Tools for targeted manipulation of the mouse genome

LOUISE VAN DER WEYDEN, DAVID J. ADAMS, AND ALLAN BRADLEY

The Wellcome Trust Sanger Institute, Hinxton, Cambs CB10 1SA, United Kingdom

Received 13 June 2002; accepted in final form 30 July 2002

van der Weyden, Louise, David J. Adams, and Allan Bradley. Tools for targeted manipulation of the mouse genome. Physiol Genomics 11: 133–164, 2002; 10.1152/physiolgenomics.00074.2002.—In the post-genomic era the mouse will be central to the challenge of ascribing a function to the 40,000 or so genes that constitute our genome. In this review, we summarize some of the classic and modern approaches that have fueled the recent dramatic explosion in mouse genetics. Together with the sequencing of the mouse genome, these tools will have a profound effect on our ability to generate new and more accurate mouse models and thus provide a powerful insight into the function of human genes during the processes of both normal development and disease.

gene targeting; homologous recombination; embryonic stem cell; mutagenesis; gene trap; mouse genome project

1) BACKGROUND

1.1) Why the Mouse?

The Human Genome Project has revealed the sequence information of many of the genes that make us what we are. However, although at least 35,000 human genes have been sequenced and mapped, adequate expression or functional information is available for only 15% of them (230). Therefore, a significant challenge for scientists over the next few decades is to annotate the human genome with functional information. This effort will enable us to gain an understanding of the molecular mechanisms and pathways underlying normal development, as well as those responsible for pathogenesis. However, to do this we need an experimental model. The mouse is an excellent experimental model for defining human gene function because of its anatomic, physiologic, and genetic similarity to humans. The mouse is also a popular model because of its relatively short life cycle and because its genome can be readily manipulated by molecular means. For example, mouse geneticists can eliminate or overexpress genes in the whole animal or in a specific tissue, introduce large pieces of self or foreign DNA into the genome, and engineer whole chromosomes. Furthermore, inbred strains of mice also provide the opportunity to study a disease trait in a defined genetic background, allowing distinction between the phenotypes conferred by a single mutation vs. the contributions of other genetic modifiers. A large genetic reservoir of potential models of human disease has been generated through the identification of spontaneous, radiation-, or chemical-induced mutant loci in mice 100 mouse models of human disease where the homologous gene has been shown to be mutated in both human and mouse (19), indicating the validity of using mice to model human disease.

In the last few years, a number of significant technological advances have dramatically increased our ability to create mouse models of human disease. These technological advances have been greatly aided by the sequencing of the mouse genome and the subsequent mouse genomic resources that have been developed.

1.2) The Mouse Genome Sequencing Project

The C57BL/6J strain has been selected as the reference strain for the production of a finished genome sequence (17). The mouse genome sequencing project has two main aims; the first is to generate a finished sequence of the mouse genome that is freely accessible to the public, and the second is to create a richly annotated information resource consisting of a database containing information about the underlying genomic sequence (reviewed in Ref. 129). Mouse genomic resources include mapping resources [such as dense genetic maps...
(51, 168), sequence tagged site (STS)-based physical maps (168), radiation hybrid (gene-based) maps (11, 88, 250), and simple sequence length polymorphism (SSLP) marker-anchored bacterial artificial chromosome (BAC) framework maps (33); DNA resources [such as the generation of expressed sequence tags (ESTs) (144), cDNAs (109, 233) and BAC libraries (175, 283); and database resources, which help to provide community access to this information (see below for URLs).

1.2.1) How is the mouse genome being sequenced?
The sequencing of the mouse genome involves two parts: a genome-wide program, and a targeted program (summarized in Ref. 129). The genome-wide sequencing program is being undertaken by the Mouse Genome Sequencing Consortium (MGSC), which is an international collaboration between four centers [the Wellcome Trust Sanger Institute; The Whitehead Institute/MIT Center for Genome Research; The Washington University Genome Sequencing Center; and Ensembl (a joint project between the Sanger Institute and the European Bioinformatics Institute)]. Incorporated into the genome-wide sequence will be the sequence of regions targeted by specific projects. These include the Trans-NIH BAC sequencing program (funding sequencing of specific BACs/regions of biological interest by sequencing facilities at Cold Spring Harbor Laboratory, University of Oklahoma, and the Albert Einstein College of Medicine), the UK-MRC Mouse Sequencing Programme (adopting a regional and functional approach to target four “core regions” on chromosomes 4, 13, 2, and X), and the Joint Genome Institute (performing comparative sequencing of mouse chromosomal regions syntenic with human chromosome 19). The annotation of the mouse genome will be facilitated by cDNA sequencing programs, including that carried out by the RIKEN Institute (who collect data on most, if not all, full-length cDNAs, their primary structures, and expression sites) and the Mammalian Gene Collection [MGC; who provide a complete set of full-length (open reading frame) sequences and cDNA clones of human and mouse genes (233)]. In addition, the MGSC will do clone-by-clone BAC sequencing to a high standard.

There are several components to the genome-wide sequencing effort. First, there is the development of a BAC map, consisting of overlapping BAC clones covering the vast majority of the mouse genome. For this, two BAC libraries have been selected: RPCI-23 (female) and RPCI-24 (male) (175). Although previously characterized by sequencing both ends of the cloned inserts (“BAC-end sequencing”) (283), as part of the MGSC they are being characterized by restriction-digest fingerprinting of each clone (“fingerprinting”), which can be used to assemble the BACs into overlapping clone “contigs” (continuous stretches of assembled sequences) covering the mouse genome. The BAC map and BAC end-sequences provide a crucial scaffold for assembling the sequence of the mouse genome. The second component uses paired-end whole-genome shotgun reads to generate light (~2.5-fold) and deeper (5- to 6-fold) coverage of the mouse genome; the sequence is obtained by preparing plasmid libraries (with inserts of 2–10 kb) and sequencing both ends of the inserts (“paired-end sequencing”). By combining the BAC map and the shotgun coverage, a “hybrid” genome assembly can be constructed, consisting of sequence contigs, linked into “supercontigs” (neighboring contigs that have been properly organized and oriented), which are further linked into “ultracontigs” (neighboring supercontigs that have been properly organized and oriented). In addition, comparative sequence information from other strains of mice, such as 129S1/SvImJ, BALB/cByJ, and C3H/HeJ will also be obtained, as such analysis is important for identifying sequence variants, such as single nucleotide polymorphisms (SNPs).

As of April 2002, two assemblies of the complete whole genome data set for the mouse genome have been generated: one by the Whitehead group using the ARACHNE program (18), and the other by the Sanger Institute using the PHUSION program. Both assemblies used the February 2002 freeze of data, and were done using the same set of files (each assembly included about 33 million reads, corresponding to ×7 coverage of the genome). Although both assemblies were evenly matched by many criteria, the ARACHNE assembly was chosen by the MGSC as the one with which to proceed for further analysis. This assembly provides 96% coverage of the euchromatic mouse genome and predicts with high confidence 22,444 genes across the genome, of which ~75% have a firm human genome counterpart. This “MGSC Version 3” assembly can be downloaded (ftp://ftp.ensembl.org/pub/assembly/mouse/mgsc_assembly_3) or searched using SSAHA (http://www.ensembl.org/Mus_musculus/ssahaview) or BLAST (http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html). In addition, the BAC map resource has been aligned to the sequence, and an initial functional annotation of these genes has been added, as well as a comparison to the human genome (at the DNA level). The final phase of the MGSC’s work is to now fill in the “gaps” and correct any errors or misassemblies (for a completed genome sequence to be available within the next 3 yr).

