


Figure 2: SNP-based mapping of the tornado mutation.
a) Distribution on the rat physical map of the 84 verified SNP markers used for mapping the tornado mutation. SNPs were selected from equally sized genomic segments (represented in alternating white and grey). X chromosome is not depicted. **b)** Graphical representation of linkage of the mutant phenotype to the central region of chromosome 1. The polymorphic SNP markers are plotted on the x-axis ordered by chromosome location. The graph shows the percentage of BN alleles, calculated from 67 genotyped tornado animals obtained from the mapping cross. The tornado mutation was introduced in the Wistar background.

Mapping the tornado mutation with whole genome SNP mapping panel

A genome-wide SNP mapping panel (Brown Norway *vs.* Wistar) was constructed to independently confirm linkage to chromosome 1. Therefore, the rat genome was distributed in 90 bins of equal size from which candidate SNPs [19] (http://cascad.niob.knaw.nl) were selected. Although some candidate SNPs could be selected on the basis of Wistar mRNA data *vs.* Brown Norway shotgun (WGS) sequencing data, the majority of the candidate SNPs were mined from Sprague Dawley EST data *vs.* Brown Norway WGS data. All candidate SNPs were tested in the six founders (two Wistar males and four BN females) of the mapping cross, resulting in a final SNP panel containing 84 verified SNPs distributed over 67 different bins (Fig. 2a).
A total of 67 tornado F2 animals (34 from the initial mapping cross and 33 from later crosses) were genotyped for these SNPs (Supporting Fig. S1), confirming strong linkage to the central region of chromosome 1, where also the albino locus is located (Fig. 2b). The tornado mutation localizes between markers rs8156543 and rs8173521 (dbSNP: http://www.ncbi.nlm.nih.gov/SNP/). With three crossing-overs in 65 animals between rs8173521 and the tornado mutation, the estimated genetic distance is ~2.3 cM proximal to rs8173521.

Table 1: Candidate deafness genes mapping to rat chromosome 1.

Locus	Gene	Human	Mouse	Rat	Reference
DFNA10	*EYA4*	6:133.5	10:22.8	1:22.8	[39]
DFNB2	*MYO7A*	11:76.6	7:86.5	1:155.4	[30,35]
DFNB7/11	*TMC1*	9:70.8	19:20.1	1:224	[40]
DFNB18	*USH1C*	11	7:34.8	1:96.8	[41,42]
DFNB22	*OTOA*	16	7:109.4	1:180.1	[43]

Candidate deafness genes that map to rat chromosome 1 were selected from Hereditary Hearing Loss Homepage (http://www.uia.ac.be/dnalab/hhh/). Albino is located on Chromosome 1 at 143.7 Mbp. Closest to albino is *Myo7a* in all 3 species. Chromosomal locations are in accordance with Ensembl Rat Genome Build release version 3.1.

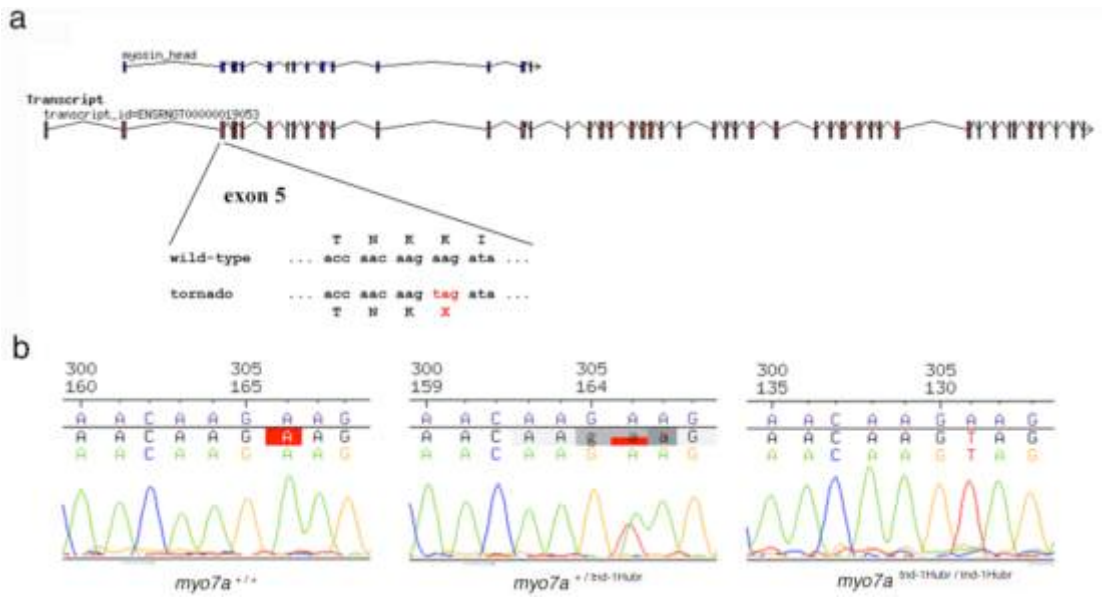


Figure 3: The tornado phenotype is caused by a premature stopcodon in *Myo7a*.
a) Schematic organization of *Myo7a* in the rat and the position and context of the identified tornado mutation (*Myo7a*$^{tnd\text{-}1Hubr}$). The exons encoding the myosin head are indicated on top of the graph. **b)** Sequence traces of the mutated position in exon 5 of a homozygous wild-type (upper), heterozygous (middle), and mutant (lower) animal.

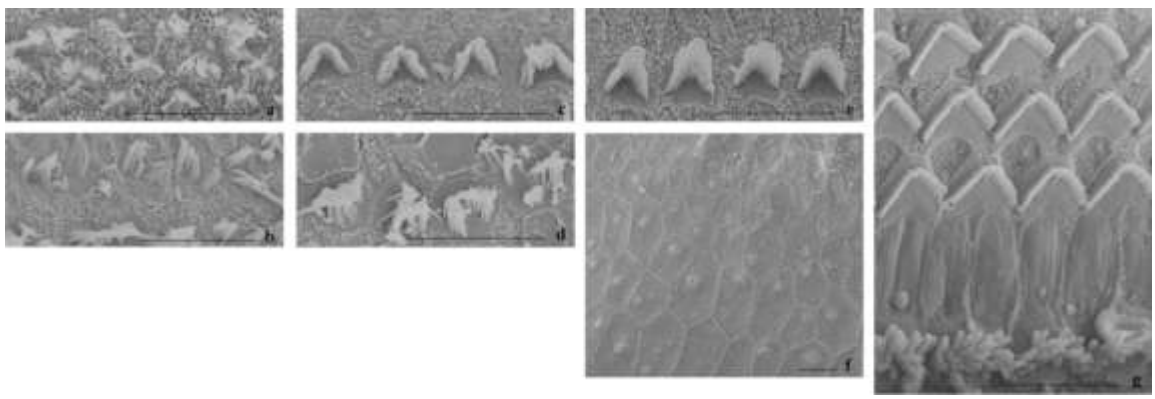
Identification of the mutation by candidate gene approach

To identify candidate genes in the remaining genomic interval, we mapped the orthologs of known human and mouse deafness genes on the rat genome (Table 1). Five human genes were found to have orthologs on rat chromosome 1, with Myo7a, encoding an unconventional myosin, closest to the albino gene and the rs8173521 SNP marker in the rat. Resequencing all 47 coding exons of Myo7a (Ensembl ID: ENSRNOT00000019053; Fig. 3a) revealed an A to T transversion (position 362) that completely segregates with the tornado phenotype in all crosses (Fig. 3b). This mutation (Myo7a$^{tnd\text{-}1Hubr}$) introduces a premature stopcodon in exon 5 in the middle of the myosin head encoding sequence and thereby most likely results in a complete loss of function of the gene. Humans carrying mutations in MYO7A are known to suffer from Usher syndrome type IB, which is characterized by a profound congenital sensorineural hearing loss, constant vestibular dysfunction and prepubertal onset of retinitis pigmentosa [30]. Most human disease causing mutations, including premature stopcodons, deletions, and missense mutations, are located in the amino-terminal end of the motor domain of the protein. The first alleles of the shaker-1 mice that were cloned, were found to be two missense mutations and a splice acceptor site mutation in the region encoding the myosin head of Myo7a [24].

