Acta Biochimica Polonica

Vol. 49 No. 3/2002

553–569

QUARTERLY

Review

Mice and humans: chromosome engineering and its application to functional genomics✪

Jan Klysik✉

Baylor College of Medicine, Department of Molecular and Human Genetics, One Baylor Plaza, Houston, TX 77030

Received: 19 June, 2002; accepted: 12 August, 2002

Key words: transgenesis, embryonic stem cell technology, gene traps, chromosome engineering, ENU mutagenesis

Functional modeling of human genes and diseases requires suitable mammalian model organisms. For its genetic malleability, the mouse is likely to continue to play a major role in defining basic genetic traits and complex pathological disorders. Recently, gene targeting techniques have been extended towards developing new engineering strategies for generating extensive lesions and rearrangements in mouse chromosomes. While these advances create new opportunities to address similar aberrations observed in human diseases, they also open new ways of scaling-up mutagensis projects that try to catalogue and annotate cellular functions of mammalian genes.

The biggest biological project in human history, The Human Genome Sequencing Project (HGSP), has come to a conclusion in the private sector (Venter *et al.*, 2001) and is expected to be finished in the public domain in the very near future (Lander *et al.*, 2001; Pennisi, 2000). Since the majority of genes are identified, and sequence information is readily available to the scientific community at large, the HGSP will affect all aspects of biological sciences including genetics and medicine. The number of genes in the human genome (40 000) is lower then previously estimated creating an exceptional opportunity to

✪This work was supported by the U.S. Public Health Service Grant P01 CA75719.

✉Address correspondence to Baylor College of Medicine, Department of Molecular and Human Genetics, One Baylor Plaza, Houston, Texas 77030, U.S.A.; tel. (713) 798 8992, fax: (713) 798 8142, e-mail: jank@bcm.tmc.edu

Abbreviations: ENU, *N*-ethyl-*N*-nitrosourea; ES, embryonic stem; GFP, green fluorescent protein; HGSP, Human Genome Sequencing Project; MGSP, Mouse Genome Sequencing Project

address genome function in a global sense. In the upcoming decades, the main research effort will inevitably shift from the description of genomic structure of DNA towards functional analyses of its genetic content. Defining new pathways, or rapidly extending those that already exist, emerges as an urgent task.

The genetic revolution began some time ago bringing an advent of transgenesis, gene targeting, embryonic stem cell technology, and chemical mutagenesis (Holschneider & Shih, 2000; Ihle, 2000; Cecconi & Meyer, 2000). The power of these technologies is well recognized as they have already produced many examples of comprehensive functional descriptions of individual loci. However, the majority of genes supplied by the HGSP awaits functional evaluation. Such analyses will require in-depth and in-breadth approaches suited for simultaneous examination of a large number of genes within a reasonable period of time (Brown & Nolan, 1998; Brennan & Skarnes 1999; Nolan *et al.*, 2000; Klysik, 2001). Progress in selected aspects of systematic genome-wide strategies, as well as large-scale mutagenesis approaches capable of uncovering multiple alleles, is presented in this review.

MOUSE EMERGES AS AN ORGANISM OF CHOICE IN FUNCTIONAL STUDIES OF MAMMALIAN GENES

Insight into the cellular function of an unknown gene can be obtained through a number of routes. Comparative sequence analysis is one of them. Useful information can be extracted based on similarities within the sequence motifs of genes, or amino-acid motifs of their protein products. Development of computational tools and their application to genetic questions will undoubtedly contribute to the overall understanding of the complexity of the human genome (Tsoka & Ouzounis, 2000; Kaminski, 2000; Sreekumar *et al.*, 2001). These tools can be applied across a broad variety of species as the repertoire of sequenced genomes grows over time. However, conclusive functional information will remain to be derived from experimental approaches applied to the vast majority of individual genomic loci.

There is no doubt that lower organisms such as *Drosophila* sp. and *Caenorhabditis elegans* will provide key initial information about many genetic pathways. However, it is unrealistic to expect that the functions of the 40 000 human genes can be condensed into a few hundred biological processes implied for these models (Miklos & Rubin, 1996). A large number of mammalian genes do not have orthologues in invertebrates. Tissue specific genes, genes related to asthma, obesity, osteoporosis, and other diseases specific to vertebrates, cannot be addressed using lower organisms. For these reasons, and because mouse genetics offers many technological advances that are not available in other mammalian organisms, many have argued that the mouse will play a key role in modeling studies of human genes and human diseases (Bottinger & Kopp, 1998; Moore, 1999; Clarke, 2000; Brayton *et al.*, 2001).

Mice and humans diverged from a common ancestor about 65 million years ago, yet most salient aspects of mammalian physiology have not diverged significantly in these lineages during this time. Both organisms have the same organ systems, similar reproductive cycles, skeletons. Physiology, biochemistry, and pathology are all very similar as well. The anatomical and physiological parallels between the two species are reflected in the comparable numbers of genes, although the size of the mouse genome is 15% smaller. Similarities are readily apparent at the level of chromosomal organization of genes. Large parts of human and mouse chromosomes are conserved. These regions, known as conserved linkage groups, can cover significant distances. For example, over 40 cM of the distal part of mouse chromosome 11 is virtually identical to human chromosome 17 (Fig. 1).

Not only the same genes can be found in these syntenic regions, but also their arrangement and relative orientation remain conserved in most cases.

Perhaps the most advantageous feature of this small mammalian model organism is its genetic tractability. Genome-wide mutations induced by chemical mutagenesis can be tracked and analyzed in parallel to corresponding phenotypes (Nolan *et al.*, 2000; Rathkolb *et al.*, 2000). Moreover, genes can be over-expressed (transgenesis) (Si-Hoe *et al.*, 2001) either or knocked out (gene targeting

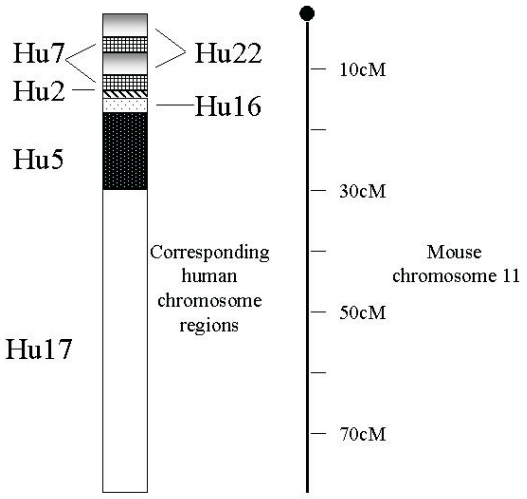

Figure 1. Synteny between mouse chromosome 11 and human chromosome 17.

and embryonic stem cell technology) (Baribault & Kemler, 1989; Pirity *et al.*, 1998) to allow for functional studies that are not possible in most other organisms.

TRANSGENESIS AND EMBRYONIC STEM CELL TECHNOLOGY

Introducing genes into the germ line by microinjecting cloned DNA into the pronuclei of fertilized mouse eggs was described many years ago by a number of investigators (Palmiter & Brinster, 1986; Hanahan, 1989; Kollias & Grosveld, 1992; Beddington, 1992). These eggs are transferred into the oviduct of pseudopregnant foster mothers for the duration of gestation. The injected DNA integrates randomly, and usually in multiple copies, causing a fraction of the mice born to be transgenic. In most of these founder mice, the tandem transgene can be detected in both somatic cells and in germ cells. Thus, the Mendelian passage of the gene to progeny can be obtained. This technology has remained almost unchanged since its original application, and it has been used extensively to assess phenotypic changes associated with over-expression of genes (Woychik & Alagramam,1998).