All information generated by the MGSC is rapidly released to the scientific community; a constantly updated and comprehensive view of the project’s data is provided at a central MGSC server (http://mouse.ensembl.org/). In addition, MGSC data is incorporated into other key genome servers, including mouse genome resources at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/genome/index.html), the MRC UK mouse sequencing program (http://mrcseq.har.mrc.ac.uk), and the mouse genomic sequence reads aligned against the human draft sequence at the University of California at Santa Cruz (http://genome.cse.ucsc.edu).

1.3) Basic Genetic Technologies for Mining the Mouse Genome

With the exponential increase in the number of genes identified by the genome sequencing projects, it
has become imperative that efficient methods be developed for determining gene function. It has now become possible to make essentially any mutation in the germ line of mice by utilizing homologous recombination (the recombining of an exogenous piece of DNA with its endogenous homologous sequence in vivo) and embryonic stem (ES) cells (the pluripotent derivatives of the inner cell mass of the blastocyst) (34). Gene targeting (using homologous recombination to alter specific endogenous genes) provides the highest level of control over producing mutations. The combination of gene targeting techniques in mouse ES cells and the Cre-loxP recombination system has resulted in the emergence of chromosomal engineering technology in mice. This advance has opened up new opportunities for modeling new diseases that are associated with chromosomal rearrangements (in humans, chromosomal abnormalities are a principle cause of fetal loss and developmental disorders (214), and chromosomal translocations are involved in the genesis of many types of human tumors (189)). Mutagenesis studies in the mouse have emphasized modeling human diseases through phenotype-driven assays. The chromosome engineering strategy for generating mice with precise chromosomal rearrangements, in combination with mutagenesis approaches, provides an invaluable tool for deciphering gene function using phenotype-based methods. In addition, gene trap approaches in ES cells also offer a route to gene discovery by providing information on gene sequence, expression, and mutant phenotype.

2) GENE TARGETING

2.1) Gene Targeting Strategies

Genetically modified mice can be generated either by direct pronuclear injection of exogenous DNA into fertilized zygotes (177) or injection of genetically modified mouse ES cells into a blastocyst (115). Direct pronuclear injection results in random integration of the injected DNA into the genome and relies on the overexpression of the transgene to produce a phenotype. In contrast, ES cells have the advantage in that they can be genetically modified by means of homologous recombination (a process by which a fragment of genomic DNA introduced into a mammalian cell can locate and recombine with the endogenous homologous sequence) prior to being injected into the blastocyst. This process is known as “gene targeting” and was first reported in mammalian cells in 1985 in erythroleukemia cells (226) and a fibroblast cell line (127). The first examples of gene targeting by homologous recombination in ES cells targeted the hypoxanthine-guanine phosphoribosyltransferase (HPRT) gene (a selectable locus on the X chromosome) (52, 239). The first reports of germ-line transmission of a targeted allele in ES cells also occurred at the HPRT locus (241), as well as the c-abl locus (212). Since then, gene targeting has been widely used in ES cells to make a variety of genetic mutations in many different loci, allowing the phenotypic consequences of the modifications to be assessed.

The procedure for the generation of mice that have been genetically modified using gene targeting strategies is essentially the same regardless of the specific targeting strategy used and is outlined in Fig. 1. Briefly, the targeting construct (containing sequence homologous to the targeted gene and a selectable marker) is electroporated into ES cells. These cells are then cultured in the presence of a selection agent to remove any cells that have not stably integrated the construct into their genome. The surviving ES cell colonies are then isolated and examined for the presence of the targeted allele (to ensure the desired recombination event has occurred) by PCR amplification or Southern blot analysis. Those containing the targeted allele are then injected into blastocysts (embryonic day 3.5) and transferred to the uterus of a foster mother. The resulting pups are then examined for their degree of chimerism (percentage of the genetic makeup of the mouse that was contributed by the ES cell). Male mice showing a high percentage of chimerism are then mated with wild-type mice to check for germ-line transmission of the targeted allele in the F1 offspring. The F1 heterozygotes may then be intercrossed to breed to homozygosity.

Important aspects of any gene targeting experiments include the vector design (which can determine the efficiency of recombination), the type of mutation generated at the target locus, selection cassettes (detailed in Table 1), and screening strategies used to identify clones of ES cells with the desired targeted modification (detailed in Ref. 38). The different types of mutations, and the vectors needed to generate such mutations, are discussed below in more detail.

2.1.1) Generating “knockout” mice: the null allele. The most common experimental strategy is to ablate the function of a target gene (generating a null allele) by introducing a selectable marker gene (for a review see Ref. 38). A targeting vector is designed to recombine with and mutate a specific chromosomal locus. The minimal components of such a vector are sequences of DNA homologous with the desired chromosomal integration site, a selection cassette, and a plasmid backbone. Two basic types of vectors can be used for targeting in mammalian cells, namely, replacement and insertion vectors. The fundamental aspects of a replacement vector are sequences of isogenic DNA with homology to the target locus (5–8 kb), a positive selectable marker (see Table 1), bacterial plasmid sequences, and a linearization site outside of the homologous sequences of the vector. The final postrecombination product is a replacement of the chromosomal homology with all components of the vector flanked on both sides by homologous sequences (see Fig. 2, A and B) (reviewed in Refs. 38, 82). The basic elements of an insertion vector are the same as those in a replacement vector. However, the major difference between the two vector types is that the linearization site of an insertion vector is made within the homologous sequences. An insertion vector undergoes single reciprocal recombination (vector insertion) with its homologous chromosomal target, which is stimulated by the double-strand...
break (or gap) in the vector. Insertion vectors have been reported to target at a 5- to 20-fold higher efficiency than replacement vectors given the same homologous sequences (81). Since the entire insertion vector is integrated into the target site, including the homologous sequences of the vector, the recombinant allele becomes a duplication of the target homology separated by the heterologous sequences in the vector backbone (see Fig. 2C).

2.1.2) Generating subtle mutations. Although major disruptions of a locus represent a valuable starting point for genetic analysis, a series of changes at the nucleotide level in both coding and control regions of the gene are important for the full understanding of a gene’s function. Thus there are techniques to allow the inclusion of small mutations in the genome, such as the “hit-and-run” and “double replacement” strategies. Hit-and-run vectors are essentially insertion vectors containing both positive and negative selectable markers outside the region of homology and the desired mutation in the region of homology (Fig. 3). This strategy requires two rounds of homologous recombination: homologous recombination and positive selection are used to generate a duplication at the target locus (via insertion of the targeting vector), followed by resolution of the duplication (excision) via a second homologous recombination event between the duplicated homologous sequences (Fig. 3). This results in removal of

Table 1. Examples of some commonly used dominant selectable markers

<table>
<thead>
<tr>
<th>Selection Cassette</th>
<th>Gene</th>
<th>Selection</th>
<th>Drug(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blasticidin S deaminase</td>
<td><em>bur</em></td>
<td>+</td>
<td>blasticidin S</td>
</tr>
<tr>
<td>Diphtheria toxin A fragment</td>
<td><em>DTA</em></td>
<td>–</td>
<td>no drugs needed</td>
</tr>
<tr>
<td>Herpes simplex virus thymidine kinase</td>
<td><em>HSVtk</em></td>
<td>–</td>
<td>gancyclovir, FIAU</td>
</tr>
<tr>
<td>Hygromycin B phosphotransferase</td>
<td><em>hph</em></td>
<td>+</td>
<td>hygromycin B</td>
</tr>
<tr>
<td>Neomycin phosphotransferase</td>
<td><em>neo</em></td>
<td>+</td>
<td>neomycin sulfate (G418)</td>
</tr>
<tr>
<td>Puromycin-N-acetyltransferase</td>
<td><em>puro</em></td>
<td>+</td>
<td>puromycin</td>
</tr>
<tr>
<td>Puromycin-N-acetyltransferase-thymidine kinase fusion</td>
<td><em>puΔtk</em></td>
<td>+/–</td>
<td>puromycin/gancyclovir, FIAU</td>
</tr>
<tr>
<td>Xanthine/guanine phosphoribosyltransferase</td>
<td><em>gpt</em></td>
<td>+/–</td>
<td>mycophenolic acid/6-thioxanthine</td>
</tr>
</tbody>
</table>