Deafness and stereocilia disorganization in the tornado mutant

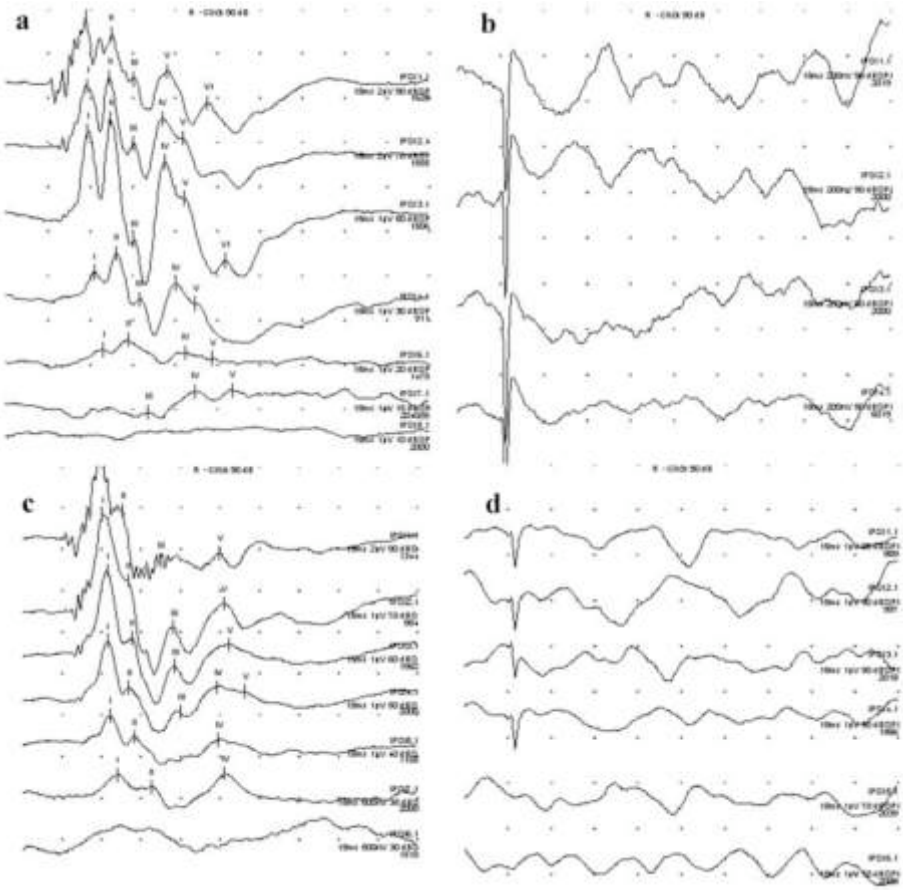
Scanning electron microscopy (SEM) on neuroepithelium of dissected cochlea, i.e. of the upper surface of the hair cells of the organ of Corti was used to investigate the cause of deafness. Normally, development of rat inner ear microanatomy and physiology leading to hearing starts around postnatal day 4 or 5 and hearing maturation is completed around postnatal day 20 [31]. In the mutant rats, we observed disorganization, destruction and fusing of the hair cell stereocilia, which was most pronounced on the outer hair cells, as early as 6 days after birth (Fig. 4a). The hair cell stereocilia bundles degenerate progressively over a short time (Fig. 4b and d; controls Fig. 4c and e, respectively), and in adult mutant rats no stereocilia can be detected anymore on the outer as well as inner hair cells (Fig. 4f; control Fig. 4g). An irregular epithelium remains in the organ of Corti region (Fig. 4f). In addition, the stereocilia degeneration progresses from the basal to apical turn throughout the neuroepithelium of the cochlea (data not shown), suggesting complete deafness over the whole audible frequency range.

Figure 4: Degeneration of the organ of Corti in tornado rats.
Scanning Electron Microscopy (SEM) micrographs of the surface of the organ of Corti of tornado animals (*Myo7a*$^{tnd\text{-}1Hubr/tnd\text{-}1Hubr}$; a, b, d, f) and non-mutant littermates (*Myo7a*$^{tnd\text{-}1Hubr/+}$ or *Myo7a*$^{+/+}$; c, e, g) at six days (a), 10 days (b, c), 20 days (d, e), and 13 weeks (f, g) of age. At six days of age the outer hair cell stereocilia are already disorganized and fused (a) and degenerate progressively (b, d, f) even on the inner hair cells (f). Eventually, an irregular epithelium remains in the organ of Corti region (f). In all panels, scale bars represent 10 μm.



Air conducted ABR can be recorded as early as postnatal day seven to eight [32]. We explored ABR measurements on tornado and wild-type rats of 10 days and four weeks of age. The wild-type rats of four weeks of age displayed normal ABR patterns, with a detectable −7 dB soundfield corrected threshold at 8 dB (SPL) (n=1) (Fig. 5a). In the mutant animals, no reproducible responses were observed at 83 dB (SPL) (n=2) (Fig. 5b). Wild-type juvenile rats of 10 days of age showed click-evoked responses between 18 and 43 dB (SPL) threshold level (n=4) (Fig. 5c: example of an auditory brainstem threshold at 28 dB), but mutant littermates failed to show any reproducible click-responses with settings up to 83 dB (SPL) (n=1) (Fig. 5d). Combining the observations of onset of stereocilia disorganization and no detectable responses to 83 dB click stimuli at 6 days of age and the severe stereocilia degeneration at 10 days of age, we conclude that tornado rats are most likely to be completely deaf throughout their life.

Figure 5: Profound deafness in tornado rats.
Typical Auditory Brainstem Responses from mutant tornado animals (*Myo7a*$^{tnd\text{-}1Hubr/tnd\text{-}1Hubr}$) and wild-type heterozygous littermates (*Myo7a*$^{tnd\text{-}1Hubr/+}$). **a)** 4 weeks of age, wild-type **b)** 4 weeks of age, mutant **c)** 10 days of age, wild-type **d)** 10 days of age, mutant. Horizontal axis: time window of 15 msec (1.5 ms/division); vertical axis: amplitude in μV. Stimulus onset at 1.5 ms. The left graphs (a, c) show good brainstem responses obtained at different stimulation levels according to standard audiometrical descending top-down procedures until the level that no reproducible responses are recognized (i.e. auditory threshold); all responses shown in the right graphs (b, d) reveal no reproducible brainstem responses at maximum stimulation level of 83 dB (SPL), including soundfield correction.



Eye phenotype

Prepubertal onset of blindness in USH1B patients is caused by degeneration of the retina (retinitis pigmentosa). The transport of opsin to the outer segment of photoreceptor cells seems especially critical for viability of these cells [33]. It has been suggested that Myo7A coparticipates in opsin transport [34], but it remains still unclear how deficiency of functional Myo7A causes retinal degeneration in human patients. Moreover, for mice lacking functional Myo7A, retinal degeneration has never been diagnosed. Currently, however, two retinal phenotypes have been

identified by microscopy in mice. First, Liu *et al.* [35] found absence of melanosomes in apical processes of the retinal pigment epithelium in Myo7A-deficient mice [35]. The second mutant phenotype is a slower rate of photoreceptive disk membrane renewal in the outer segments [34], caused by perturbed phagocytosis of disk membranes [36].
Similar to Liu *et al.*, we found absence of melanosomes in the apical processes of retinal pigment epithelium in a tornado background at five weeks of age (Fig. 6a; control Fig. 6b). However, both pigmented and albino tornado rats did not show any signs of retinal degeneration at 20 weeks of age (data not shown).

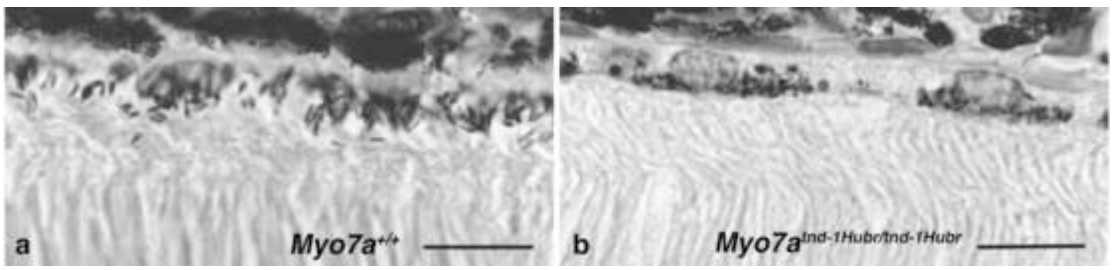


Figure 6: Melanosome migration defect in tornado rats.
Histological staining (hematoxylin/eosin) of the eyes of pigmented non-mutant (a) and tornado (b) animals, revealing the absence of melanosomes in the apical processes of the retinal pigment epithelium. Scale bars represent 10 μm.