At the beginning of 1980's, yet another modern technique was born: embryonic stem cell technology (Fig. 2). As early as 1974, Brinster showed that cells derived from embryonal carcinoma, when injected into the blastocyst, could contribute to the development of somatic tissues of chimaeric mice (Brinster, 1974). This important observation was eventually confirmed by other groups (Papaioannou *et al.*, 1975; Mintz & Cronmiller, 1978), yet the use of these carcinoma cells for the germ line transmission of new genetic functions has never materialized, in part due to difficulties encountered at the experimental reproducibility level. Several years later, embryonic stem (ES) cells were derived from the inner mass of the blastocyst (Evans & Kaufman,1981; Martin, 1981) to become the key element of the system that allows the site-specific genetic alterations to be efficiently transmitted into the germ line. The wide spread and unrestricted availability of ES cells revolutionized mouse genetics over the past 10 years. This is an extraordinarily powerful technology. In combination with site specific gene targeting and chemical mutagenesis, embryonic stem cells became a part of many elegant functional strategies capable of addressing every imaginable locus within the mouse genome (Baribault & Kemler, 1989; Pirity *et al.*, 1998; Woychik &

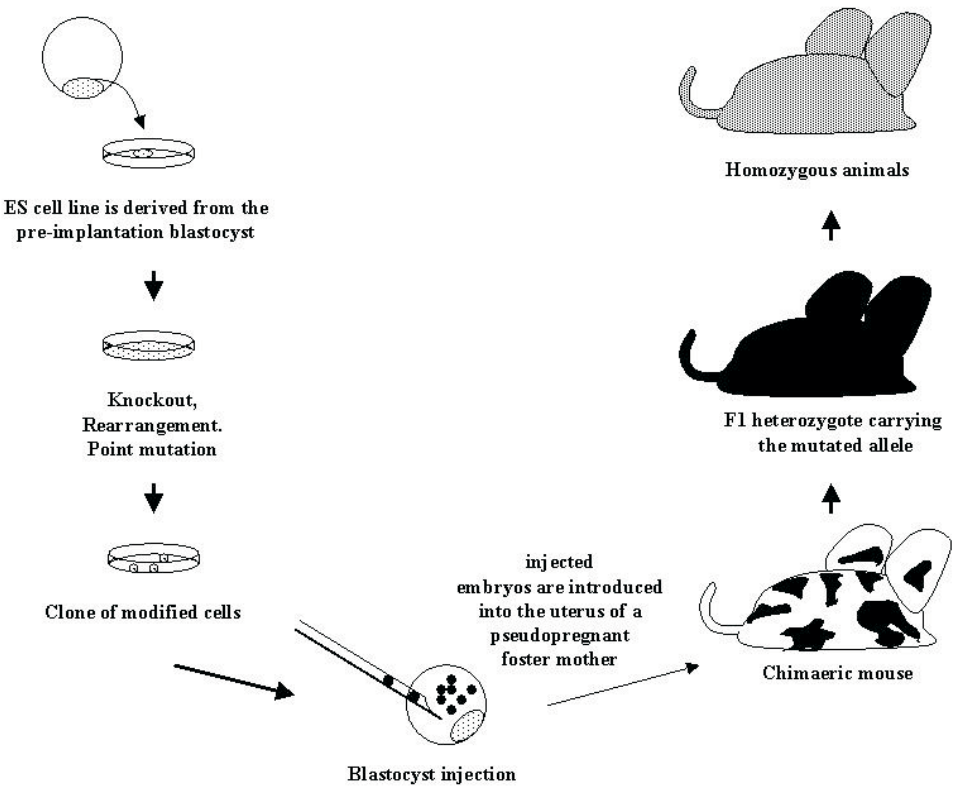

Figure 2. Embryonic stem cell technology.

A diagram illustrating how ES cells can be used to establish a targeted allele in the germ line. The isolated ES cells are pluripotent and have the ability to contribute to several tissues of the fetus including extraembryonic membranes. Undifferentiated proliferation of ES cells in culture is achieved by using feeder cells (Evans & Kaufman, 1981; Wobus *et al.*, 1984; Robertson, 1987) that secrete the differentiation inhibiting factor (leukemia inhibitory factor, LIF) (Smith *et al.*, 1988). ES cells can be targeted at defined loci *via* homologous recombination. Alterations may be as subtle as single nucleotide changes, or more extensive such as gene knock-outs or robust chromosomal rearrangements. The genetically modified cells are then injected into the blastocyst. Injected blastocysts are surgically introduced into the uterus of the pseudopregnant foster mother. The chimaeras obtained are able to transmit the modified allele to F1 progeny. The F1 intercross leads to F2 homozygous animals having modified both alleles of the targeted gene.

Alagramam, 1998; Brennan & Skarnes, 1999; Clarke, 2000; Cecconi & Meyer, 2000).

GENOTYPE-DRIVEN APPROACHES TO FUNCTIONAL GENOMICS

In recognition of the importance of the mouse model in the studies of human genes, the Mouse Genome Sequencing Project (MGSP) was initiated in 1999. About 95% of the mouse genome is now sequenced, and this sequence information is freely accessible through public databases (Rogers & Bradley, 2001). The MGSP generates an invaluable resource for systematic comparative gene ablation experiments. These have already provided a wealth of knowledge with regard to many biological processes including development (Capecchi, 1989; Friedrich & Soriano, 1991; Brandon *et al.*, 1995a; Brandon *et al.*, 1995b; Brandon *et al.*, 1995c), carcinogenesis (Palapattu *et al.*, 1998; McClatchey & Jacks, 1998; Pitot *et al.*, 2000), metabolism (Postic *et al.*, 1999; El-Sohemy & Archer, 1999), and neurodegenerative pathways (Brusa, 1999; Baum *et al.*, 2000). Since many genes play an important role in both embryogenesis and in

adulthood, simple disruption of these loci can lead to an embryonic-lethal phenotype. To facilitate studies in such cases, inducible gene ablation systems (conditional knock-outs) were devised. They permit for guiding an intact allele through developmental stages and to induce the inactivation process in a controlled manner. First tissue-specific, conditional gene inactivation taking advantage of the Cre/*loxP* recombination system was pioneered by K. Rajewsky (Gu *et al.*, 1994). Many other binary systems, such as *TetR* or *Gal4* based systems, are now available (reviewed in Lewandoski, 2001).

It has been argued that the gene-by-gene targeting through homologous recombination in mouse ES cells is not particularly well suited for large-scale functional projects (Brown & Nolan, 1998). Even with the support of the complete genomic sequence information, this technology may not offer a sufficient throughput to become the main route in gene analyses. Instead, other approaches appear to be more promising. One of them is the gene trap technique. Introduction of retroviral DNA into the mouse germ line was reported in 1976 (Jaenisch, 1976). Ten years later, the first gene traps were designed and tested for integration near enhancer sequences (Allen *et al.*, 1988; Kothary *et al.*, 1988). More recently, retroviral traps were deployed to rescue clones in which the integration event took place selectively within actively transcribed genes (Skarnes *et al.*, 1995; Chowdhury *et al.*, 1997), transcriptionally inactive genes (Palapattu *et al.*, 1998), transiently expressed genes (Thorey *et al.*, 1998), or genes that code for cell surface proteins (Skarnes *et al.*, 1995). Since retroviral integration frequently leads to a hypomorphic allele, and since the trapping cassette provides to a unique sequence tag permitting easy isolation and identification of the flanking gene sequence, the gene trap approaches are well suited to large-scale mutagenesis projects. Global analyses of mammalian genes will benefit greatly from these approaches once complete libraries of gene-trapped ES cells are readily available and equipped with efficient screening systems. There are at least three ongoing large-scale projects devoted to this task. One is lead by Lexicon Genetics (Texas, U.S.A.) (Zambrowicz & Friedrich, 1998), another one is arranged by a German consortium (Thorey *et al.*, 1998), while the most recent program was put in place by Skarnes (Brennan & Skarnes, 1999). The growing list of examples of genes successfully identified and analyzed recently using gene traps includes: transcriptional enhancer factor *Tef-1* (Chen *et al.*, 1994), brain malformation causing gene *jmj* (Takeuchi *et al.*, 1995), anti-apoptotic factor *bcl-w* (Ross *et al.*, 1998), neuropilin 2, a receptor for a class 3 semaphorin (Skarnes *et al.*, 1995), *Apaf1*, an apoptotic protease activating factor 1 (Cecconi & Meyer, 2000), Aquarius gene, *Aqr* (Sam *et al.*, 1998), hematopoietic genes *Hzf* and *Hhl* (Hidaka *et al.*, 2000), and PRDC gene (protein related to differential screening-selected gene aberrative in neuroblastoma (DAN) and cerberus) (Minabe-Saegusa *et al.*, 1998).