Positive selection markers serve to isolate rare transfected cells that have stably integrated DNA (which occurs at a frequency of ~1 in 10⁴ treated cells), whereas negative selection markers serve to eliminate various subpopulations of these transfected cells. It is important to note that selectable markers associated with targeted alleles carry promoter and enhancer sequences, which can interfere with other genes linked to the targeted locus (173, 182). Therefore, to generate mutations without having to interpret the effects of selectable markers on the phenotype, marker removal is routinely performed, such as with the use of site-specific recombinases (see Section 2.2) or homologous recombination techniques (see Section 2.1.3). FIAU, 1/1-2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil.
the plasmid sequences, selection markers, and one complement of the duplication from the target locus. Hit-and-run targeting procedures have been successfully used to generate mutations in ES cells in the HPRT locus (247), the Hox-2.6 locus (80), the \( \beta \)-amyloid precursor protein locus (73), the \( G_{i2o} \) locus (202), and the cystic fibrosis transmembrane conductance regulator (\( C_{fr} \)) locus (49) (with germ-line transmission of the mutated allele being demonstrated in the latter two examples). Although hit-and-run targeting procedures are typically performed in ES cells in vitro, with the modified ES cells subsequently being injected into blastocysts to generate mice carrying the mutation, one group has modified this technique to generate mice carrying a subtle mutation in the \( K_{-ras} \) gene, whereby the targeted insertion allele was created in cultured ES cells, and the excision was performed in mice in vivo (95).

The “double replacement” (or “tag-and-exchange”) technique also relies on two rounds of homologous recombination, although the second round of recombination is also a gene targeting event (i.e., vector-chromosome recombination rather than chromosome-chromosome recombination). Double replacement vectors are replacement vectors containing both positive and negative selectable markers inside the region of homology (Fig. 4). Using this strategy, the targeted gene is first replaced with a selectable marker to inactivate the allele (“tagging”), followed by retargeting with a second vector to reconstruct the inactivated allele, and introduce the engineered mutation (“exchanging”). This strategy has been used to introduce subtle mutations in mouse ES cells in the \( Col1a-1 \) gene (269), \( \alpha 2 \) Na,K-ATPase gene (7), and the Huntington’s disease homolog (\( Hdh \)) gene (36) (with germ-line transmission of the mutated allele demonstrated in the latter example). A modified approach, termed “stable tag exchange,” incorporates an additional positive selection cassette to circumvent the problem of high levels of nonhomologous recombinants, which survive selection due to loss of tag gene expression (266). The advantage of this technique is that any number of different vectors can be used in the second step to generate an array of mutations at the same locus, as has been used in ES cells to generate a series of different targeted alterations to the prion protein gene (153). However, as with the hit-and-run method, a limitation of this procedure is the spontaneous loss of sensitivity to the negative selection agent due to events other than homologous recombination, such as mutation of the negative selectable marker (82).

In contrast to conventional targeting techniques, which generate ES cells with a selectable marker in the mutated (targeted) locus, both the hit-and-run and double replacement techniques generate ES cells with mutations, yet no selectable marker. This is important, as most of these cassettes contain a promoter/enhancer and polyadenylation signal, all of which have the potential to interfere with transcription of neighboring genes.

---

1 Alternatively, a single marker such as the \( Hprt \) mini-gene can be used, allowing for both positive and negative selection in the appropriate genetic background (228).
genes and hence generate a phenotype that is difficult to interpret (89, 182). Another way of eliminating the problem associated with the presence of the selectable marker in the targeted locus is to use a site-specific recombination system, such as loxP sites flanking the marker to allow its excision through the actions of Cre recombinase (discussed in more detail in Section 2.2).

2.1.3) Generating conditional mice. Conditional activation (“gain-of-function”) or inactivation (“loss-of-function”) of gene expression in vivo can be used to control gene expression in a temporal-, spatial-, or tissue-specific manner. The tight regulation of gene expression offered by conditional gene targeting means avoidance of any potential embryonic lethality associated with the mutation and allows for a more accurate mouse model of human disease and sporadic cancer initiation and progression (reviewed in Ref. 97).

Such control can be achieved using binary transgenic systems, in which gene expression is regulated by the interaction of an effector protein product on a target transgene. These interactions can be controlled by crossing mouse lines or by adding/removing an exogenous inducer. Binary transgenic systems fall into two categories: transcriptional transactivation (used to activate transgenes in gain-of-function experiments) and site-specific DNA recombination (used to activate transgenes or to generate tissue-specific gene knockouts). Transcriptional transactivation is more widely used than DNA recombination to bring about transgene activation in mice, because the latter is irreversible, whereas the former occurs only when the transactivator is present (thus if an inducible system is used, then transgene activation can be reversed at a given time, providing greater control over transgene expression) (reviewed in Ref. 123).

2.1.3.1) TRANSCRIPTIONAL TRANSACTIVATION. The most widely used binary transcription transactivation systems are the tetracycline-dependent regulatory systems (“Tet-on” and “Tet-off”; see http://www.clontech.com/tet/index.shtml or http://www.zmbh.uni-heidelberg.de/bujard/Homepage.html) (70). These systems use a chimeric transactivator to control transcription of the gene of interest from a silent promoter and are based on two regulatory elements derived from the E. coli tetracycline-resistance operon: the Tet repressor protein (TetR) and the Tet operator sequence (tetO). The Tet-off system uses a plasmid that expresses a fusion protein known as the tetracycline-controlled transactivator (tTA), composed of TetR and the VP16 activation domain of herpes simplex virus (Fig. 5A). tTA binds to the tetracycline response element (TRE) and activates transcription of the target gene in the absence of the inducer, doxycyclin (Dox).2 The Tet-on system uses a form of TetR containing four amino acid changes that result in altered binding characteristics and create the reverse TetR (rTetR). As shown in Fig. 5B, rTetR binds the TRE and activates transcription of the target gene in the presence of Dox. Recent examples

\footnote{Dox is the inducer of choice because of its low cost, commercial availability, and efficient activation/inactivation of rTet/tTA at doses well below cytotoxic levels (16, 71).}
of use of the Tet system in transgenic mice include the regulation of delta-FosB expression to control bone formation and bone mass (221), regulation of the SV40 T-antigen oncogene to control 

This page contains a diagram illustrating the process of double replacement targeting. The diagram shows a sequence of steps involving gene replacement, selection, and recombination. The text describes the process in detail, mentioning the use of Cre or Flp recombinases to inactivate endogenous genes and the ability to control gene expression temporally.

Another transcriptional transactivation system is the Gal4-based system, from Saccharomyces cerevisiae. The transcriptional activator Gal4 directs the expression of responsive genes by binding to upstream activating sequences (UAS), such that UAS-linked target genes can be activated (174). The system can be made inducible if Gal4 is linked to ligand binding domain of progesterone receptor and VP16 activator (GLVP). GLVP can be activated by binding to synthetic steroids, such as RU-486 or ZK-98.734. However, the use of the GLVP system has generally been limited to controlling gene expression in adult mice (183, 260–261), as the synthetic steroids have been shown to induce abortions.