Discussion

We initially established ENU mutagenesis conditions for the rat for setting up gene-driven knockout technology using target-selected mutagenesis [4]. However, the same F1 animals are also suited for forward genetic, phenotype-driven, approaches. Indeed, several phenotypes caused by dominant mutations were readily identified in our experiments [4] as well as by others [5]. Gould and colleagues [5] identified 74 visually aberrant mutants in a screen for dominant phenotypes in nearly 5,000 F1 progeny from ENU-treated Sprague Dawley rats. About half of them were found to be fertile and inherit the phenotype. Here, we describe a small-scale study on recessive phenotypes after ENU mutagenesis in the rat and the subsequent cloning of the mutated gene using a novel SNP mapping panel. Although, the SNP mapping panel is of relatively low density and specifically designed for mapping crosses between Wistar and Brown Norway, it can easily be adapted or expanded with more SNPs, since the CASCAD and public databases now harbor over 50,000 candidate and validated SNPs.
The genome scan carried out using our SNP panel pointed towards a region on chromosome 1 containing *Myo7a*, the ortholog of human Usher syndrome type 1B gene. Resequencing of the coding region of this gene in the tornado rat identified a nonsense mutation in the middle of the core domain, the myosin head. Human patients with Usher syndrome type 1B, also harboring mutations in the myosin head, suffer from profound congenital deafness [30]. Shaker-1, the mouse model for Usher syndrome type 1B, also displays severe congenital hearing impairment due to typical neuroepithelial-type cochlear defects [24]. In 1998, Self *et al.* [37] described the correlation between the nature of the mutation, the severity of stereocilia disorganization, and electrophysiological response in three different alleles of *Myo7a* in the mouse. The most severe

disruption of development of the stereocilia was observed in *Myo7a*^{816SB}, a mutant lacking 10 amino acid residues of the core of the motor head. These animals showed complete absence of stimulus-related cochlear potentials. Mice with a relatively mild shaker-1 phenotype (*Myo7a*^{6J}) already show prenatal disorganization of stereocilia in the organ of Corti, whereas the microvilli at the upper surface of the hair cell, from which the stereocilia develop, are unaffected at embryonic day 16.5. The tornado rat allele described here (*Myo7a*$^{tnd\text{-}1Hubr}$), is likely to resemble the severe mouse and the common human alleles, as the premature stopcodon in the beginning of the gene is expected to result in complete loss of gene function. Indeed, in the tornado rat, stereocilia disorganization is obvious at 6 days of age and results in complete degeneration of cochlear hair cells within 1 month after birth. Furthermore, measurements of brainstem responses upon auditory stimuli suggest that tornado animals are deaf throughout life.

Taken together, the identification of the first rat model for Usher syndrome type 1B may contribute to the further understanding of the molecular mechanisms underlying Usher syndrome type IB and healthy and diseased inner ear and eye function in general. A major advantage of a rat model over a mouse model for studying inner ear function may be the size and accessibility of the the organ, for example for cochlear implant studies [38]. Finally, the *Myo7a*$^{tnd\text{-}1Hubr}$ mutant described here, is the first rat model induced by ENU mutagenesis that is cloned by forward genetics, providing proof of principle for systematic forward genetics in the rat. Large-scale phenotype-driven screens in the rat have the potency to result in important new insights in the function of genes in the development of complex disease and higher brain functions, for which the rat currently is the best-suited model organism.

Acknowledgements

We thank J. Korving for histological sections on retinas and J. Curfs for help with the ABR measurements. This work was supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics and NV Organon.

References

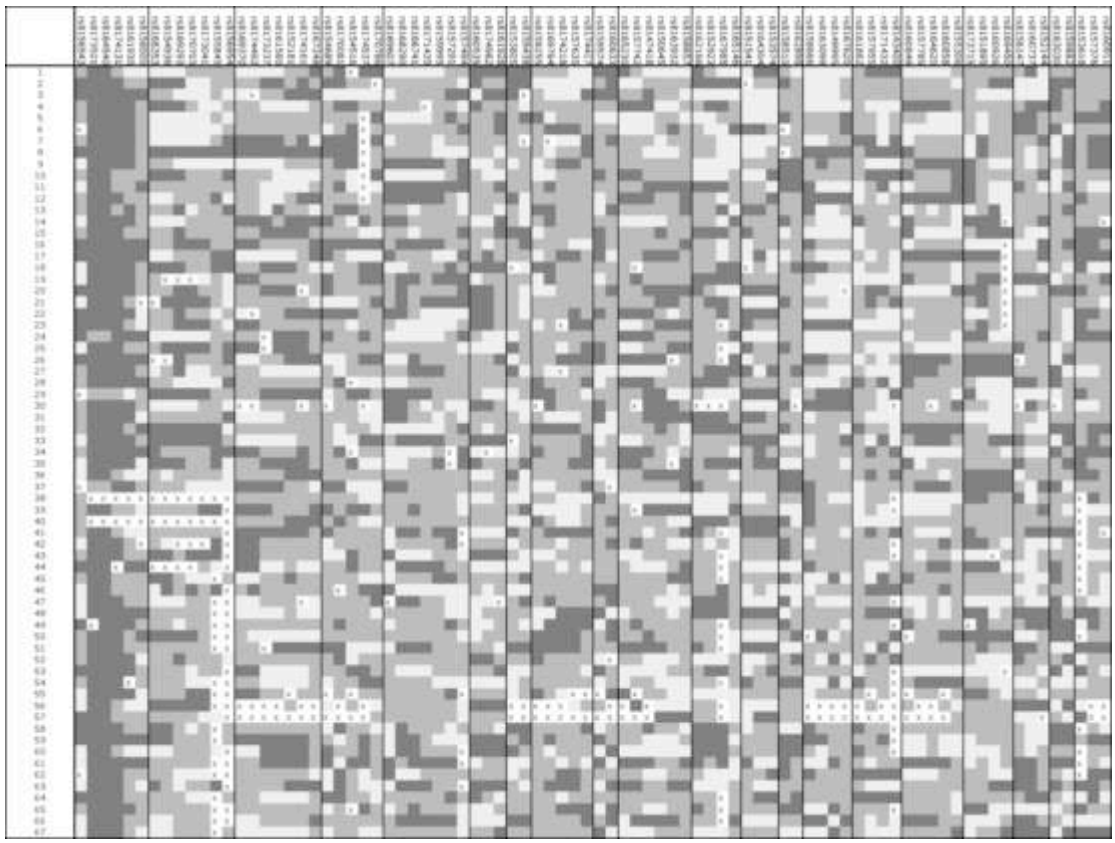
1. Gibbs, R.A. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428, 493-521.
2. Bockamp, E. *et al.* (2002) Of mice and models: improved animal models for biomedical research. *Physiol Genomics*, 11, 115-132.
3. van der Weyden, L. *et al.* (2002) Tools for targeted manipulation of the mouse genome. *Physiol Genomics*, 11, 133-164.
4. Smits, B.M. *et al.* (2004) Target-selected mutagenesis of the rat. *Genomics*, 83, 332-334.
5. Zan, Y. *et al.* (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol*, 21, 645-651.
6. O'Brien, T.P. and Frankel, W.N. (2004) Moving forward with chemical mutagenesis in the mouse. *J Physiol*, 554, 13-21.
7. Keays, D.A. and Nolan, P.M. (2003) N-ethyl-N-nitrosourea mouse mutants in the dissection of behavioural and psychiatric disorders. *Eur J Pharmacol*, 480, 205-217.
8. Cox, R.D. and Brown, S.D. (2003) Rodent models of genetic disease. *Curr Opin Genet Dev*, 13, 278-283.