PHENOTYPE-BASED APPROACHES

Chemical mutagenesis is a potent approach to generate a large mutant mouse resource. Induced phenotypes can be traced parallel to the transmitted mutations and analyzed in genetic crosses. *N*-Ethyl-*N*-nitrosourea (ENU) (Fig. 3) is a chemical compound that causes single point mutations in a wide variety of organisms. The ethyl group of ENU can be trans-

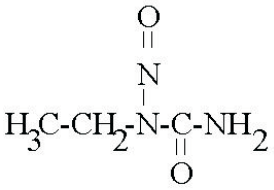

Figure 3. *N*-Ethyl-*N*-nitrosourea.

ferred to oxygen and nitrogen atoms of bases with the most reactive sites identified as N1, N3, N7 of adenine, O6, N3, and N7 of guanine, O2, O4 and N3 of thymine, and O2 and O3 of cytosine (Shibuya & Morimoto, 1993; Noveroske *et al.*, 2000). It also modifies proteins through carbamoylation of amino-acid moieties. Although modifications of cellular components other that DNA add to the toxicity of this chemical, they are believed to have no inheritable consequences.

In mice, the highest mutation rates were observed in pre-meiotic spermatogonial stem cells, with single locus mutation frequencies $1.5–6 \times 10^{-3}$ (Hitotsumachi *et al.*, 1985; Shedlovsky *et al.*, 1993; Hong *et al.*, 2001). This translates into 200–600 gametes that need to be screened in order to obtain a mutation at any single genomic locus of interest. ENU predominantly induces single point changes such as A/T → T/A transversions, (44%), A/T → G/C transitions (38%), A/T → C/G transitions (5%), G/C → A/T transitions (8%), G/C → C/G transversions (3%), and C/G → T/A transversions. It is the most potent mutagen available today to mouse geneticists.

Since ENU causes single base mutations in DNA, it can be used to induce multiple alleles for a single gene. Many genes code for multifunctional proteins, or are represented by a series of protein isoforms. In contrast to gene targeting, gene traps, and other mutagenesis techniques, ENU produces randomly distributed point mutations which offer much greater chance for uncovering phenotypic diversity of a given locus. One example of a multiallelic series is the quaking (*qk*) function. Initially, a single phenotype was known (Sidman *et al.*, 1964) manifested through severe dismyelination of the central nervous system (quaking and seizures). In later experiments, four independent ENU mutants were obtained and found to be homozygous lethal (Shedlovsky *et al.*, 1988; Justice & Bode, 1988; Soewarto *et al.*, 2000), indicating the importance of *qk* in embryogenesis. The list of multiallelic genes grows rapidly (Klysik *et al.*, 2002). Based on the high efficiency of ENU mutagenesis protocols, several large-scale mouse programs have been launched including those at The Mouse Genome Centre, Harwell (U.K.); The National Research Center for Environment and Health (Neuherberg, Germany); The Jaxon Laboratory; University of Pennsylvania (Philadelphia, U.S.A.); Oak Ridge National Laboratory (Oak Ridge, U.S.A.), Tennesse Medical Genome Centre; The Australian National University (Canberra, Australia); Mouse Functional Genomics Research Group, RIKEN Genomic Science Center (Yokohama, Japan); and Baylor College of Medicine (Houston, TX, U.S.A.).

Although the power of ENU to induce dominant and recessive mutations and to generate multiple alleles is an important advantage in large-scale genetic studies, the detection of underlying mutations is somewhat more difficult and requires extensive backcrossing to follow the phenotype in parallel to polymorphic simple sequence length polymorphism (SSLP) markers. Labor associated with genotyping is a major bottleneck in this and all other large-scale genetic approaches. For these reasons, improved strategies that reduce or virtually eliminate the need for genotyping begin to emerg. It is the chromosome engineering combined with embryonic stem cell technology that has the capacity to alleviate many difficulties by providing novel genetic tools and engineered mouse strains.

DEFINED CHROMOSOMAL LESIONS THAT CAN BE USED IN ENU MUTAGENESIS STUDIES

Defined segmental deletions and inversions are particularly well suited for genetic crosses used to uncover recessive phenotypes induced by chemical mutagenesis. The use of deletions was first proposed by Rinchik (Rinchik *et al.*, 1990; Rinchik, 1991) and it takes advantage of the pseudodominat behavior of a recessive

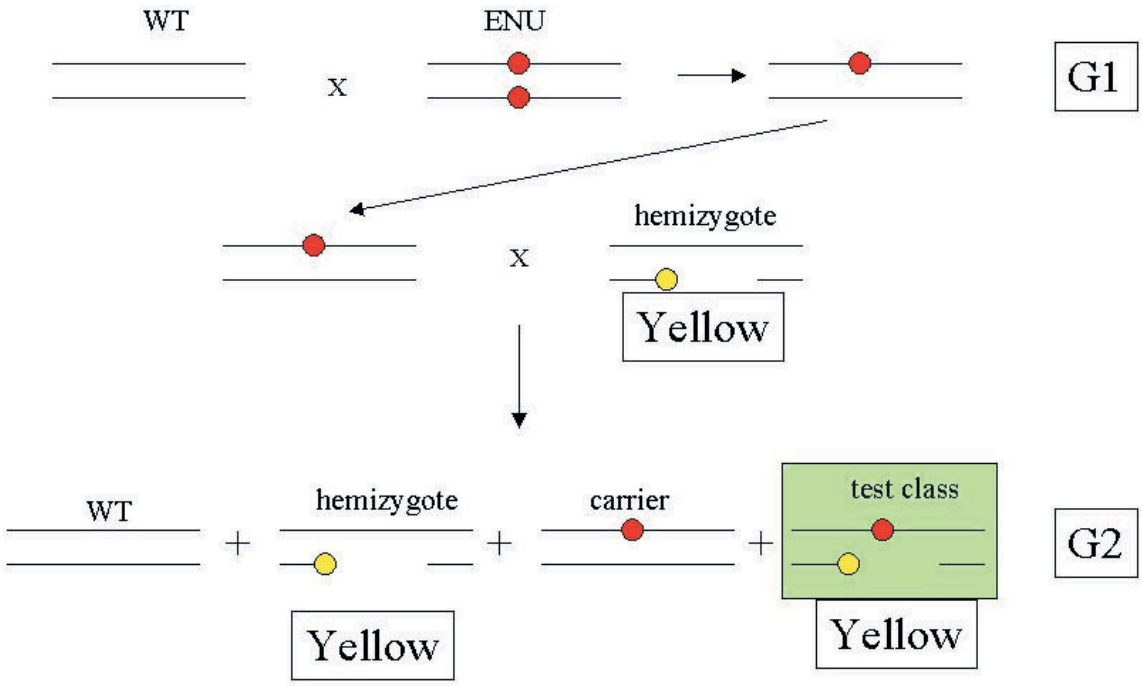

Figure 4. Two-generation cross to uncover recessive phenotypes induced by chemical mutagenesis.