2.1.3.2 Site-specific DNA Recombination. The most widely used site-specific DNA recombination system uses the Cre recombinase from bacteriophage P1, and more recently the Flp recombinase from S. cerevisiae has begun to show utility (discussed in more detail in Section 2.2). By using gene targeting techniques to produce mice with modified endogenous genes that can be acted on by Cre or Flp recombinases expressed under the control of tissue-specific promoters, site-specific recombination can be used to inactivate endogenous genes in a spatially controlled manner. Cre/Flp activity can also be controlled temporally by 1) delivering cre/FLP-encoding transgenes in viral vectors, 2) administering exogenous steroids to mice that carry a chimeric transgene consisting of the cre gene fused to a mutated ligand-binding domain, 3) using transcriptional transactivation to control cre/FLP expression, and 4) using soluble Cre. There are numerous examples of conditional mice that have been generated using the loxP/Cre recombinase system (19b).
and Otx1 in development of the rostral head) (236). The most widely used knock-in strategy is the replacement of a gene by a reporter gene such as LacZ or GFP, to monitor expression pattern of the gene during development and in the adult mouse, both in a spatial and temporal manner. For example, LacZ has been knocked into the Pax3 locus (to examine the role of Pax3 in the neuroepithelium and somites) (143). A knock-in allele usually results in the loss-of-function of the endogenous gene.

Although the early knock-in studies in mouse ES cells utilized the double replacement technique (Section 2.1.2) (69, 228), more recent knock-in experiments have been performed using conventional gene targeting approaches (Fig. 6A), often in combination with the Cre-loxP system to remove the selection cassette (detailed in Section 2.2; Fig. 6B) (76). As shown in Fig. 6, such knock-in vectors are essentially replacement targeting vectors containing the transgene and a positive selectable marker and are designed such that after homologous recombination, the transgene is transcriptionally regulated by the endogenous promoter of the locus. This method is based on the production of a fusion protein between the endogenous and knocked in products. However, if a fusion product is undesirable (such as in the case of secreted proteins), then 1) the transgene can be knocked into the untranslated region upstream of the endogenous translational start site (although splicing may occur around the inserted transgene) or 2) an internal ribosome entry site (IRES) element (181) can be used to generate a bicistronic mRNA [the encephalomyocarditis virus IRES is most commonly used in mammalian cells (110)] (113).

2.1.5) Generating mosaic mice. As discussed above, the most widely used analysis of gene function in mice is through the generation of a null allele, which allows the generation of mice homozygous for the desired mutation. However, it is important to note that such a strategy can only be used to assess the earliest function of the gene, as mutational loss of a gene essential for embryogenesis can cause lethality and hence prevent examination of its role in adult mice. To overcome this limitation, conditional mice (see Section 2.1.3) or mosaic mice can be generated. Mosaics are individuals containing cells of more than one genotype. This allows for the generation of mice containing mutant homozygous cells (which would be lethal in a homozygous mouse) on an otherwise heterozygous background (i.e., the genetic composition of the vast majority of cells in the mouse is heterozygous; hence, the animal is viable). Conversely, mosaic mice have been used to study the function of c-src in osteoclasts and other cell types; Src-deficient mice (src<sup>−/−</sup>) crossed with transgenic mice expressing src under the control of tartrate resistant acid phosphatase promoter (a gene highly expressed in osteoclasts) allowed the generation of src<sup>−/−</sup> mice that express Src in osteoclasts, thereby allowing the analysis of Src function in other tissues in mice free from the morbidity of osteopetrosis (211).

Genetic mosaics can be generated when mitotic recombination between homologous chromosomes occurs during the G2 phase of the cell cycle and the recombinant chromatids segregate to different daughter cells (X segregation). Recombinant chromatids produced in G2 can also segregate to the same daughter cell, and the nonrecombinant chromatids to the other daughter
cell (Z segregation). It is X segregation that is useful for genetic mosaic analysis, because it produces clones of homozygous mutant daughter cells from heterozygous mothers (whereas Z segregation produces daughter cells that are phenotypically indistinguishable from the parent cell or from cells produced by G1 recombination).

There are numerous approaches to generate mosaics in mice (see Fig. 7). The most common is the use of site-specific Cre/loxP technology (described in Section 2.2) in a conditional knockout strategy, in which mosaics are generated in specific tissues by the activity of Cre on a loxP-flanked allele (Fig. 7A) (74). Alternatively, homozygous mutant ES cells can be produced in culture, and these can be used to generate chimeras. For example, direct selection of loss of heterozygosity (LOH) events in neo cassette-containing ES cells can be achieved using high concentrations of G418 (which causes chromosomal nondisjunction followed by reduplication of the remaining chromosome; Fig. 7B) (154).

The mechanism that duplicates the targeted neo cassette (thereby providing enhanced G418 resistance) also allows mutations on the same chromosome to become homozygous. Homozygous clones can also be generated by repetitive neo targeting, followed by puro targeting of the second allele. Another strategy (Fig. 7C), based on the principles of the Flp-FRT site-specific recombination system in Drosophila (22), uses the spatial and temporal regulation of Cre expression, which can be used to trigger mitotic recombination (with recombination rates of $10^{-4} - 10^{-2}$ in ES cells transiently transfected with Cre) (134). Alternatively, Bloom's mice can be used, as they possess a genetic background that displays increased rates of mitotic recombination (Fig. 7D). Bloom's mice (Blm$^{-/-}$) are deficient in the RecQ DNA helicase/Blm protein (140), analogous to the human condition, Bloom's syndrome [a rare cancer-prone disorder in which the cells of affected persons have a high frequency of somatic mutation and genomic instability due to high frequency of sister chromatid ex-

Fig. 6. Knock-in strategies. A: knock-in of a transgene into the target locus in-frame with the target. The knock-in vector (black line) contains the transgene, polyadenylation signal (pA), a positive selection cassette (+), and DNA (green line; exons shown as boxes) that has homology to the target locus (gray line; exons shown as boxes). Using this vector, the transgene must be inserted in-frame with the upstream exon. B: knock-in of a transgene using a splice acceptor sequence and floxed selection cassette. The knock-in vector (black line) contains the splice acceptor signal (SA), transgene, polyadenylation signal (pA), a positive selection cassette (+) flanked by two loxP sites (black triangles), and DNA (green line; exons shown as boxes) that has homology to the target locus (gray line; exons shown as boxes). The use of a splice acceptor allows the transgene to be in-frame with the target locus irrespective of where it integrates within the locus. The use of loxP sites to flank (“floxed”) the selection cassette allows for excision of the cassette from the locus upon exposure to Cre recombinase. X represents recombination between the vector and the genome.
change (68)]. Bloom’s mice and ES cells exhibit greatly elevated rates of mitotic recombination, allowing homozygous mutant daughter cells to be generated from heterozygous mothers both in vitro and in vivo. The rate of mitotic recombination in Blm-deficient cells is sufficient to enable homozygous mutant clones to be recovered for any gene, regardless of its location in the genome (140).

2.2) Site-Specific Recombination

2.2.1) Recombinases: Cre and Flp. The simplest site-specific recombination systems are those composed of a recombinase enzyme and its target sequence. These systems allow for the deletion, insertion, inversion, or translocation of specific regions of DNA. Two such recombinase systems (both members of the λ-integrase superfamily) include the Cre-loxP system from the bacteriophage P1 and the Flp-FRT system from the budding yeast S. cerevisiae. Both Cre and Flp cleave DNA at a distinct target sequence (Fig. 8A) and ligate it to the cleaved DNA of a second identical site, to generate a contiguous strand. The orientation of these target sites relative to each other directs the type of modification catalyzed by the recombinase (detailed in Fig. 8B). When compared with one another, Cre appears more effective in recombining a transgene array in vitro and in vivo than Flp (198). However, mutated versions of Flp, such as the temperature-sensitive FlpL (30) and the enhanced Flp (Flpe) (31), show improved activity over wild-type Flp. Indeed, the generation of high-efficiency Flpe deleter mice [the ROSA26 locus was targeted to create a mouse strain with generalized expression of FLPe, termed the FLPeR ("flipper") strain (56)] show that Flpe is an alternative to Cre-loxP (199). In general, Cre and Flpe are used in situations requiring high-efficiency recombination (30–31, 56, 245); Flp and FlpL are used when tight regulation is needed (255). Currently, the most widely used site-
specific DNA recombinase system in both ES cells and mice is the Cre-loxP system.