9. Hedrich, H.J. (2000) History, Strains and Models. In *The Laboratory Rat* (Krinke, G.J., ed.), pp. 3-16, Academic Press.
10. Jacob, H.J. and Kwitek, A.E. (2002) Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet*, 3, 33-42.
11. Greenhouse, D.D. *et al.* (1990) Catalogue of Inbred Strains of Rats. In *Genetic monitoring of inbred strains of rats* (Hedrich, H.J., ed), pp. 410-480, Gustav Fisher.
12. Jacob, H.J. *et al.* (1992) Genetic dissection of autoimmune type I diabetes in the BB rat. *Nat Genet*, 2, 56-60.
13. Shepel, L.A. *et al.* (1998) Genetic identification of multiple loci that control breast cancer susceptibility in the rat. *Genetics*, 149, 289-299.
14. Rapp, J.P. (2000) Genetic analysis of inherited hypertension in the rat. *Physiol Rev*, 80, 135-172.
15. Brown, S.D. and Nolan, P.M. (1998) Mouse mutagenesis-systematic studies of mammalian gene function. *Hum Mol Genet*, 7, 1627-1633.
16. Brown, S.D. and Balling, R. (2001) Systematic approaches to mouse mutagenesis. *Curr Opin Genet Dev*, 11, 268-273.
17. Chwalisz, W.T. *et al.* (2003) The circling behavior of the deafblind LEW-ci2 rat is linked to a segment of RNO10 containing Myo15 and Kcnj12. *Mamm Genome*, 14, 620-627.
18. Kuramoto, T. *et al.* (2004) Rat neurological mutations cerebellar vermis defect and hobble are caused by mutations in the netrin-1 receptor gene Unc5h3. *Brain Res Mol Brain Res*, 122, 103-108.
19. Guryev, V. *et al.* (2004) Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res*, 14, 1438-1443.
20. Nickerson, D.A. *et al.* (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res*, 25, 2745-2751.
21. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-386.
22. Jewett, D.L. and Williston, J.S. (1971) Auditory-evoked far fields averaged from the scalp of humans. *Brain*, 94, 681-696.
23. Wang, A. *et al.* (1998) Association of unconventional myosin MYO15 mutations with human nonsyndromic deafness DFNB3. *Science*, 280, 1447-1451.
24. Gibson, F. *et al.* (1995) A type VII myosin encoded by the mouse deafness gene shaker-1. *Nature*, 374, 62-64.
25. Wilson, S.M. *et al.* (2001) Mutations in Cdh23 cause nonsyndromic hearing loss in waltzer mice. *Genomics*, 74, 228-233.
26. Alagramam, K.N. *et al.* (2001) The mouse Ames waltzer hearing-loss mutant is caused by mutation of Pcdh15, a novel protocadherin gene. *Nat Genet*, 27, 99-102.
27. Mburu, P. *et al.* (2003) Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31. *Nat Genet*, 34, 421-428.
28. Canzian, F. (1997) Phylogenetics of the laboratory rat Rattus norvegicus. *Genome Res*, 7, 262-267.
29. Thomas, M.A. *et al.* (2003) Phylogenetics of rat inbred strains. *Mamm Genome*, 14, 61-64.
30. Weil, D. *et al.* (1995) Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature*, 374, 60-61.

31. Pujol, R. *et al.* (1998) Development of sensory and neural structures in the mammalian cochlea. In *Development of the Auditory System* (Rubel, A., Popper, N., and Fay, R.R., eds.), Springer-Verlag.
32. Geal-Dor, M. *et al.* (1993) Development of hearing in neonatal rats: air and bone conducted ABR thresholds. *Hear Res*, 69, 236-242.
33. Marszalek, J.R. *et al.* (2000) Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell*, 102, 175-187.
34. Liu, X. *et al.* (1999) Myosin VIIa participates in opsin transport through the photoreceptor cilium. *J Neurosci*, 19, 6267-6274.
35. Liu, X. *et al.* (1998) Mutant myosin VIIa causes defective melanosome distribution in the RPE of shaker-1 mice. *Nat Genet*, 19, 117-118.
36. Gibbs, D. *et al.* (2003) Abnormal phagocytosis by retinal pigmented epithelium that lacks myosin VIIa, the Usher syndrome 1B protein. *Proc Natl Acad Sci U S A*, 100, 6481-6486.
37. Self, T. *et al.* (1998) Shaker-1 mutations reveal roles for myosin VIIA in both development and function of cochlear hair cells. *Development*, 125, 557-566.
38. Vischer, M. *et al.* (1997) Effect of high-frequency electrical stimulation of the auditory nerve in an animal model of cochlear implants. *Am J Otol*, 18, S27-29.
39. Wayne, S. *et al.* (2001) Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus. *Hum Mol Genet*, 10, 195-200.
40. Kurima, K. *et al.* (2002) Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. *Nat Genet*, 30, 277-284.
41. Ouyang, X.M. *et al.* (2002) Mutations in the alternatively spliced exons of USH1C cause non-syndromic recessive deafness. *Hum Genet*, 111, 26-30.
42. Ahmed, Z.M. *et al.* (2002) Nonsyndromic recessive deafness DFNB18 and Usher syndrome type IC are allelic mutations of USHIC. *Hum Genet*, 110, 527-531.
43. Zwaenepoel, I. *et al.* (2002) Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB22. *Proc Natl Acad Sci U S A*, 99, 6240-6245.

Supporting Material

Supporting Figure S1: Detailed overview of marker distribution in 67 mutant animals of the mapping cross. Markers are organized by chromosomal position.
Dark grey squares represent homozygous Wistar alleles. Intermediate grey squares represent heterozygous alleles. Bright grey squares represent homozygous BN alleles. Squares with x are not analyzed. It is obvious that around the centre of chromosome 1, there is an overrepresentation of Wistar alleles, indicating linkage to that subchromosomal region.



CHAPTER 8

Discussion: Speculation on Increasing the Efficiency of ENU Mutagenesis in the Rat

Speculation on Increasing the Efficiency of ENU Mutagenesis in the Rat

Bart M. G. Smits

Functional Genomics Group, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

ENU mutagenesis is now available for the rat as a tool to interfere with endogenous gene expression. The efficiency, however, is not high enough yet to generate any knockout on demand. In this chapter, I will discuss possibilities to increase the efficiency ENU mutagenesis in the rat.

The rat model system is primarily of great value for the study of human complex diseases [1], like hypertension, diabetes, and many psychiatric disorders. QTL-mapping approaches using affected strains and unaffected strains have identified many genomic regions harboring potential functional polymorphisms [2]. While fine-mapping approaches are continuously ongoing to narrow down QTL regions even further, the final step of identifying the causal polymorphism is the most challenging part (this thesis; introduction). For this last stage, technologies to manipulate the genome, like transgenesis, RNAi-mediated gene silencing, and ENU mutagenesis, are required. In this thesis, three pioneering studies on the use of ENU mutagenesis as a tool for genome manipulation in the rat are described.

Currently, ENU mutagenesis is the only approach to generate genetic knockouts in the rat. The success of the ENU mutagenesis approach largely depends on two factors: ENU mutagenesis efficiency and mutation screening capacity. Although we [3] and others [4] demonstrated that ENU mutagenesis can be used in rat to generate knockouts and mutants, the efficiency of the mutagenesis does not allow the generation of a specific knockout in any gene of interest yet. In chapter 3, we calculated that with the current mutation rate of 1 in 1,2 million bp, about 50,000 animals need to be generated to obtain a specific knockout of an average-sized gene with 96% certainty. Since no lab in the world can afford breeding such a number of rats in a relatively short time period, and sperm freezing is still not routinely accessible, the need for an increased mutation rate seems unavoidable.

Both chapters 2 and 3 describe optimal mutagenesis conditions for several rat strains. The optimal doses for these strains are determined as the highest dose at which the injected males still regain fertility. We demonstrated that injection of higher doses of ENU would result in permanent sterility, caused by depletion of sperm cells [3]. It is unknown if spermatogonial cell death is related to mutagenesis and mutation rates [5], however, in zebrafish a three to five times higher mutation rate in the F1 population can be achieved [6], illustrating that vertebrate genomes can tolerate much higher mutation loads without going into apoptosis. Although the explanation for higher mutation rate in zebrafish may be found in the different way of ENU administration, there is ample evidence that the DNA repair system is involved. Mouse data has shown that after saturating the DNA repair system, the relation between ENU dose and induced mutations is linear, implicating a crucial role of DNA repair in mutagenesis [7]. The exact role of the DNA repair system is complicated, since different repair systems with different efficiencies and precision may be used within a spermatogonial stem cell population depending on the type of lesion and stage of the cell cycle [7,8].

In our last mutagenesis screen (Chapter 3), we isolated a rat mismatch repair protein 6 (MSH6) knockout. MSH6 is part of the mismatch repair (MMR) system, a DNA repair system primarily responsible for post-replication correction of nucleotide mispairs and extra-helical loops [9]. By employing ENU mutagenesis in MSH6-deficient background, we hypothesize to achieve a higher mutation rate, possibly reaching up to the zebrafish mutation rate. Evidence for this is provided by MSH6-deficient mice cell extracts that are capable of repairing 1-, 2-, and 4-base IDLs (insertion-deletion loops), but are deficient in the repair of single-base mismatches [10,11]. In addition, MSH6-deficient mouse small intestinal epithelium is found to have a strong elevated mutation frequency, predominantly consisting of G:C to A:T transitions [12]. This type of mutations currently comprises 13% of the ENU-induced mutation frequency and is expected to increase substantially in the MSH6 knockout rat.

MSH6-deficient mice show normal fertility, but also a reduction of life expectancy, due to increased tumor formation [13]. It remains to be determined whether MSH6-deficient rats live long enough to undergo ENU mutagenesis and generate sufficient offspring to establish an F1 library. To test this hypothesis, the optimal ENU mutagenesis conditions in MSH6-deficient background needs to be determined.