Dominant phenotypes show up in G1 while recessive phenotypes are collected in G2. Since the deletion chromosome is dominantly marked (*K14-agouti*, yellow), these animals can be recognized in the litter with a limited genotyping.

mutation when located on the wild type chromosome opposite the deletion interval. In a typical cross (Fig. 4), an ENU mutagenized male is mated with a wild type female to produce founders in generation 1 (G1). G1 animals are then crossed with deletion carriers to derive the test class of animals that show a phenotype. The underlying mutation is likely to be found within the deletion interval. This regional screen can be carried out effectively with limited genotyping, provided that the deletion endpoints are well defined and visibly marked by a dominant marker gene such as *K14-agouti*.

Before chromosome engineering techniques became available in mice, segmental deletions were produced mainly through radiation or chemical mutagenesis. This animal resource proved to be extremely powerful in scanning large segments of the genome for recessive lesions. Using series of X-ray induced deletions within the 6–11 cM region of chromosome 7 spanning the albino locus (*Tyr*), it was possible to uncover 31 recessive mutations at 10 different loci through genetic crosses involving more than 4 500 pedigrees (Rinchik & Carpenter, 1999). There is little doubt that future development of a genome wide collection of more defined segmental deletions, particularly those that are dominantly tagged, would constitute a powerful genetic tool for functional analyses of genes in mice (Rinchik, 2000).

Chromosomal engineering to produce planned rearrangements in ES cells, including segmental deletions, was pioneered in mid 1990's (Ramirez-Solis *et al.*, 1995). Recently, this technology has been refined and modified to generate a large deletion resource in ES cells and in mice (Lindsay *et al.*, 1999; Zheng *et al.*, 2001; Klysik *et al.*, 2002). Figure 5A outlines basic principles of the method. In the

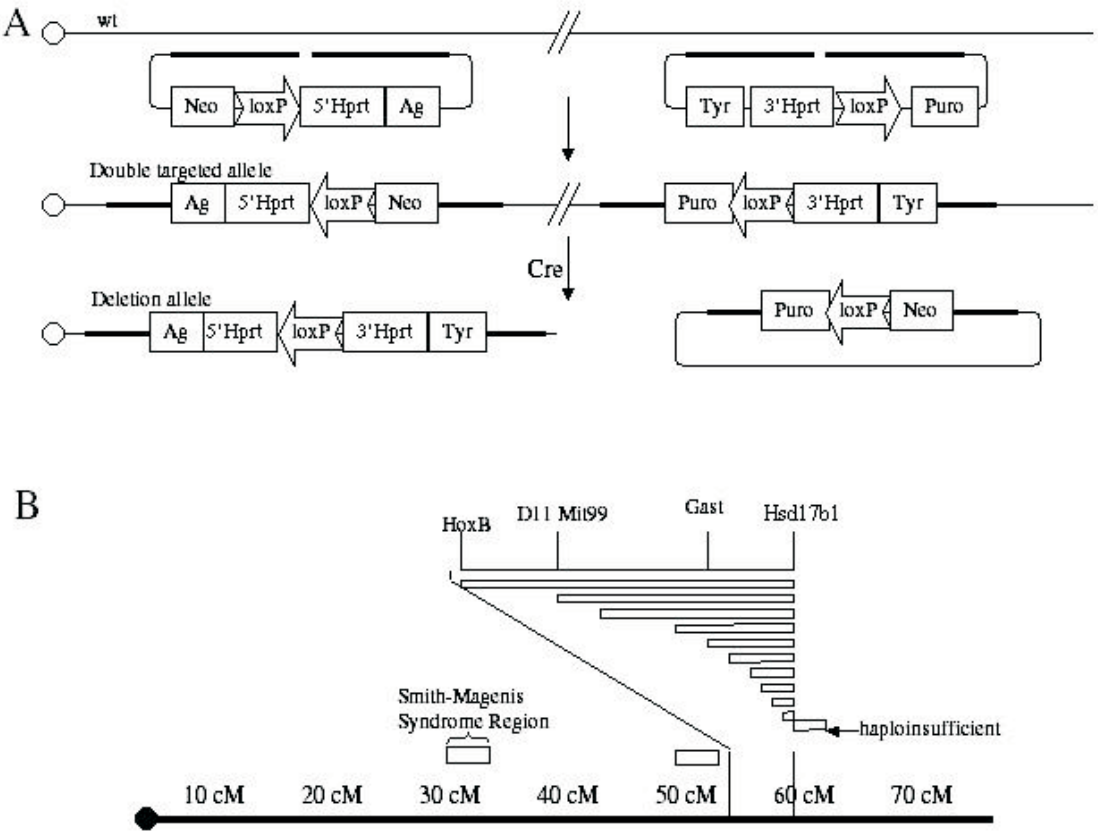

Figure 5. Segmental deletions.

A. A multistep strategy for generating a deletion between chosen loci using the Cre/*loxP* recombination system.

Two independent site-specific targeting events establish functional cassettes at the desired sites. These cassettes carry drug resistance markers (*Neo* and *Puro*), nonfunctional parts of the *Hprt* gene, *loxP* sites, and coat color transgenes (*Tyr* or *K14-agouti*). loxP-dependant recombination is induced in double targeted cells using Cre. Since *loxP* sites are configured as direct repeats along the chromosome, products of Cre/*loxP* recombination are a deletion and an acentric portion of the excised chromosome. The acentric circle is lost in subsequent divisions while ES cells carrying segmental deletion are selectable using HAT (hypoxanthine, aminopterin and thymidine) media due to the reconstitution of a fully functional *Hprt* gene. In adult mice, deletion carriers are recognizable through a visible phenotype (*K14-agouti* or *Tyr*). *K14-agouti* is a particularly useful marker because it is penetrant on most pigmented mouse coat color backgrounds. The *Tyr* minigene confers pigmented coat color only on albino backgrounds. *Tyr* works at some loci but fails to cause coat color change at others.

B. Segmental deletions for chromosome 11 established in the the germ line.

first step, two targeting cassettes are introduced into the pre-selected loci on a chosen chromosome. These cassettes share one common feature, the *loxP* site, which is used at the stage of the deletion induction. Other elements are different and provide nonfunctional parts of the *Hprt* gene (3′ or 5′), selective markers (*Puro* or *Neo*), dominant coat color markers (*K14-agouti* or *Tyr*) and genomic homology segments. *Puro* and *Neo* functions are used to select for each individual targeting event to produce a double targeted ES cell line. This cell line is then subjected to the transient expression of Cre, a prokaryotic recombinase that catalyzes recombination between the two *loxP* sites. Provided that the *loxP* sites are positioned as direct repeats along the chromosome, the prod-

uct of Cre recombination is a deletion selectable on HAT media due to the reconstitution of the fully functional *Hprt* gene. ES cells carrying segmental deletions are then established in the germline using standard embryonic stem cell technology.

The mouse deletion resource currently available at Baylor College of Medicine, as shown in Fig. 5B, is restricted to chromosome 11. It will be expended in the coming years genome-wide. Not only can these animals be useful in ENU mutagenesis projects, but they also constitute invaluable animal models for chromosomal aberrations frequently found in many human genetic disorders. Just recently, the power of hemizygous deletions was demonstrated through successful modeling of the DiGeorge syndrome and Prader-Willi syndrome for which the key genetic elements were identified (Lindsay *et al.*, 1999; Tsai *et al.*, 1999; Merscher *et al.*, 2001; Lindsay *et al.*, 2001).