2.2.2) Cre: advantages and disadvantages. Advantages of the Cre recombinase system include its 1) simplicity (no cofactors required), 2) fidelity (recombination is carried out such that there is no loss or gain of nucleotides), 3) small 34-bp target site (which does not perturb the surrounding genes when positioned in chromosomal DNA), and 4) broad utility (acting on supercoiled, relaxed, or linear DNA substrates; functioning over large megabase distances; and functional in a wide range of cell types) (reviewed in Refs. 53,
Accordingly, Cre-mediated recombination is 4-OHT de-modified (4-OHT), has been added to Cre (29, 59, 65, 90). Activation of tamoxifen or its derivative 4-hydroxytamoxifen can cause chromosomal rearrangements/aberrations and increased number of sister chromatid exchanges when expressed at very high levels (137, 210, 220), most likely by way of pseudo-loxP sites. High level of Cre expression has also been shown to reduce cell proliferation in mouse embryonic fibroblasts (MEFs) (48, 137, 220) and is speculated to be involved in causing cell-cycle arrest (1).

2.2.4 Systems for delivery of Cre in vitro and in vivo.

The methods used to introduce Cre into ES cells are numerous. The usual in vitro introduction method for transient expression or stable integration is calcium precipitation or electroporation [which can reach as high as 70% efficiency (163)]. Alternative methods of delivery include infection of ES cells with a recombinant Cre adenovirus (2, 5, 259). Adenoviral transfection of Cre results in transient Cre expression, but mosaic mice can be produced, which transmit the targeted allele to their offspring with high frequency (106). The method of introduction of Cre in vivo depends heavily upon the specific application. The simplest solution to achieve overall excision in the developing embryo is crossing of the floxed mice with transgenic mice carrying the Cre recombinase gene (a general Cre expresser transgenic line) (119, 206). Another variation of this approach is pronuclear injection of a Cre expression vector [with the resultant transient Cre expression inducing recombination during preimplantation development (8)] or injecting Cre RNA or protein.

2.3) Chromosome Engineering

2.3.1) Chromosomal rearrangements. A large array of mice containing chromosomal rearrangements (deletions, inversion, duplications, and translocations) have been generated by exposure to chemical mutagens [cyclophosphamide, ethylene oxide, chlorambucil, and hydrophilic peptides from the Kaposi fibroblast growth factor (FGF-4) (94) or basic peptides derived from HIV-TAT (98, 180) have been produced to promote cellular uptake of recombinant Cre. Recombination has been observed in a variety of cultured cell types and in specific tissues examined in mice following intraperitoneal administration. This new cell-permeable form of Cre will likely open up new opportunities for genetically manipulating cells both in vitro and in vivo (40).

In certain applications, the retroviral or adenoviral delivery systems have been used. For example, after intranasally administered adenoviral Cre, mice expressing a mutated K-ras oncogene gene placed downstream of a floxed transcriptional termination stop element developed lung tumors by 2–16 wk of age (due to Cre-mediated removal of the floxed stop element, and hence expression of the mutant K-ras) (95). An alternative is the combination of avian retrovirus with the TVA (EnvA receptor) delivery system, which provides the possibility of viral delivery of Cre in a tissue- or cell-specific manner, as there is a requirement for the expression of a special avian receptor (tv-a) in the mouse cells from a cell-type-specific transgene to make these cells susceptible to infection (85). More recently, an adeno-associated virus [used to transfer foreign genes into the adult and neonatal central nervous system in animals (23, 136, 146)] expressing Cre recombinase has been shown to mediate extensive in vivo recombination in neural cells of defined brain regions in the mouse (108).
and N-ethyl-N-nitrosourea (ENU) or radiation (X-rays) (47, 124, 192, 205, 234, 237). However, although these mutagens have generated some valuable mouse models for human diseases, such as the mouse model of trisomy 21 (47, 192), their usefulness for inducing rearrangements is limited by the fact that the endpoints of the induced rearrangements cannot be predetermined (see SECTION 3, below, for details on mutagenesis strategies). To this end, gene targeting-based strategies have been developed to introduce defined chromosomal rearrangements into the mouse genome by engineering them in ES cells using the Cre/loxP site-specific recombination system. This technology, known as “chromosome engineering,” has successfully generated numerous mouse models that accurately recapitulate human chromosomal rearrangements, such as the chromosomal deletion within band 22q11 (del22q11) causing DiGeorge syndrome (130–131), the paternal deficiency of chromosome 15q11-q13 causing Prader-Willi syndrome (244), the translocation between chromosomes 8 and 21 (t(8;21)) found in acute myeloid leukemia (AML) (32), and the reciprocal translocation between chromosomes 9 and 11 (t(9;11)q22; q23) associated with acute leukemia (42).

### 2.3.2 Chromosome engineering technology

#### 2.3.2.1 Selection of endpoints

Generating chromosomal rearrangements in the mouse involves the sequential insertion of two targeting vectors into two separate loci in the ES cell genome. Thus an important decision is to define the two endpoints of the rearrangement. Endpoints can be chosen at any genomic region. For example, SSLP microsatellite markers make useful endpoints as they have been genetically mapped (see the Whitehead Institute STS Physical Map of the Mouse at [http://www-genome.wi.mit.edu/cgi-bin/mouse/index](http://www-genome.wi.mit.edu/cgi-bin/mouse/index)). Numerous SSLP markers have been successfully used as the endpoints for engineering chromosomal deletions on mouse chromosome 11 (286). In addition, the high-resolution mapping information available for the mouse genome (see the MGSC Ensembl Mouse Genome Server at [http://www.ensembl.org/Mus_musculus/](http://www.ensembl.org/Mus_musculus/)) means that genes of known chromosomal location may also be used as endpoints. Genes used as endpoints have included the epidermal growth factor receptor (Egfr) gene (242), the amyloid precursor protein gene (125), the myeloperoxidase (MPO) gene (10), Wnt3, p53, and Hoxb9 (190, 286), p63 (149), and the HoxB cluster (148), to name a few.

#### 2.3.2.2 Generating the chromosomal rearrangement

Once the endpoints have been selected, the first step involves the targeting of an insertion vector containing a loxP site, a positive selection cassette (e.g., neomycin), and one of two complementary, but nonfunctional, fragments of the Hprt gene into the desired locus (selected endpoint) of the ES cell genome (see Fig. 9A) (285). ES cell clones with a loxP site targeted to a first endpoint can be identified by positive selection and Southern blot analysis. The second step involves the
targeting of a second insertion vector containing a loxP site, a different positive selection cassette (e.g., puromycin), and the complementary fragment of the Hprt gene into the second endpoint (see Fig. 9B). The generation of these targeting vectors is detailed below. The doubly targeted ES cell is then transiently transfected (electroporated) with a vector expressing Cre, which facilitates recombination between loxP sites such that the intervening DNA is deleted. Commonly used Cre-expression vectors include pOG231 (171), pTurboCre (GenBank accession no. AF334827), pCrePAC (238), and pBS185 (125). The type of chromosomal rearrangement derived from the double-targeted ES cells is determined by the loxP configuration (reviewed in Fig. 8B). As shown in Fig. 10A, loxP sites in the same orientation generate a chromosomal deletion (or duplication event; not shown). Alternatively, loxP sites in the opposite orientation generate a chromosomal inversion (Fig. 10B). The generation of chromosomal translocations is discussed below (in Section 2.3.4).