Another component of the DNA repair system that upon blocking forms a good candidate to enhance the mutation rate is O6-methylguanine-DNA alkyltransferase (MGMT), since the enzyme is efficient at repairing alkylated guanine adducts in mammals [14]. A rat knockout for this gene does not exist yet, but mouse knockouts are hypersensitive towards the methylating agent MNU leading to almost 10-fold lower LD50 value [15]. This also suggests that the use of ENU in MGMT-deficient background may cause a major reduction in life expectancy. Nevertheless, pharmacological suppression of the enzyme by O6-benzylguanine (O6-BG) together with ENU mutagenesis may cause temporal sensitivity towards ENU, possibly not resulting in excessive tumor formation. ENU is highly effective in spermatogonia and such combinatory treatment may give substantially increased mutation rates.

Besides high mutation rates, efficient screening technology is a prerequisite for a successful target-selected mutagenesis approach. In this thesis, two different techniques have been used for mutation as well as SNP discovery, namely CEL I-mediated heteroduplex cleavage [16] and resequencing [17,18]. We have used CEL I-mediated heteroduplex cleavage successfully to discover the first 17 ENU-induced mutations (Chapter 2; [3]) and 103 SNPs in 96 strains (Chapter 3; [19]). This method was initially used as a resequencing surrogate, since it was more cost-effective at that time. In addition, pooling of PCR products up to four times allowed higher throughput compared to resequencing. However, the increased throughput and spectacularly more cost efficient operation of the next generation 96-capillary sequencing machines have put dideoxy resequencing back into competition with the CEL I method. Additional advantages, like higher accuracy and robustness, increased possibilities for automation and superior logistic characteristics, have put resequencing forward as the method of choice for mutation discovery for the high-throughput target-selected mutagenesis setup at the Hubrecht Laboratory. Hence, we found 120 novel mutations (Chapter 3) and 861 novel SNPs in 36 strains (Chapter 5) using the resequencing approach.

Despite the advantages and declining costs of PCR-based capillary resequencing, emerging new technologies may revolutionize the field of resequencing and mutation discovery [20]. Recently, novel array-based sequencing methods have been developed, which have in common a cyclic sequencing-by-synthesis method on spatially separated oligonucleotide features occurring massively in parallel [21,22]. Currently, only prokaryote genomes of 2 to 3 Mbp have been successfully resequenced. Although full resequencing of much more complex mammalian genomes of over 2,5 Gbp raises technological as well as computational challenges, these technologies provide promising future directions. Reducing the complexity of the genome to the regions of interest, e.g. by amplifying coding regions (in rat 28,4 Mbp), followed by pooling large

numbers of individuals and analysis in one go, may make these technologies currently already interesting.

Another emerging resequencing technology is hybridization-based sequencing. Affymetrix and Perlegen have pioneered on this approach by hybridizing sample DNA to microfabricated arrays of immobilized oligonucleotide probes [23,24]. For each base pair of a reference genome to be resequenced, there are four features on the chip, allowing discriminating between the four bases. Limiting the features on the array to the positions of interest would enable a researcher to scan a full genome within one hybridization experiment. In case of rat knockouts, we are only interested in positions in coding regions that could mutate into a premature stop codon given the rat ENU mutation spectrum. In Chapter 3, we calculated that roughly 1,6 million coding bases in the rat genome meet these criteria. Although current microarrays do not permit this amount of features, genotyping-based screening of multiple sets of positions of interest for every individual, possibly in a pooled fashion, would greatly enhance mutation discovery efficiency.

To summarize, proof-of-concept for ENU mutagenesis in the rat has been demonstrated, but to make the approach common technology for making knockouts for any gene of interest, additional innovations will be required. We expect that an increase in the mutation frequency and/or improved mutation screening technology will be able to meet these requirements in the near future.

References

1. Jacob, H.J. and Kwitek, A.E. (2002) Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet*, 3, 33-42.
2. Flint, J. *et al.* (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet*, 6, 271-286.
3. Smits, B.M. *et al.* (2004) Target-selected mutagenesis of the rat. *Genomics*, 83, 332-334.
4. Zan, Y. *et al.* (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol*, 21, 645-651.
5. Justice, M.J. *et al.* (2000) Effects of ENU dosage on mouse strains. *Mamm Genome*, 11, 484-488.
6. Wienholds, E. *et al.* (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res*, 13, 2700-2707.
7. Favor, J. (1998) The mutagenic activity of ethylnitrosourea at low doses in spermatogonia of the mouse as assessed by the specific-locus test. *Mutat Res*, 405, 221-226.
8. Vogel, E.W. and Natarajan, A.T. (1995) DNA damage and repair in somatic and germ cells in vivo. *Mutat Res*, 330, 183-208.
9. Bernstein, C. *et al.* (2002) DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat Res*, 511, 145-178.
10. Edelmann, W. *et al.* (2000) The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. *Cancer Res*, 60, 803-807.
11. Edelmann, W. *et al.* (1997) Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. *Cell*, 91, 467-477.
12. Mark, S.C. *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.
13. Wei, K. *et al.* (2002) Mouse models for human DNA mismatch-repair gene defects. *Trends Mol Med*, 8, 346-353.
14. Lindahl, T. *et al.* (1988) Regulation and expression of the adaptive response to alkylating agents. *Annu Rev Biochem*, 57, 133-157.
15. Glassner, B.J. *et al.* (1999) DNA repair methyltransferase (Mgmt) knockout mice are sensitive to the lethal effects of chemotherapeutic alkylating agents. *Mutagenesis*, 14, 339-347.

16. Oleykowski, C.A. *et al.* (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*, 26, 4597-4602.
17. Beier, D.R. (2000) Sequence-based analysis of mutagenized mice. *Mamm Genome*, 11, 594-597.
18. Wienholds, E. *et al.* (2002) Target-selected inactivation of the zebrafish rag1 gene. *Science*, 297, 99-102.
19. Smits, B.M. *et al.* (2004) Genetic variation in coding regions between and within commonly used inbred rat strains. *Genome Res*, 14, 1285-1290.
20. Shendure, J. *et al.* (2004) Advanced sequencing technologies: methods and goals. *Nat Rev Genet*, 5, 335-344.
21. Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*.
22. Shendure, J. *et al.* (2005) Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*.
23. Lipshutz, R.J. *et al.* (1999) High density synthetic oligonucleotide arrays. *Nat Genet*, 21, 20-24.
24. Patil, N. *et al.* (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 294, 1719-1723.

SAMENVATTING

Functionele Genoomanalyse in de Rat

Inleiding

Ratten worden door velen beschouwd als brutale, bijtgrage knaagdieren, die verantwoordelijk waren voor het verspreiden van ziekten als de builenpest en typhus. Anderen daarentegen houden tegenwoordig tamme ratten als huisdier. De rat wordt ook beschouwd als het eerste dier dat werd gefokt voor wetenschappelijke doeleinden. Er zijn namelijk rapporten uit begin negentiende eeuw bewaard gebleven, waarin voedingsexperimenten met ratten worden beschreven. Ook werd de rat als één van de eerste dieren onderworpen aan genetische experimenten om het overerven van bepaalde eigenschappen aan te tonen. In 1877 en 1885 werden ratten al gebruikt voor de studie naar erfelijkheid van vachtkleur. Later zou de laboratoriummuis de rol van genetisch model organisme overnemen, omdat zijn beperkte grootte zich meer leent voor het kweken in grotere aantallen. De rat zou vanwege zijn grotere omvang vooral een model dier voor fysiologie, anatomie en chirurgie worden.
Tegenwoordig is de laboratoriumrat (*Rattus norvegicus*) één van de meest gebruikte proefdieren voor het bestuderen van vele verschillende aspecten van menselijke gezondheid en ziekten, zoals hoge bloeddruk, diabetes, neurologische aandoeningen, als depressiviteit, schizofrenie, angststoornissen, maar ook aanleg voor kanker. Daarnaast worden ratten veelvuldig gebruikt in de geneesmiddelenindustrie om de effecten en bijeffecten van nieuwe geneesmiddelen te testen.
Van de laboratorium rat bestaan meer dan 230 verschillende inteeltstammen, waarvan de eerste al dateert uit 1909. Alle individuen van een inteeltstam worden als genetisch identiek beschouwd. Inteeltstammen zijn gemaakt door ratten uit een zeer diverse populatie te selecteren op een bepaalde eigenschap, bijvoorbeeld hoge bloeddruk of hoog bloedsuikergehalte. De ratten met de hoogste bloeddruk worden onderling gekruist en de nakomelingen zullen weer sterk variëren in bloeddruk. De ratten met de hoogste bloeddruk zullen weer onderling worden gekruist, waarna weer een ronde van selectie plaatsvindt. Uiteindelijk is door de inteelt de genetische variatie tussen de nakomelingen zeer gering geworden. Omdat steeds is geselecteerd zullen de individuen steeds minder variëren in bloeddruk en zijn de genetische factoren die aanleg voor hoge bloeddruk bepalen in deze ratten verzameld. De volgende stap is bepalen welke gebieden in het genetisch materiaal (DNA) dit zijn.