SEGMENTAL INVERSIONS IN ENU MUTAGENESIS SCREENS

Chromosomal inversions are also extremely powerful tools in functional analyses of animal genomes (Justice *et al.*, 1997; Rinchik, 2000). The useful feature of an inversion allele is the apparent suppression of meiotic recombination observed within the inversion region (Zheng *et al.*, 1999). If a single recombination event was to occur between the inversely positioned homologous segments, nonviable acentric or dicentric chromosomes would form (Fig. 6A) and such animals never appear in the litter. Therefore, this effect is equivalent to the suppression of recombination and can be conveniently used in genetic crosses to maintain induced mutations within the same chromosome.

Uncovering recessive mutations using segmental inversions requires a three generation crossing scheme and, to avoid labor associated with genotyping, the inversion chromosome and its wild type homologue should be tagged with visible dominant coat color markers (i.e. *K14-agouti*, green fluorescent protein, etc.). Fig. 6B and 6C provide examples of crosses that can be used for this purpose. The scheme in Fig. 6B takes advantage of females homozygous for inversion (Inv) and tagged with the *K14-agouti* minicassette. *K14-agouti* causes yellow tails and yellow ears (yellow). These animals are mated with wild type ENU mutagenized males that carry yet another dominant marker within the chromosomal segment that is inverted in the yellow mice. The green fluorescent protein (GFP) gene is used in this example, but any other visible tag can be considered. The GFP expressing mice are easily distinguishable from their littermates upon UV illumination by emitting green fluorescent light (Hadjantonakis *et al.*, 1998; Hadjantonakis & Nagy, 2001). Yellow and fluorescent carriers are generated in G1. Since these animals may carry a different repertoire of ENU-induced mutations, each of them is taken for the second cross with mice homozygous for the inversion, to generate more yellow, fluorescent carriers in G2. These G2 carriers are intercrossed to obtain the test class of animals that have a wild type appearance in G3. When showing a phenotype, the underlying mutation is likely to be found in these mice within the inversion region. Absence of the test class of animals indicates a recessive lethal mutation that can be easily maintained using yellow, fluorescent animals obtained in G2 and G3. The great advantage of this kind of screen is that it relies totally on visual genotyping.

Another example of a three generation genetic screen takes advantage of engineered balancer chromosomes. Balancer chromosomes were first described at the beginning of the 20th century to be responsible for maintaining heterozygous, lethal mutations in self-perpetuating *D. melanogaster* stocks (Muller, 1918). There are two important features of the balancer chromosome: (i) Recombination events within the chromosome interval

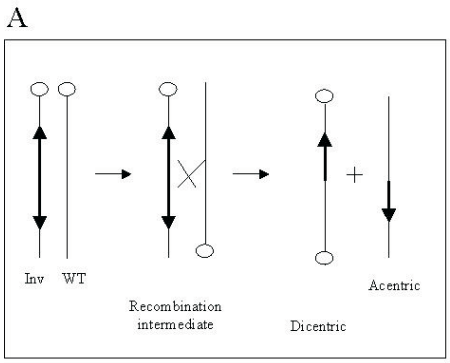

Figure 6A. Meiotic recombination.

How recombination between a wild type and an inversion carrying homologue leads to nonviable dicentric and acentric chromosomes.

must be suppressed (this is most conveniently achieved by segmental inversion), and (ii) a recessive lethal mutation must be present within the recombination suppressive segment. As a consequence, animals homozygous for the balancer will never appear in the stock. Mice carrying engineered balancer chromosomes became available only recently and some of them entered the ENU-mutagenesis screens (Klysik *et al.*, 2002) with the goal to uncover and correlate the induced phenotypes with corresponding mutations. One such ongoing screen is illustrated in Fig. 6C. Heterozygous mice, having a 25 cM *Trp53-Wnt3* inversion tagged by *K14-agouti* and recessive le-

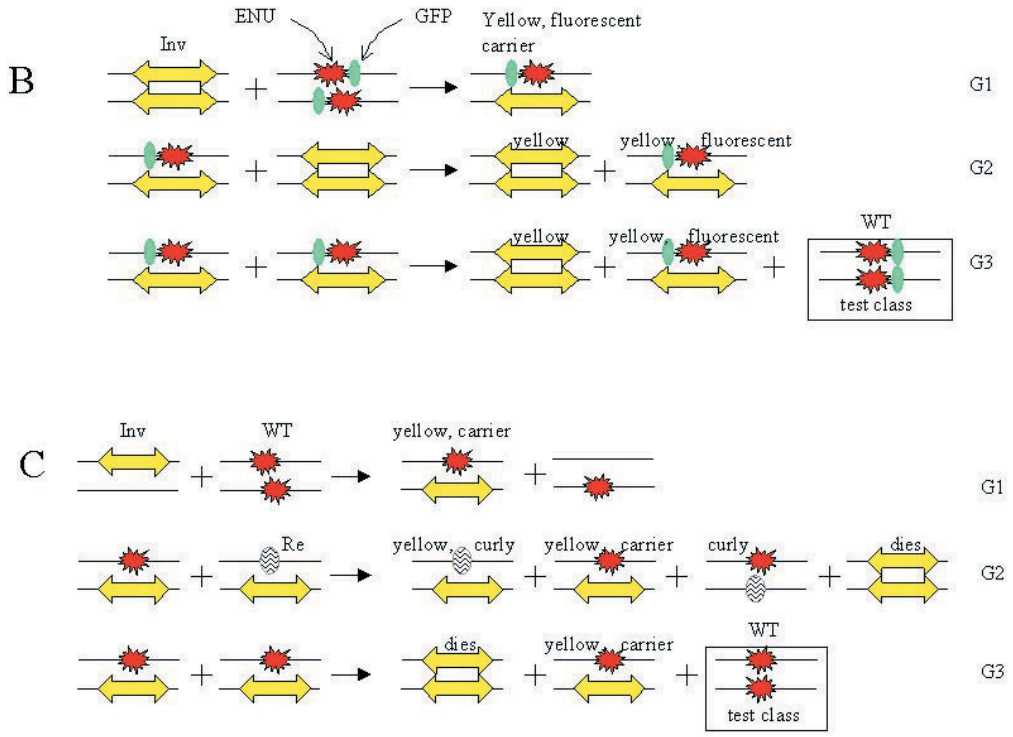

Figure 6B. Genetic screen to uncover recessive, ENU-induced phenotypes using animals homozygous for an inversion.

Although dominant phenotypes are also induced by ENU, they are detectable in G1. Yellow arrow indicates the inversion region marked by the dominant *K14-agouti* minigene. Green circles stand for the enhanced green fluorescent protein dominant transgene (GFP) responsible for emitting green fluorescent light upon UV illumination. Randomly distributed ENU mutations are represented by red explosion symbols.

Figure 6C. Genetic screen that takes advantage of a 25 cM balancer inversion on chromosome 11 (*Trp53-Wnt3*).

The inversion is marked by the dominant *K14-agouti* transgene (yellow) and a recessive lethal mutation at *Wnt3*. *Re* is a dominant mutation at the *Rex* locus that confers wavy coat appearance and curly whiskers. Other symbols as in Fig. 6B.

thal mutation (the *Wnt3* locus is recessive lethal and it was disrupted in the inversion construction process), are mated with ENU mutagenized males to produce yellow founders in G1. The second-generation cross is performed by mating yellow G1 founders with balancer carriers that have the homologous chromosome mutated at the *Rex* locus. In addition to having yellow tails and years, *Rex* animals are marked by showing curly whiskers and wavy coats (Crew & Auerbach, 1939; Carter, 1951; Klysik *et al.*, 2002). All progeny in G2 are distinguishable while inversion homozygous mice die and do not appear in the litter. Yellow carriers from G2 are intercrossed to generate the test class of animals in G3. No genotyping is required in this protocol.