This Cre-mediated recombination event can be selected for in culture because a functional Hprt cassette is reconstituted, which confers resistance to the drug hypoxanthine-aminopterin-thymidine (HAT). Thus, using selection in HAT, Southern blot analysis, and fluorescent in situ hybridization analysis (FISH), ES cell clones carrying the desired chromosomal rearrangement can be identified. As with traditional gene targeting strategies shown in Fig. 1, these ES cells are injected into mouse blastocysts to generate chimeras, from which the progeny that carry the engineered chromosomes are derived.

2.3.2.3) GENERATION OF GENE TARGETING VECTORS FOR CHROMOSOME ENGINEERING. Gene targeting vectors, such as those shown in Fig. 9, can be generated in the conventional way by sequentially inserting various genetic components into a plasmid construct (133, 190). However, the “two library” system (285) greatly reduces the number of cloning steps required for generating gene targeting vectors. This system is composed of two complementary libraries of pre-made gene tar-

---

Fig. 10. Generation of defined chromosome rearrangements. The endpoints (red stars, X and Y) selected for the interval of genomic DNA to be rearranged are modified using two consecutive steps of gene targeting in the same ES cell. Both steps involve a positive selection scheme using the Hprt minigene to identify clones containing rearrangements induced by Cre recombinase. The first step involves targeting of a linearized insertion vector containing a loxP site (black triangle), a nonfunctional 5'-Hprt cassette, and a neomycin resistance cassette (N), to endpoint X. The second step involves targeting of a linearized insertion vector containing a loxP site, a nonfunctional 3'-Hprt cassette, and a puromycin resistance cassette (P), to endpoint Y. The final step uses Cre recombinase to catalyze site-specific recombination between the loxP sites. A: if the loxP sites are in the same orientation, then the intervening region is deleted. B: if the loxP sites are in the opposite direction, then a chromosomal inversion occurs. Cre-mediated recombination to unite the 5' and 3' cassettes restores Hprt activity, making the ES cells resistant to HAT-containing medium (HATr).
targeting insertion vectors. The 5′-Hprt library was generated by cloning a genomic library into a vector backbone that contained the 5′-Hprt cassette, a loxp site, and neomycin (PGKNeo bpA) as a positive selectable marker. The complementary 3′-Hprt library was generated by cloning a genomic library into a vector backbone that contained the 3′-Hprt cassette, a loxp site, and puromycin (PGKpurop bpA) as a positive selectable marker. In addition, both of these libraries are equipped with genes encoding visible coat color markers. The tyrosinase minigene (Ty) has been used to “tag” transgenics with a visible pigment marker in albino mice (176, 275), and the K14-Agouti transgene (Ag) uses the keratin-14 promoter to constitutively express agouti (114). Mice carrying the Ty transgene have a grayish coat on an otherwise albino background, and the Ag confers a “butterscotch” coat color in black agouti or non-agouti mice (285).

To make a defined rearrangement between any two desired loci, clones for each endpoint are first isolated from the appropriate library. Linearization of the construct within the genomic insert generates a gene targeting insertion vector that integrates at the target locus. Targeting can be assessed by Southern blot analysis using a DNA fragment that was removed from the genomic insert before gene targeting, by an external probe, or by PCR amplification using primers specific for the gap and vector. An additional feature of this gene targeting library system is that clones isolated from these libraries can also be used for analyzing single gene function via the knockout (null allele) approach, as demonstrated by targeted disruption of the p63 locus (149).

2.3.3) Chromosomal deletions. 2.3.3.1) USES FOR THE ENGINEERING OF CHROMOSOMAL DELETIONS. The chromosomal engineering of deletions can be used to identify TSGs without prior knowledge of the gene function. For example, mice possessing a deletion encompassing a putative TSG should exhibit increased tumorigenesis, as they only possess a single copy of a TSG, and tumor-specific loss or inactivation of the remaining allele can be used to clone the causative gene. Chromosome engineering can also be used to generate mouse models of human microdeletion syndromes. For example, mice heterozygous for a 1.2-Mb deletion between Es2 and Ufd1l on mouse chromosome 16 show cardiovascular abnormalities resembling those found in DiGeorge syndrome patients (130). When these mice were crossed to a strain harboring a duplication of the same region, the mutant phenotype was functionally rescued, indicating that the defects seen in DiGeorge patients are due to haploinsufficiency (loss of one functional copy of a gene). Thus the potential embryonic lethality of a particular deletion can be functionally determined by assessing the ability of the duplicated allele to rescue the embryonic lethality of the deletion (as mice containing both the deleted and duplicated alleles are genetically balanced).

The major factor limiting the generation of deletions in ES cells is the size of the rearranged interval, with deletions >22 cM leading to ES cell lethality or severe growth disadvantage of the cells in culture (286). Although Cre/loxP recombination has been shown to occur over such large distances, clones can emerge from these experiments that have undergone a compensatory genetic change, such as a chromosomal duplication (286). X-ray- and UV-induced mutagenesis have also been successfully performed to generate deletion complexes in ES cells. Radiation-induced deletions can be localized and made selectable by targeting a vector that carries a negative selection cassette to a predetermined locus; for example, homologous recombination can be used to insert a herpes simplex virus thymidine kinase (HSV-th) gene into a specific locus, followed by irradiation and then selection for clones that have deleted the tk cassette (240, 276). In addition, deletion alleles in ES cells of several centimorgans have been successfully transmitted through the germ line (190, 276).

2.3.3.2) GENERATION OF NESTED CHROMOSOMAL DELETIONS. An extension of the chromosome engineering strategy is the generation of nested chromosomal deletions, a series of variably sized, overlapping deletions surrounding a predetermined genomic locus. This strategy has been used to identify that haploinsufficiency of the Tbx1 gene contained within the 1.2-Mb deletion interval of DiGeorge mice was responsible for the aortic arch defects seen in these patients (131). Furthermore, if the genomic locations of the endpoints are known, then nested deletions can be extremely useful for mapping novel recessive mutations (278). However, to get around the task of having to generate targeting vectors for the nested endpoints, retroviral integration of a second loxp site and selection cassette can be used (235). More specifically, deletion complexes can be anchored to a predetermined location in the genome by targeting the 5′-Hprt-loxp cassette. The 3′-Hprt-loxp cassette is then randomly inserted into the ES cell genome by retrovirus-mediated integration, generating a library of ES cell clones with the same targeted endpoint and a collection of random endpoints (Fig. 11). This method has been used to generate nested deletions extending from a few kilobases to several megabases at the Hprt locus, also on chromosome 11 (235). Electroporation has also been used to randomly insert the second loxp site and selection cassette into the ES cell genome (122). However, insertion by electroporation increases the risk of genomic rearrangements occurring at the insertion site, and tandem repeats of a vector may be introduced into the insertion site (although these should be reduced to a single locus by the activity of Cre on a head-to-tail concentrate) (278). Endpoints generated by random insertion can be defined by cloning the genomic DNA that flanks the deletion endpoints and by mapping these junction fragments onto a physical map of the region. Nested deletions can also be efficiently generated by irradiation (116, 240, 276), although they require additional extensive characterization to define each deletion interval (278).