DNA, RNA en eiwitten

Het DNA van de rat is georganiseerd in 21 chromosomen en bestaat uit bijna 2,5 miljard baseparen. Het DNA bestaat uit vier verschillende basen, A, C, G, en T, waarvan A en T een paar vormen en G en C ook. Uiteindelijk zijn er dus twee strengen met A-tjes, C-tjes, G-tjes, en T-tjes, die precies op elkaar plakken. In 2004 is de gehele basenvolgorde van het ratten DNA bekend geworden. Sommige van deze baseparen in het DNA vormen een code om een eiwit te maken. Reeksen van baseparen die zo'n code vormen, worden genen genoemd. Door het aflezen van een gen wordt eerst messenger RNA gemaakt en door het aflezen van dit messenger RNA wordt het corresponderende eiwit gemaakt. De rat heeft tussen de 25.000 en 30.000 van deze genen. Ter vergelijking: het menselijk DNA bestaat uit 23 chromosomen en ongeveer 2,8 miljard baseparen en bevat ook tussen de 25.000 en 30.000 genen. Van vele van deze genen is het bekend welk eiwit ervan wordt gemaakt en wat ongeveer de functie daarvan is in het lichaam, maar van het overgrote deel van de genen is niet precies bekend wat ze doen. Bovendien zitten er in het DNA ook nog reeksen van baseparen die wel een functie hebben, maar niet voor een eiwit coderen (functionele, niet-coderende gebieden). Als laatste zitten er in het DNA stukken, die geen

enkele functie hebben ('junk-DNA'). Het functionele genoomonderzoek richt zich erop de functie van genen en andere gebieden in het DNA te achterhalen.

Polymorfismen

Terug naar de ratten. Omdat het DNA van de rat in sterke mate lijkt op het menselijk DNA, wordt verwacht, dat als we de functie van gedeelten van het ratten DNA te weten kunnen komen, we ook iets leren over vergelijkbare gebieden in het menselijk DNA. Net als mensen onderling, verschillen rattenstammen ook op bepaalde posities in hun DNA van elkaar. De basen op zulke posities worden polymorfismen genoemd. Het is dus heel belangrijk om de polymorfismen in kaart te brengen om zo iets meer over de genetische basis van die aandoening waarvoor een bepaalde rattenstam is geselecteerd, te weten te komen. Mogelijk spelen soortgelijke polymorfismen ook een rol bij dezelfde aandoening in mensen.
Polymorfismen komen voor in genen, in functionele, niet-coderende gebieden, en in junk-DNA. Het is op dit moment technisch en financieel niet mogelijk om voor alle rattenstammen de gehele DNA sequentie te bepalen en zo in één keer alle polymorfismen op te sporen. We verwachten, dat dit binnen 5 tot 10 jaar wel mogelijk is. Om toch deze polymorfismen op te sporen, hebben we gebruik gemaakt van twee verschillende technieken.

Polymorfismen opsporen met bleekselderij?

De eerste techniek is gebaseerd op een klievingsreactie met een enzym (CEL I) dat we uit bleekselderij hebben gehaald. We hebben hiervoor stukken uit het ratten DNA specifiek vermenigvuldigd in 96 verschillende rattenstammen en deze vermenigvuldigingsprodukten allemaal afzonderlijk één-op-één gemengd met het vermenigvuldiginsprodukt van hetzelfde stukje DNA van de referentiestam (de stam Brown Norway, die ook gebruikt is om de hele basenvolgorde van het ratten DNA te bepalen). Als er nu van één van de stammen ten opzichte van de referentiestam een verschil in dat bepaalde stukje DNA zit, en we halen alle baseparen uit elkaar en plakken ze weer aan elkaar, dan zullen dus strengen DNA op elkaar plakken, die op de positie van het polymorfisme niet precies passen. Er ontstaat een bobbeltje in het aan elkaar geplakte produkt. Deze bobbeltjes kunnen heel specifiek worden doorgeknipt door het CEL-I enzym en gescheiden worden van de produkten zonder bobbeltjes. Op deze manier konden we voor 96 veel gebruikte laboratorium stammen, die een model zijn voor bepaalde aandoeningen, polymorfismen bepalen ten opzichte van de referentiestam Brown Norway. Dit is beschreven in hoofdstuk 4 en we hebben ons hierbij gericht op polymorfismen in genen. In hoofdstuk 6 is dezelfde techniek gebruikt voor het bepalen van polymorfismen in 96 rattenstammen in een promotor gebied (het gebied voorafgaande aan een gen) van een belangrijk gen, de dopamine receptor D3, dat bij mensen is betrokken bij aandoeningen als schizofrenie, ziekte van Parkinson en verslaving.

De 'wild side' van de laboratoriumrat

Met een tweede techniek, die bestaat uit de directe bepaling van de basenvolgorde van vermenigvuldigingsprodukten ('direct sequencing'), hebben we eerst willekeurige stukken in het DNA van een wilde rat gelezen en daarin de polymorfismen ten opzichte van de referentiestam Brown Norway bepaald. We zijn toen specifiek deze polymorfismen gaan bekijken in 36 laboratorium stammen en kwamen tot de conclusie dat 85% van de polymorfismen tussen een willekeurige wilde rat, die door de Ongediertebestrijding van de Gemeente Utrecht was gevangen

in één van de grachten, ook voorkomt in laboratoriumstammen! Hierdoor kwamen we op het idee om met behulp van wilde ratten DNA fragmenten heel efficient naar polymorfismen te gaan zoeken in laboratoriumstammen, zonder dat we daarvoor de gehele basenvolgorde van elk van de stammen hoeven te bepalen. Bovendien krijgen we hierdoor een idee over hoe sterk rattenstammen in genetisch opzicht van elkaar verschillen en hoeveel ze genetisch gezien nog lijken op wilde ratten. In hoofdstuk 5 is deze studie beschreven.

Zelf polymorfismen in ratten aanbrengen om 'knockouts' te maken

In hoofdstukken 4, 5 en 6 hebben we dus natuurlijk voorkomende polymorfismen (genetische variatie) in verschillende gebieden in het DNA van laboratoriumstammen bepaald, om een idee te krijgen van wat moeder natuur allemaal in de loop der jaren heeft veranderd in het ratten DNA. Omdat we van bepaalde stammen weten, dat ze een menselijke aandoening nabootsen, bijvoorbeeld hoge bloeddruk, kunnen we bepaalde polymorfismen die vaak voorkomen in die stam, gaan linken met deze aandoening. In werkelijkheid ligt het natuurlijk veel complexer. Het is meestal zo, dat niet één polymorfisme verantwoordelijk is voor hoge bloeddruk, maar een heleboel polymorfismen allemaal een klein beetje bijdragen aan het ziektebeeld. Dergelijke ziekten, zoals hoge bloeddruk en diabetes, worden dan ook 'complexe ziekten' genoemd.
Soms bestaat er aanwijzing dat een bepaald gen te maken heeft met deze ziekte. Een veel gebruikte benadering om de rol van dit gen in een organisme te bepalen (functionele genoomanalyse), is om het gen uit te schakelen door er gericht een mutatie in aan te brengen (knockout). Deze benadering wordt 'reverse genetics' genoemd. Je muteert een gen en gaat dan bekijken wat voor effect deze uitschakeling heeft op het organisme.
In ratten was het niet mogelijk om gericht genen uit te schakelen ('knockouts' te maken). In hoofdstukken 2 en 3 hebben we beschreven hoe we dit toch in de rat hebben gerealiseerd. Onze benadering is gebaseerd op het moedwillig aanbrengen van willekeurige genetische variatie. Dat kan worden gedaan door geslachtsrijpe mannelijke ratten te behandelen met een chemisch stofje ENU (*N*-ethyl-*N*-nitrosourea). ENU zorgt ervoor dat willekeurige veranderingen (mutaties) in het DNA van de spermacellen ontstaan. Door deze behandelde mannen te laten paren met onbehandelde vrouwen, zullen deze willekeurige mutaties in het DNA van de nakomelingen terecht komen. Het enige wat wij moeten doen, is van alle nakomelingen een DNA-monstertje afnemen en het DNA screenen, m.b.v. hierboven beschreven technieken, op geïnduceerde genetische veranderingen in het gen, dat mogelijk betrokken is bij een bepaalde aandoening. Wij hebben vastgesteld, dat 1 op de 20 mutaties in een gen de functie van dat gen compleet verstoord, waardoor gesproken kan worden van een 'knockout'. Op deze manier hebben wij zes van de in totaal acht 'knockouts' in de rat in wereld gemaakt.
Omdat alleen de mannelijke helft van het totale DNA in de nakomelingen genetisch is veranderd, en de nakomelingen twee kopieën hebben van elk gen (de andere kopie komt van de onbehandelde moeder), komen de mutaties in de directe nakomelingen alleen voor in heterozygote vorm (slechts één van de twee kopieën is geraakt). Door zo'n heterozygoot uit te kruisen met een onbehandeld vrouwtje en hun nakomelingen vervolgens weer met elkaar te laten paren, kunnen uiteindelijk dieren verkregen worden, die beide kopieën van het geraakte gen bezitten en dus geen normaal functionerend kopie meer hebben. Dit zijn de echte 'knockouts'.