With the power of chromosome engineering technology, a genome-wide collection of segmental inversions and balancers is under construction to facilitate future saturated ENU mutagenesis screens in mice. The basic principle of the construction process is outlined below.

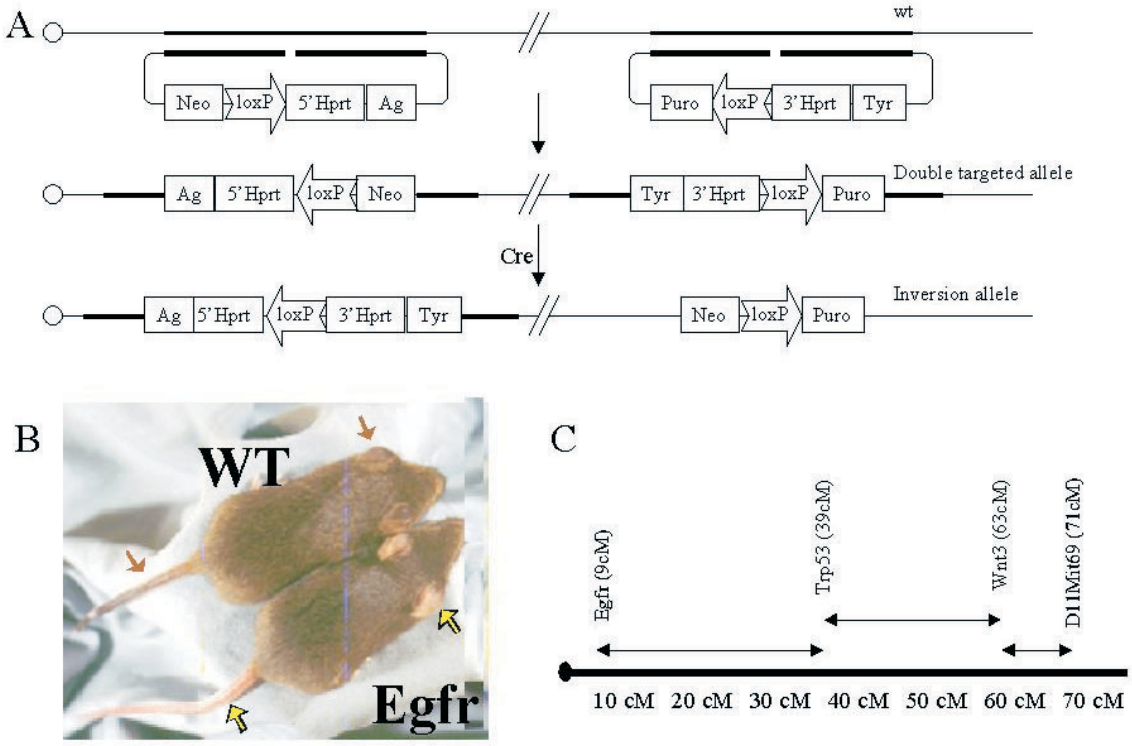

Figure 7. Engineered chromosomal inversions.

A. A diagram for generating segmental inversions.

Elements of the targeting cassettes remain the same as in Fig. 5A. Since the double targeted allele has *loxP* sites positioned in an inverted repetition configuration, the product of Cre recombination is an inversion of the chromosomal segment between the two *loxP* endpoints. Inversion alleles are tagged with coat color markers (*K14-agouti* and *Tyr*) and animals carrying the inversion are easily distinguishable from wild type littermates.

B. Phenotypic manifestation of *K14-agouti*.

Mice display yellow ears and tails. However, these differences are not apparent in *albino* mice. The *K14-agouti* minicassette is located at the epidermal growth factor receptor (*Egfr*) inversion endpoint in this case.

C. The inversion resource (balancers) available for mouse chromosome 11 and generated using chromosome engineering technology.

ENGINEERING CHROMOSOMAL INVERSIONS AND BALANCERS

Chromosomal manipulations that lead to segmental inversions in ES cells are shown in Fig. 7A. While basic components of targeting cassettes remain the same as in the case of generating segmental deletions (cf. Fig. 5A), the protocol differs by establishing *loxP* sequences in an inverted repetition configuration along a chromosome. This configuration dictates that the Cre-catalyzed recombination event leads to the inversion. A tagging function that has not yet failed to produce a visible phenotype is *K14-agouti*, while *Tyr* expression has proven to be more position dependant. Cells with the correct rearrangement are used to generate chimeras and the inversion allele is established in the germ line. Fig. 7B illustrates a typical phenotypic appearance of mice carrying the *K14-agouti* minigene.

Although there are other ways to induce segmental inversions in eukaryotic cells, the site specific targeting of chromosomal loci allows the control of several key parameters: (i) The inversion endpoints can be carefully selected to optimize the length of Cre-induced rearrangement. This is important because the efficiency of ENU screens increases with the increasing length of the inversion interval. However, double cross-overs may occur over distances longer than 30 cM, preventing these constructs from acting as segmental recombination suppressors (Lyon *et al.*, 1982; Zheng *et al.*, 1999; Klysik *et al.*, 2002). Inversions not exceeding 10–30 cM seem to be optimal. (ii) Rearrangements can be conveniently tagged by dominant coat color markers that limit, or virtually eliminate, genotyping labor associated with genetic screens. Finally, (iii) inversion endpoint(s) can be designed to breake in a known recessive lethal function of the region under analysis. Disruption of a recessive lethal gene is required to produce balancer chromosomes.

In mice, the engineered inversion resource is in its infancy and the largest collection of strains carrying these rearrangements is currently available for chromosome 11 (Fig. 7C). Assuming an average inversion length to be about 25 cM, 60 different strains must be developed to cover the entire mouse genome. By facilitating the design of targeting vectors, genomic probes, and markers (Zheng *et al.*, 1999; Klysik *et al.*, 1999), the emergence of a fully annotated mouse genomic sequence will undoubtedly help to accelerate this process (Rogers & Bradley, 2001).

COLLECTING PHENOTYPES AND THEIR CORRELATION WITH CAUSATIVE GENES

The *Wnt3-Trp53* inversion was first to enter the production pipeline of ENU-induced phenotypes. Since a single laboratory or program cannot embrace all aspects of mouse physiology and pathology, the selection of tests and assays in phenotype screens seems particularly important. The tests employed in our work include a complete blood cell count to screen for hyper- and hypoproliferative hematological disorders. The level of protein and glucose is determined in urine to select for kidney abnormalities. Using tandem mass spectrometry, a large number of biochemical metabolites are screened in blood and urine. Skeletal abnormalities are detected using a high resolution X-ray machine.

The collection of phenotypes obtained thus far falls into 8 main categories which can be classified as those that affect embryonic development, eye function, growth, neuro-motor functions, metabolism, hematopoesis, skeleton, skin and coat, and urogenital functions. A more detailed description of phenotypes and the generated mouse mutant resource can be accessed through http://www.mouse-genome.bcm.tmc.edu.

An important element of phenotype-driven genetic analyses is the correlation of the observed phenotype with the underlying gene and mutation. Screens that rely on balancer

chromosomes instantly narrow the area of search to the inversion interval. Further confinement will rely on genetic crosses performed using small, 2–5 cM segmental deletions spanning the inversion area. For this purpose, the deletion mouse resource is under development in parallel to the inversion mice. Finally, the emergence of a fully annotated mouse genomic sequence will eliminate the requirement for further mapping and will alleviate many labor-intensive circumstances associated with finding point-mutated genes using conventional means.