2.3.4) Chromosomal translocations. Chromosomal translocations are involved in the genesis of many types of human tumors, often as a result of the abnor-
Fig. 11. Nested chromosomal deletions induced with a retroviral vector. The first deletion endpoint is fixed by targeting the 5′-Hprt cassette, the neomycin resistance gene (N), and a loxP site (black triangle) to a predetermined locus. The second step involves the integration of the 3′-Hprt cassette, the puromycin resistance cassette (P), and a second loxP site into a random site in the ES cell genome, using a recombinant retroviral vector. Here is depicted G1 recombination events from retroviral integration of the Hprt gene and hence survival in HAT medium, for ES cell clones carrying the recombinant chromosomes. The nested deletions can be identified from a pool of HAT-resistant clones on the basis of their sensitivity to G418 and puromycin. LTR, long terminal repeat.

mal expression of cellular oncogenes or by creating novel fusion genes (152, 189). Although mouse models for several human leukemias have been established by tissue-specific expression of fusion proteins transgenes (138) and knock-in constructs expressing a fusion protein under the control of the appropriate endogenous promoter (44), they fail to recapitulate the situation found in human translocation-induced tumors. For example, in these mice the fusion gene is present from the inception of embryogenesis, whereas in humans the chromosomal translocation is believed to occur at later stages (172, 274). In addition, these mice possess only one fusion protein, whereas in balanced chromosomal translocations two fusion proteins are generated (186). Thus to accurately recapitulate chromosomal translocations found in human tumors, chromosome engineering is most effective (224, 249). Numerous mouse chromosomal translocations with predetermined breakpoints have been created using this strategy, including models of human leukemia-associated translocations such as t(8;21)(q22;q22) (32) and t(9;11)(p22;q23) (42).

Using the Cre-loxP system, chromosomal translocations are generated using the same techniques as for deletions, duplications and inversions (see Section 2.3.2), except the loxP sites, orientated in the same direction relative to their respective centromeres, are targeted to nonhomologous (different) chromosomes. If the loxP sites are in opposite directions, then recombination will result in acentric chromosomes (without a centromere) and dicentric chromosomes (containing two centromeres), which will be unable to survive. Thus, to prevent the generation of anacentric and dicentric chromosomes, only pairs of genes with the same transcriptional orientations relative to their centromeres can be engineered to produce fusion proteins (278). To generate a fusion protein from a chromosomal translocation, the targeting vectors need to be designed to allow the two genes to be linked through their introns (with the loxP site embedded in the breakpoint), so RNA splicing will generate an in-frame fusion mRNA and protein (278). To induce the translocation, mice carrying both targeted endpoints can be crossed with transgenic mice expressing a tissue- and/or temporal-specific Cre. In addition, this strategy circumvents the problem of transmitting the translocation through the male germ line, as the presence of chromosomal translocations in male germ cells can cause infertility (141).

2.3.5) Balancer chromosomes. A balancer chromosome is one containing an inverted region(s). Balancer chromosomes can be generated using chromosomal engineering technology, as chromosomal inversions can be generated in mouse ES cells by successive gene targeting of a loxP site to the two endpoints, followed by Cre-mediated recombination between the two loxP sites (Fig. 10B) (284). In addition, the inversion can be marked with a dominant marker (such as the dominant K14-Agouti coat-color gene) so that progeny carrying the balancer can be readily identified.

As inversions suppress crossing over during mitotic recombination (the genetic exchange of information between sister chromatids), balancer chromosomes are genetic reagents that can be used for stock maintenance (to maintain the integrity of mutagenized chromosomes). Balancer chromosomes can also be used for large-scale mutagenesis screens (see Section 3.1.2). For example, in intercrosses between siblings that have inherited the balancer chromosome and a mutagenized chromosome, absence of non-balancer-carrying progeny indicates the presence of one or more recessive lethal mutations on the mutagenized chromosome (284, 278). Indeed, the first mouse balancer chromosome was constructed on chromosome 11, to facilitate...
the isolation of ENU-induced recessive mutations on mouse chromosome 11 (http://www.mouse-genome.bcm.tmc.edu/ENU/ENUHome.asp) (284). This balancer chromosome is based on a 24-cM inversion between the Trp53 gene and the Wnt3 gene (in addition, a coat color marker, K14-Agouti, has been inserted into the mutated Wnt3 locus). Balancers will also be useful to facilitate analysis and maintenance of other types of knockouts.

2.3.6) Homologous recombination in *E. coli*. To simplify the generation of knockout constructs, recombining technologies have been developed. This form of DNA engineering utilizes methods based on homologous recombination in *E. coli* that enable large segments of genomic DNA in BACs or P1 artificial chromosomes (PACs) to be modified and subcloned, without the need for restriction enzymes or DNA ligase.

2.3.6.1) *RecA* homologous recombination. Several recombination pathways have been identified in *E. coli*, including the RecA pathway. A homologous recombination-based system allowing modification of BACs in recombination-deficient *E. coli* is the temperature-sensitive shuttle-vector-based system (75, 170). This temperature-sensitive plasmid replicates in cells growing at the permissive temperature (30°C) but is lost in cells growing at the restrictive temperature (42–44°C) because its origin of replication cannot function (79). Introduction of the *E. coli* RecA gene into the temperature-sensitive shuttle vector allows the RecA host *E. coli* strain containing the BAC to become competent to perform homologous recombination of the resident BAC in vivo (150, 272). Using this method, transgenic mice have been generated by pronuclear injection of the modified BAC, and germ-line transmission of the intact BAC obtained (272).

2.3.6.2) ET recombination. Another recombination pathway in *E. coli* is the RecBCD pathway involved in repairing double-strand breaks (225). RecBCD unwinds and degrades DNA to generate 3’ single-stranded DNA (ssDNA) tails, which are used by RecA to initiate recombination. However, *E. coli* possessing wild-type RecBCD degrade the introduced linear DNA molecules before recombination has proceeded; therefore, to restore recombination activity, a suppressor mutation that activates expression of a nuclease that produces 3’ overhangs may be used (117). RecBCD can be inactivated by 1) the sbcA mutation, which removes a repressor for the endogenous lambid prophage, Rac, which in turn induces the expression of *recE* and *recT* [two Rac genes that encode homologous recombination functions (135)], or 2) the gam protein of *λ* phage [in the presence of gam, *λ* phage-encoded recombination function stimulates homologous recombination by the *λ red* genes (156, 188)].

Engineering DNA by homologous recombination mechanisms involving the RecE/RecT and Redo/Redβ proteins in *E. coli* is termed “ET recombination” (also known as Red recombination and lambda-mediated cloning). ET recombination was pioneered by Stewart and colleagues (280), who showed that a PCR-amplified fragment of linear dsDNA, flanked by short regions of homology (42 bp) to a plasmid or BAC can be efficiently targeted to a plasmid or BAC by electroporating the dsDNA into *recBC sbcA* strains (see http://www-db.embl-heidelberg.de/dfs/servlet/de.embl.bk.wwTools.GroupLeftEMBL/ExternalInfo/stewart/ETcloning-textonly.html) (280). As shown in Fig. 12, ET recombination involves two steps: first, the amplification of a fragment of DNA by PCR with flanking regions of homology, plus the introduction of phage recombination functions into a BAC-containing bacterial strain; and second, the transformation of the cassette into the bacterial cells that contain a BAC and recombinase functions (the bacterial cells generate a recombinant in vivo, and detection of the recombinant is done by selection, counter-selection, or by direct screening). ET recombination allows alterations such as point mutations, sequence insertions and/or deletions to be carried out at any position on a target DNA molecule and has greatly simplified the generation of

---

**Fig. 12. Recombineering steps to generate a bacterial artificial chromosome (BAC) recombinant.** The first step in recombineering is to amplify the cassette of interest (such as a selectable marker or reporter gene; gray box) by PCR, using primers containing 40–50 bp of homology to the target site on the BAC (white box). The next step is to introduce phage recombination functions into a BAC-containing bacterial strain or introduce a BAC into a strain that carries recombination functions. The cassette can then be transformed into cells that contain the BAC and recombination functions, to allow generation of a recombinant clone [which are those containing the PCR-amplified DNA cassette inserted into the target gene (black box)]. Recombinants can be detected by selection, counterselection or by direct screening (colony hybridization).
transgenic and knockout constructs (reviewed in Refs. 43, 160).