Syndroom van Usher

Om het allemaal nog iets ingewikkelder te maken, hebben we in hoofdstuk 7 beschreven hoe we een genetische aandoening, geïnduceerd door de behandelingen beschreven in hoofdstukken 2 en 3, hebben opgehelderd met behulp van natuurlijk voorkomende polymorfismen, zoals

geïdentificeerd in hoofdstukken 4, 5, en 6. Door de behandelingen met ENU zijn dus willekeurige mutaties geïnduceerd in de nakomelingen. In sommige van deze nakomelingen hebben we een interessante mutatie gevonden door middel van de twee screeningsmethoden. Door met deze nakomlingen te gaan fokken met onbehandelde partners en hun nakomelingen weer op elkaar te kruisen, kan de betreffende mutatie homozygoot gemaakt worden. Sommige van deze dieren hebben nu dus twee kopieën van het veranderde gen en geen enkele normale kopie meer. Doordat een rat met een interessante mutatie ook nog vele andere willekeurige mutaties bevatten, kunnen ook andere, niet bekende mutaties homozygoot gemaakt worden. Dit is gebeurd bij één van de interessante mutanten. We vonden in de nakomelingen een aantal dieren, dat in rondjes liep, daardoor 'Tornado' genoemd. We hebben bekeken of deze dieren de interessante mutatie bevatten waarvoor we ze homozygoot wilden maken, maar dit bleek niet het geval. Er moest dus bij toeval een andere mutatie homozygoot gemaakt zijn, die dit gedrag veroorzaakt.
Door 'Tornado' dieren te laten paren met vrouwtjes van de referentiestam Brown Norway, en hun nakomelingen weer onderling te kruisen, kunnen nakomelingen worden verkregen, die weer 'Tornado' zijn en overal in hun DNA Brown Norway/'Tornado' polymorfismen hebben, behalve op het chromosoom rond de 'Tornado'-mutatie. We kunnen nu dus in deze derde generatie 'Tornado' dieren bepalen in welke stukken het DNA afkomstig is van Brown Norway en waar alleen maar 'Tornado'-DNA voorkomt. Het gebied met voornamelijk 'Tornado'-DNA zal dus dicht bij de mutatie in de buurt liggen, omdat de mutatie uit slechts 'Tornado'-DNA bestaat. Dit hebben we bepaald voor 67 derde generatie 'Tornado'-dieren en we kwamen uit op een gebied in het midden van het eerste chromosoom van de rat.
In de tussentijd hebben we de 'Tornado'-dieren verder bestudeerd en gevonden, dat naast het 'rondjes lopen', de dieren ook doof zijn, evenwichtsstoornissen hebben en hyperactief zijn. In dat gebied op chromosoom 1 liggen een aantal genen, waarvan er één, genaamd *myosin7a*, betrokken is bij de vorming van het binnenoor in mensen. We hebben in de 'Tornado'-mutanten in dit gen een mutatie gevonden, die de functie van dat gen compleet plat legt. Mensen met mutaties in dit gen lijden aan het Syndroom van Usher (type 1B), een ernstige aangeboren afwijking, die lijdt tot aangeboren doofheid, evenwichtsstoornissen, en progressieve blindheid in de vroege puberteit. We hebben nu een uniek diermodel in handen om deze menselijke ziekte verder te bestuderen. Bovendien hebben we aangetoond, dat polymorfismen in de rat te gebruiken zijn om geïnduceerde genetische aandoeningen te identificeren.

Algemene conclusie

Met de in dit proefschrift beschreven studies hebben we bijgedragen aan het in kaart brengen van genetische variatie tussen rattenstammen en hebben we technieken ontwikkeld om door middel van de 'reverse genetic'-benadering genen te veranderen en uit te schakelen in de rat. Bovendien hebben we natuurlijk voorkomende polymorfismen gebruikt voor een 'forward genetic'-benadering in de rat. In de introduktie staat beschreven, wat het belang van genetische benaderingen is voor het onderzoek naar complexe ziekten in de rat. Ten eerste weten we nu iets meer van de natuurlijk voorkomende genetische variatie in veel gebruikte laboratoriumstammen, maar om uiteindelijk de genetische oorzaak van een complexe ziekte te begrijpen, moeten we nog veel meer, zoniet alle genetische variatie tussen laboratoriumstammen in kaart brengen. Bovendien moeten we met verschillende genetische experimenten, waarvan 'forward' en 'reverse genetics' er twee zijn, bewijzen dat een gen of ander functioneel gebied in het DNA invloed heeft op het ontstaan van het ziektebeeld. Er is dus nog veel meer onderzoek nodig op het gebied van rattengenetica.

Dankwoord

Tot slot nog even een paar alinea's voor de personen die tijdens de afgelopen vier jaar op één of andere manier hebben bijgedragen aan dit proefschrift. In mijn eentje was ik echt niet zo ver gekomen, dus voordat ik iemand vergeet: alvast bedankt iedereen!
Het spits wordt afgebeten door Edwin. Cuppen, ik heb in een dronken bui in de kroeg herhaaldelijk mijn dank betuigd en dat doe ik nu nog even in (redelijk) nuchtere toestand: Bedankt voor al je tips, correcties op papers, discussies, dagelijkse slappe praat, en gezelligheid tijdens drinkactiviteiten! Ook al valt het niet mee om de internationale muizenmuur te slechten, toch blijven we allebei nog even in het 'rattenveld' actief, dus waarschijnlijk treffen wij elkaar nog vaak genoeg tijdens congresborrels en afterparties...
En natuurlijk Ronald, bedankt voor je support bij onze projecten en opmerkingen op mijn papers en hoofdstukken.

Rondom mij heen in het lab hebben ook wat mensen een zeer essentiële bijdrage geleverd. Allereerst de Josjes (Josine en José) en Mark: ik kon altijd bij jullie terecht met alle 'high-throughput'-experimenten en dat leverde dan altijd zeer spoedig de gewenste resultaten op, wat natuurlijk ook blijkt uit een aantal publicaties met jullie namen erboven. Heel veel dank voor jullie hulp, inzet en lol in het lab en tijdens borrels!
Als tweede wil ik Esther bedanken, voor alle ingewikkelde pipeteerklusjes, die je voor mij heel secuur hebt uitgevoerd en tot op de dag van vandaag nog steeds uitvoert!
Ten derde dan de andere leden en oud-leden van het Cuppen-lab. Harma, Judith, Marieke en Geert, bedankt voor de samenwerkingen bij diverse projecten en 'small talks' in het lab. Eelke, Bastiaan, en Dimitri, mannen van het eerste uur: ook bedankt!

Wat mensen ook over studenten zeggen, ik hou er een zeer positief gevoel aan over. Beide studenten die ik heb begeleid voor hun stage, hebben ook een publicatie opgeleverd en daarmee ook een hoofdstuk in dit proefschrift.
Ik wil Joram en Dimphy heel hartelijk danken voor hun inzet en geduld. Dimphy succes in Singapore en Van Mullenstein: NIETSCH!
Langzaam kom ik dan bij twee andere opvallende leden van het Cuppen-lab: Victor and Zhenya, locally known as 'The Nerds'. Thanks a lot for discussions, collaborations, talks during borrels and for being my roommate! I specially want to thank Victor for reformatting and processing a lot of my (sometimes crappy) data. It was worth it, referring to four papers with your essential contribution. You'll finally get rid off me!
Lang geleden, in vervlogen tijden was er een persoon, die het heeft aangedurfd om met Cuppen en mij in één kamer te zitten: Romke! Rompje heel hartelijk dank voor jou technische ondersteuning, droge humor, wilde rat en veilige terugritten vanuit Winterberg.