FUTURE OUTLOOK

Mouse genetics has entered the phase of rapid deployment of new tools and novel technologies to efficiently address the function of mammalian genes. ENU mutagenesis combined with the use of genome-wide deletions and inversions is just one example. While identification of almost all the genes in the human genome has already materialized, studies of the functional manifestation of individual genes performed in the context of an entire organism are gaining in impetus. These studies will provide key ingredients in the comprehensive description and understanding of mono- and polygenic disorders and molecular pathways. The number of large-scale mutagenesis programs around the world grows rapidly, yet many challenges still lay ahead. One of them is an easy access to, and timely and systematic cataloging of, the fast growing phenotypic information to complement the already existing databases in the U.S.A. (http://www.jax.org) and in Europe (http://www.mgu.har.mrc.ac.uk/mutabase/). Based on the dynamics of the current progress in functional analyses of mouse genes, expectations rise high, momentum increases rapidly, and the future looks very promising indeed.

The author thanks Drs. Monica Justice, Allan Bradley, James Lupski, Craig Chinault and David Nelson for encouragement and comments.

REFERENCES

Allen ND, *et al.* (1988) Transgenes as probes for active chromosomal domains in mouse development. *Nature.*; **333**: 852–5.

Baribault H, Kemler R. (1989) Embryonic stem cell culture and gene targeting in transgenic mice. *Mol Biol Med.*; **6**: 481–92.

Baum L, *et al.* (2000) Apolipoprotein E isoforms in Alzheimer's disease pathology and etiology *Microsc Res Tech.*; **50**: 278–81.

Beddington RSP. (1992) Transgenic strategies in mouse embryology and development. In *Transgenic animals.* Grosveld FG, Kollias G. eds, pp 79–98. Academic Press, San Diego.

Bottinger EP, Kopp JB. (1998) Lessons from TGF-beta transgenic mice. *Miner Electrolyte Metab.*; **24**: 154–60.

Brandon EP, Idzerda RL, McKnight GS. (1995a) Knockouts. Targeting the mouse genome: a compendium of knockouts (Part I). *Curr Biol.*; **5**: 625–34.

Brandon EP, Idzerda RL, McKnight GS. (1995b) Targeting the mouse genome: a compendium of knockouts (Part II) *Curr Biol.*; **5**: 758–65.

Brandon EP, Idzerda RL, McKnight GS. (1995c) Targeting the mouse genome: a compendium of knockouts (Part III). *Curr Biol.*; **5**: 873–81.

Brayton C, Justice M, Montgomery CA. (2001) Evaluating mutant mice: anatomic pathology. *Vet Pathol.*; **38**: 1–19.

Brennan J, Skarnes WC. (1999) Gene trapping in mouse embryonic stem cells. *Methods Mol Biol.*; **97**: 123–38.

Brinster RL. (1974) The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exp Med.*; **140**: 1049–56.

Brown SD, Nolan PM. (1998) Mouse mutagenesis-systematic studies of mammalian gene function. *Hum Mol Genet.*; **7:** 1627–33.

Brusa R. (1999) Genetically modified mice in neuropharmacology. *Pharmacol Res.*; **39**: 405–19.

Capecchi MR. (1989) Altering the genome by homologous recombination. *Science.*; **244**: 1288–92.

Carter TC. (1951) Wavy-coated mice: phenotypic interactions and linkage tests between rex and (a) waved-1, (b) waved-2. *J Genet.*; **50**: 268–76.

Cecconi F, Meyer BI. (2000) Gene trap: a way to identify novel genes and unravel their biological function. *FEBS Lett.*; **480**: 63–71.

Chen Z, Friedrich GA, Soriano P. (1994) Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.*; **8**: 2293–301.

Chowdhury K, *et al.* (1997) Evidence for the stochastic integration of gene trap vectors into the mouse germline *Nucleic Acids Res.*; **25**: 1531–6.

Clarke AR. (2000) Manipulating the germline: its impact on the study of carcinogenesis. *Carcinogenesis.*; **21**: 435–41.

Crew FAE, Auerbach C. (1939) Rex: a dominant autosomal monogenic coat texture character in the mouse. *J Genet.*; **38**: 341–4.

El-Sohemy A, Archer MC. (1999) Regulation of mevalonate synthesis in low density lipoprotein receptor knockout mice fed n-3 or n-6 polyunsaturated fatty acids. *Lipids.*; **34**: 1037–43.

Evans MJ, Kaufman MH. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature.*; **292**: 154–6.

Friedrich G, Soriano P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.*; **5**: 1513–23.

Gu H, *et al.* (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science.*; **265**: 103–6.

Hadjantonakis AK, Nagy A. (2001) The color of mice: in the light of GFP-variant reporters. *Histochem Cell Biol.*; **115**: 49–58.

Hadjantonakis AK, *et al.* (1998) Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech Dev.*; **76**: 79–90.

Hanahan D. (1989) Transgenic mice as probes into complex systems. *Science.*; **246**: 1265–75.

Hidaka M, *et al.* (2000) Gene trapping of two novel genes, Hzf and Hhl, expressed in hematopoietic cells. *Mech Dev.*; **90**: 3–15.

Hitotsumachi S, Carpenter DA, Russell WL. (1985) Dose-repetition increases the mutagenic effectiveness of *N*-ethyl-*N*-nitrosourea in mouse spermatogonia. *Proc Natl Acad Sci U S A.*; **82**: 6619–21.

Holschneider DP, Shih JC. (2000) Genotype to phenotype: challenges and opportunities. *Int J Dev Neurosci.*; **18**: 615–8.

Hong HK, *et al.* (2001) The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice. *Genesis.*; **29**: 163–71.

Ihle JN. (2000) The challenges of translating knockout phenotypes into gene function. *Cell.*; **102**: 131–4.

Jaenisch R. (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc Natl Acad Sci U S A.*; **73**: 1260–4.

Justice MJ, Bode VC. (1988) Three ENU-induced alleles of the murine quaking locus are recessive embryonic lethal mutations. *Genet Res.*; **51**: 95–102.

Justice MJ, Zheng B, Bradley A. (1997) Using targeted large deletions and high-efficiency *N*-ethyl-*N*-nitrosourea mutagenesis for functional analyses of the mammalian genome. *Methods.*; **13**: 423–36.

Kaminski N. (2000) Bioinformatics A user's perspective. *Am J Respir Cell Mol Biol.*; **23**: 705–11.

Klysik J, *et al.* (1999) An integrated gene and SSLP BAC map framework of mouse chromosome 11 *Genomics.*; **62**: 123–8.

Klysik J. (2001) Concept of immunomics: a new frontier in the battle for gene function? *Acta Biotheor.*; **49**: 191–202.

Klysik J, Justice M, Bradley A. (2002) Chromosome engineering and ENU mutagenesis: their use for defining gene function. *Curr Genomics.*; (in press).

Kollias F, Grosveld FG. (1992) Transgenic strategies in mouse embryology and development. In *Transgenic Animals.* Grosveld FG, Kollias G. eds, pp 47–77. Academic Press, San Diego.

Kothary R, *et al.* (1988) A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature.*; **335**: 435–7.

Lander ES, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature.*; **409**: 860–921.

Lewandoski M. (2001) Conditional control of gene expression in the mouse. *Nat Rev Genet.*; **2**: 743–55.

Lindsay EA, *et al.* (1999) Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature.*; **401**: 379–83.

Lindsay EA, *et al.* (2001) *Tbx1* haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature.*; **410**: 97–101.

Lyon MF, Phillips RJ, Fisher G. (1982) Use of an inversion to test for induced X-linked lethals in mice. *Mutat Res.*; **92**: 217–28.

Martin GR. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A.*; **78**: 7634–8.

McClatchey AI, Jacks T. (1998) Tumor suppressor mutations in mice: the next generation. *Curr Opin Genet Dev.*; **8**: 304–10.

Merscher S, *et al.* (2001) TBX1 is responsible for cardiovascular defects in velo-cardio-facial/ DiGeorge syndrome. *Cell.*; **104**: 619–29.

Miklos GL, Rubin GM. (1996) The role of the genome project in determining gene function: insights from model organisms. *Cell.*; **86**: 521–9.

Minabe-Saegusa C, *et al.* (1998) Sequence and expression of a novel mouse gene PRDC (protein related to DAN and cerberus) identified by a gene trap approach. *Dev Growth Differ.*; **40**: 343–53.

Mintz B, Cronmiller C. (1978) Normal blood cells of anemic genotype in teratocarcinoma-derived mosaic mice. *Proc Natl Acad Sci U S A.*; **75**: 6247–51.

Moore KJ. (1999) Utilization of mouse models in the discovery of human disease genes. *Drug Discov Today.*; **4**: 123–8.

Muller HJ. (1918) Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics.*; **3**: 422–99.

Nolan PM, *et al.* (2000) Implementation of a large-scale ENU mutagenesis program: towards increasing the mouse mutant resource. *Mamm Genome.*; **11**: 500–6.

Noveroske JK, Weber JS, Justice MJ. (2000) The mutagenic action of *N*-ethyl-*N*-nitrosourea in the mouse. *Mamm Genome.*; **11**: 478–83.

Palapattu GS, *et al.* (1998) Transgenic mouse models for tumor suppressor genes. *Cancer Detect Prev.*; **22**: 75–86.

Palmiter RD, Brinster RL. (1986) Germ-line transformation of mice. *Annu Rev Genet.*; **20**: 465–99.

Papaioannou VE, *et al.* (1975) Fate of teratocarcinoma cells injected into early mouse embryos. *Nature.*; **258**: 70–3.

Pennisi E. (2000) Human genome. Finally, the book of life and instructions for navigating it. *Science.*; **288**: 2304–7.

Pirity M, Hadjantonakis AK, Nagy A. (1998) Embryonic stem cells, creating transgenic animals. *Methods Cell Biol.*; **57**: 279–93.

Pitot HC, *et al.* (2000) Review article: the stages of gastrointestinal carcinogenesis — application of rodent models to human disease. *Aliment Pharmacol Ther.*; **14**: (Suppl 1) 153–60.

Postic C, *et al.* (1999) Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic β cell-specific gene knock-outs using Cre recombinase. *J Biol Chem.*; **274**: 305–15.

Ramirez-Solis R, Liu P, Bradley A. (1995) Chromosome engineering in mice. *Nature.*; **378**: 720–4.

Rathkolb B, *et al.* (2000) Large-scale *N*-Ethyl-*N*-nitrosourea mutagenesis of mice — from phenotypes to genes. *Exp Physiol.*; **85**: 635–44.

Rinchik EM. (1991) Chemical mutagenesis and fine-structure functional analysis of the mouse genome. *Trends Genet.*; **7**: 15–21.

Rinchik EM. (2000) Developing genetic reagents to facilitate recovery, analysis, and maintenance of mouse mutations. *Mamm Genome.*; **11**: 345–51.

Rinchik EM, Carpenter DA. (1999) *N*-ethyl-*N*-nitrosourea mutagenesis of a 6- to 11-cM subregion of the Fah-Hbb interval of mouse chromosome 7: Completed testing of 4557 gametes and deletion mapping and complementation analysis of 31 mutations. *Genetics.*; **152**: 373–83.

Rinchik EM, Carpenter DA, Selby PB. (1990) A strategy for fine-structure functional analysis of a 6- to 11-centimorgan region of mouse chromosome 7 by high-efficiency mutagenesis. *Proc Natl Acad Sci U S A.*; **87**: 896–900.

Robertson EJ. (1987) Embryo-derived stem cell lines. In *Teratocarcinomas and embryonic stem cells: a practical approach.* Robertson EJ. ed, pp 71–112. IRL Press, Oxford.

Rogers J, Bradley A. (2001) The mouse genome sequence: status and prospects. *Genomics.*; **77**: 117–8.

Ross AJ, *et al.* (1998) Testicular degeneration in *Bclw*-deficient mice. *Nat Genet.*; **18**: 251–6.

Sam M, *et al.* (1998) Aquarius, a novel gene isolated by gene trapping with an RNA-dependent RNA polymerase motif. *Dev Dyn.*; **212**: 304–17.

Shedlovsky A, *et al.* (1993) Mouse models of human phenylketonuria. *Genetics.*; **134**: 1205–10.

Shedlovsky A, King TR, Dove WF. (1988) Saturation germ line mutagenesis of the murine t region including a lethal allele at the quaking locus. *Proc Natl Acad Sci U S A.*; **85**: 180–4.

Shibuya T, Morimoto K. (1993) A review of the genotoxicity of 1-ethyl-1-nitrosourea. *Mutat Res.*; **297**: 3–38.

Sidman RL, Dicke MM, Appel SH. (1964) Mutant mice (quaking and jimpy) with deficient myelination in the central nervous system. *Science.*; **144**: 309–11.

Si-Hoe SL, Wells S, Murphy D. (2001) Production of transgenic rodents by the microinjection of cloned DNA into fertilized one-cell eggs. *Mol Biotechnol.*; **17**: 151–82.

Skarnes WC, *et al.* (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc Natl Acad Sci U S A.*; **92**: 6592–6.

Smith AG, *et al.* (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature.*; **336**: 688–90.

Soewarto D, *et al.* (2000) The large-scale Munich ENU-mouse-mutagenesis screen. *Mamm Genome.*; **11**: 507–10.

Sreekumar KR, Aravind L, Koonin EV. (2001) Computational analysis of human disease-associated genes and their protein products. *Curr Opin Genet Dev.*; **11**: 247–57.

Takeuchi T, *et al.* (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev.*; **9**: 1211–22.

Thorey IS, *et al.* (1998) Selective disruption of genes transiently induced in differentiating mouse embryonic stem cells by using gene trap mutagenesis and site-specific recombination. *Mol Cell Biol.*; **18**: 3081–8.

Tsai TF, *et al.* (1999) Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene

contributing to Prader-Willi syndrome. *Hum Mol Genet.*; **8**: 1357–64.

Tsoka S, Ouzounis CA. (2000) Recent developments and future directions in computational genomics. *FEBS Lett.*; **480**: 42–8.

Venter JC, *et al.* (2001) The sequence of the human genome. *Science.*; **291**: 1304–51.

Wobus AM, *et al.* (1984) Characterization of a pluripotent stem cell line derived from a mouse embryo. *Exp Cell Res.*; **152**: 212–9.

Woychik RP, Alagramam K. (1998) Insertional mutagenesis in transgenic mice generated by the pronuclear microinjection procedure. *Int J Dev Biol.*; **42**: 1009–17.

Zambrowicz BP, Friedrich GA. (1998) Comprehensive mammalian genetics: history and future prospects of gene trapping in the mouse. *Int J Dev Biol.*; **42**: 1025–36.

Zheng B, Mills AA, Bradley A. (1999) A system for rapid generation of coat color-tagged knockouts and defined chromosomal rearrangements in mice. *Nucleic Acids Res.*; **27**: 2354–60.

Zheng B, Mills AA, Bradley A. (2001) Introducing defined chromosomal rearrangements into the mouse genome. *Methods.*; **24**: 81–94.

Zheng B, *et al.* (1999) Engineering a mouse balancer chromosome. *Nat Genet.*; **22**: 375–8.