Rats, worms, mice and fish, there is always a suitable animal to work on for the Hubrecht researchers population. I would like to thank everybody of the Hubrecht Lab for the pleasant time I had in the lab and during numerous borrels.
Uiteraard zijn er mensen die ik speciaal wil bedanken. Allereerst zijn dat de leden en oud-leden van het Plasterk-lab. Ik ga jullie niet allemaal opnoemen, want dan ga ik zeker iemand vergeten. René en Marcel, bedankt voor de squashlessen. Tijsterman, gelukkig heb ik wel een zeer bescheiden aantal potjes van je kunnen winnen! Maar ook ernstig vaak verloren... Oh ja, Robert van Olympos wil ik bedanken voor de soepele huurtarieven en voor het afnemen van 30 ton speelgoed...

Leden van de AIO-commissie 2002-2003: Bas, Joris, Berber, Ellen, Armando, en Marijn. Ik wou maar even zeggen: 'We hebben het top gedaan...' En advies aan alle volgende AIO-commissies: 'Kies nooit een congreszaal naast een razende rollercoaster!'
En zoals beloofd in een eigen alinea: Ira! Bedankt voor al je hulp, smakelijke reacties en dwingende opmerkingen. Erg moedig dat je elke ochtend nog lachend langs onze kamer durft te lopen!

Toch ook nog even een paar woordjes voor de ondersteunende mensen in het Hubrecht Lab: civiele dienst, automatisering, financiële administratie, personeelszaken, dierenverzorgers, en natuurlijk de schoonmakers: bedankt!
'**H**et was weer eens zover! ...' Een zeer gevleugelde uitspraak als we weer eens uitgebreid geborreld hadden (waarbij het overigens nooit uit de hand is gelopen...) Voorlopig kunnen jullie opgelucht ademhalen: ik zal de komende tijd geen borrel meer geven. Maar jullie kunnen natuurlijk wel op mijn promotiefeestje wraak nemen!
En als we het toch over borrelen hebben: dat doe ik natuurlijk het liefst alleen, thuis. Maar ik wil toch graag de vrijdagmiddag, -avond, en -nacht borrelclub hartelijk danken voor alle steun tijdens drinkactiviteiten. Daar komt-ie dan: Porno, Flopsch, Ouwe, Cobus, Adonis, Tommie, Ieniemienie, Herder, Schapeneuker, Cup E, Henkie, Don Karel, Shoharma, Half, Nietsch, en iedereen die geen bijnaam heeft en toch borrelt: bedankt!

Wordt het nu tijd voor de mensen die niks met het lab te maken hebben. Allereerst wil ik mijn voormalige zaalvoetbalteam bedanken voor de plezierige veldslagen die we samen hebben ondernomen. Totdat het noodlot toesloeg en ik niet meer mee kon doen. Maar jullie hebben mij niet gemist, toch?
Ook al wordt het steeds moeilijker om iedereen weer bij elkaar te krijgen, we hebben het aardig lang volgehouden om elke maand bij iemand te eten! Roel, Maayke, Linda, Lise en Osje, bedankt voor alle etentjes, weekendjes, feestjes en Beerenburg. Culemburg en Leiden is nog wel te overzien, maar Madison... ik denk dat ik voorlopig moet afhaken, maar ik kom terug!
Roosendaal, waar ligt dat ook al weer? Ik moet er af en toe heen om die vraag nog te kunnen blijven beantwoorden. Gelukkig staan daar steevast twee gasten mij op te wachten met een goudgele unit in de knuist geklemd: Rens en Geert! We kunnen die traditie natuurlijk voortzetten in de VS en misschien hebben ze daar ook wel een Moeder Overste en een Sjoes!
Langzaam kom ik dan bij de meest belangrijke mensen uit. Allereerst mijn ouders: pa en ma, heel veel dank voor jullie steun bij alles wat ik doe. Misschien begrijpen jullie het niet allemaal, maar zolang ik er nog lol in heb, vinden jullie het goed, toch? En natuurlijk ook nog de Moulaaaard. Ik wil je wel graag bedanken, maar ik weet niet zo goed waarvoor... je blijft toch gewoon mijn broer! Kom maar naar Madison, dan gaan we daar fietsen en zuipen.
Dan als allerlaatste, Marijntje! Ik ben nog altijd zo gelukkig bij jou en ik ben heel blij, dat jij en Lotje mee willen naar de VS. We gaan het gewoon doen! Ik hou van jou...

Bart

Curriculum Vitae

Bartolomeus Mathijs Godefridus Smits is geboren op 28 juni 1978 in Roosendaal. In 1996 behaalde hij het VWO diploma op het St. Gertrudislyceum te Roosendaal, en ging hij Biologie studeren aan de Universiteit Utrecht. In het kader van de doctoraalfase van zijn studie deed hij een onderzoeksstage bij de vakgroep Moleculaire Microbiologie (Faculteit Biologie), Universiteit Utrecht, onder begeleiding van Dr. Sally Hoffer en Prof. Dr. Jan Tommassen. Een tweede onderzoeksstage werd uitgevoerd bij de Functional Genomics Groep van het Hubrecht Laboratorium, onder begeleiding van Dr. Celine Moorman en Prof. Dr. Ronald Plasterk. In december 2001 behaalde hij het doctoraal eindexamen. In november 2001 is hij gestart als Onderzoeker in Opleiding in de nieuw opgerichte onderzoeksgroep van Dr. Edwin Cuppen in het Hubrecht Laboratorium, waarvan de in dit proefschrift beschreven studies het resultaat zijn. Begin 2006 zal hij starten als Postdoc in het laboratorium van Prof. Dr. M.N. Gould van de afdeling Oncologie, Universiteit van Wisconsin in Madison, Verenigde Staten.

Bartholomeus Mathijs Godefridus Smits was born on June 28th 1978 in Roosendaal. He attended the St. Gertrudislyceum in Roosendaal, graduating in 1996. He studied Biology at Utrecht University. As an undergraduate, he performed a research project with the department of Molecular Microbiology (Faculty of Biology), Utrecht University, under supervision of Dr. Sally Hoffer and Prof. Dr. Jan Tommassen. A second research project was performed at Functional Genomics Group of the Hubrecht Laboratory, under supervision of Dr. Celine Moorman and Prof. Dr. Ronald Plasterk. In December 2001 he graduated. From November 2001, he started his PhD-project at the Hubrecht Laboratory in the newly formed research group of Dr. Edwin Cuppen, from which the work described in this thesis is the result. In the beginning of 2006, he will start as a postdoctoral researcher in the laboratory of Prof. Dr. M.N. Gould of the department Oncology, University of Wisconsin, Madison, USA.

List of Publications

Smits, B.M.G., Mudde, J.B., Plasterk, R.H.A., and Cuppen, E. (2004) Target-selected Mutagenesis of the Rat. *Genomics* **83:** 332-334

Smits, B.M.G., D'Souza, U.M., Berezikov, E., Cuppen, E., and Sluyter, F. (2004) Identifying Polymorphisms in the *Rattus norvegicus* D_3 Dopamine Receptor Gene and Regulatory Region. *Genes, Brain and Behavior* **3:** 138-148

Smits, B.M.G., Van Zutphen, L.F.M., Plasterk, R.H.A., and Cuppen, E. (2004) Genetic Variation in Coding Regions between and within Commonly Used Inbred Rat Strains. *Genome Research* **14:** 1285-1290

Smits, B.M.G., Peters, T.A., Mul, J.D., Croes, H.J., Fransen, J.A.M., Beynon, A.J., Guryev, V., Plasterk, R.H.A., and Cuppen, E. (2005) Identification of a Rat Model for Usher Syndrome type 1B by ENU Mutagenesis-driven Forward Genetics. *Genetics* **170(4):** 1887-96

Smits, B.M.G., Mudde, J.B., Van de Belt, J., Verheul, M., Olivier, J., Homberg, J., Guryev, V., Cools, A.R., Ellenbroek, B.A., Plasterk, R.H.A., and Cuppen, E. (2005) Generation of Gene Knockouts and Mutant Models in the Laboratory Rat by ENU-driven Target-selected Mutagenesis. *Pharmacogenetics and Genomics, in press*

Smits, B.M.G., Guryev, V., Zeegers, D., Wedekind, D., Hedrich, H.J., and Cuppen, E. (2005) Efficient Single Nucleotide Polymorphism Discovery in Laboratory Rat Strains using Wild Rat-derived SNP Candidates. *BMC Genomics, in press*

Smits, B.M.G. and Cuppen, E. (2005) Rat Genetics: The Next Episode. Invited Review *Trends in Genetics, submitted*