Since its conception in 1998, many variations of phage-based *E. coli* homologous recombination (recombineering) systems have been developed (4, 121, 156, 158, 159, 161, 164, 277, 280, 281). Examples include the use of a temperature-dependent repressor that tightly controls prophage expression; recombination functions can thus be transiently supplied by shifting cultures to 42°C for 15 min (277). Another system, termed the univector plasmid-fusion system (UPS), uses Cre-lox site-specific recombination to catalyze plasmid fusion between the univector (a plasmid containing the gene of interest) and host vectors containing regulatory information (133). More recently, a new method termed “recombination cloning” (REC), uses as little as 80 bp of total sequence homology to screen for a specific gene from a genomic library (in plasmid or phage form) (282). The advances which REC has over existing recombineering technologies include: 1) improved vector design (AKO-1), 2) enhanced recombination using λ red and gam (which together with recA show a 100-fold enhancement in recombination frequencies compared with those obtained using recBC mutants), and 3) the in vivo generation of recombination substrates (which also improves in the recombination efficiency) (282).

3) MUTAGENESIS TECHNOLOGIES

The low frequency at which spontaneous mouse mutants occur (5 × 10⁻⁶ per locus) makes them unfeasible for use in teasing out an entire genetic pathway (230). Furthermore, as spontaneous mutations may be caused by small base changes in coding sequences, retroviral insertions, or by other events such as deletions, it can be very hard to identify and characterize the molecular lesion. Many mouse mutagenesis strategies have therefore been developed, and each generates mutations of a different molecular nature and at varying frequencies.

3.1) Phenotype-Driven Mutagenesis

In gene targeting experiments, the gene is known at the beginning of the experiment, and a phenotype can sometimes be predicted. In contrast, phenotype-driven mutagenesis starts with a specific phenotype, and only at the end is the responsible gene identified. In phenotype-driven screens, the animal is exposed to a mutagen that acts randomly on the entire genome, and the resulting mutants are then subjected to a wide array of tests including those for blood chemistry, lipid profile, hematology, and for more obvious features such as physical deformities. A large number of animals are screened to identify individuals that display the specific phenotype of interest. Following the isolation of an animal with the desired trait, a candidate gene approach may be used to identify the mutated gene, which is confirmed to be heritable by test matings. There are several methodologies employed in phenotype-driven mutagenesis screens.

3.1.1) Irradiation. X-rays induce mutations at a frequency of 13–50 × 10⁻⁵ per locus and create chromosomal rearrangements ranging from simple deletions and inversions to complex translocations (35, 203, 230). Although X-ray mutagenesis causes chromosomal rearrangements, which can provide a molecular landmark for identifying the affected gene(s), several genes are often affected by these chromosomal rearrangements, and as a result it is difficult to dissect individual gene function.

3.1.2) Chemical mutagenesis. 3.1.2.1) ENU-INDUCED MUTATIONS. ENU is the most potent mutagen in mice (204), with a mutagenesis frequency of 1.5 × 10⁻³ per locus (230). It is an alkylating agent that transfers its ethyl group to oxygen or nitrogen radicals in DNA, resulting in mispairing and base pair substitution if not repaired. In contrast to other chemical agents such as chlorambucil, which induces deletions, inversions, or translocations and often involves more than one gene (61, 205), ENU primarily produces point mutations (most commonly AT-to-TA transversions and AT-to-GC transitions) (187). ENU affects germ cells, particularly spermatogonial stem cells, and induces point mutations, leading to a wide range of missense, nonsense, and splice-site mutations (104). In addition to loss-of-function mutations, ENU mutagenesis can also result in gain-of-function mutations and hypomorphic alleles (179, 195–196, 253).

3.1.2.2) ENU MUTAGENESIS STRATEGIES. ENU mutagenesis is a powerful tool (“genetic reagent”) for systematic analysis of the mouse genome (reviewed in Refs. 13, 104, 197). ENU mutagenesis in the mouse has been used to obtain multiallelic series of single genes at complex loci to further define function. For example, the five ENU-induced alleles of the quaking (*qk*) locus (*qk*<sup>1</sup>–<sup>5</sup>, *qk*<sup>kt1</sup>, *qk*<sup>kt2</sup>, *qk*<sup>kt3/4</sup>, and *qk*<sup>e5</sup>), each show different phenotypes, making them a valuable resource for fine structure/function studies (45–46, 104, 217). ENU mutagenesis can also aid in dissecting biochemical or developmental pathways (24, 100–102, 107, 195). ENU has also been used to generate new recessive mutations in a defined region of the mouse chromosome, such as saturation germ-line mutagenesis of the murine t region (216, 217).

3.1.2.3) Genome-Wide Screens for Dominant or Recessive Mutations. A number of different strategies can be used in ENU-mutagenesis experiments to screen for phenotypically interesting mice, including genome-wide screens for dominant or recessive mutations, and sensitized screens (some examples are given in Table 2). Mutagenesis is initiated by injecting male mice with ENU, resulting in the generation of mutations in their spermatogonial stem cells. These mice then enter a latent period during which they are sterile as the gonads are repopulated with new mutant sperm. After recovery these mice are mated with wild-type females to produce a large number of F1 offspring. Each of the F1 animals carries a unique complement of altered alleles. The F1 animals are either examined directly (dominant screen) or enter various breeding regimes to generate recessive phenotypes. The heritability of a
REVIEW

Table 2. Large-scale dominant and recessive ENU-induced mutation screens

<table>
<thead>
<tr>
<th>Location</th>
<th>Genetic Screen</th>
<th>Phenotypes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse Mutagenesis Programme (MRC, UK) (<a href="http://www.mgu.har.mrc.ac.uk/mutabase/">http://www.mgu.har.mrc.ac.uk/mutabase/</a>)</td>
<td>dominant (genome wide)</td>
<td>neurological</td>
<td>200</td>
</tr>
<tr>
<td>GSF-National Research Center for Environment and Health (Germany) (<a href="http://www.gsf.de/sg/group/enm-mouse.html">http://www.gsf.de/sg/group/enm-mouse.html</a>) and <a href="http://www.ukl.uni-freiburg.de/mmz/homede.htm">http://www.ukl.uni-freiburg.de/mmz/homede.htm</a></td>
<td>recessive (mouse Chr 13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse Genome Center (BCM, USA) (<a href="http://www.mouse-genome.bcm.tmc.edu">http://www.mouse-genome.bcm.tmc.edu</a>)</td>
<td>dominant (genome wide)</td>
<td>clinical-chemical;</td>
<td>57, 86, 87</td>
</tr>
<tr>
<td>The Jackson Laboratory (Univ. of Penn., USA) (<a href="http://www.jax.org/nmf/documents/overview.html">http://www.jax.org/nmf/documents/overview.html</a>)</td>
<td>recessive (mouse Chr 4, 11)</td>
<td>dysmorphology; immunology; behavior; lysosomal defects;</td>
<td></td>
</tr>
<tr>
<td>Oak Ridge National Laboratory (USA) (<a href="http://bio.lsd.ornl.gov/mouse/">http://bio.lsd.ornl.gov/mouse/</a>)</td>
<td>recessive (mouse Chr 5)</td>
<td>gene function; development; human disease models;</td>
<td>103</td>
</tr>
<tr>
<td>McLaughlin Research Institute (USA) (<a href="http://www.montana.edu/wwwmri/enump.html">http://www.montana.edu/wwwmri/enump.html</a>)</td>
<td>recessive (mouse Chr 7)</td>
<td>neurological; behavioral</td>
<td>209</td>
</tr>
<tr>
<td>Medical Genome Centre (ANU, Australia) (<a href="http://csmr.anu.edu.au/group_pages/mge/MedGenCen.html">http://csmr.anu.edu.au/group_pages/mge/MedGenCen.html</a>)</td>
<td>dominant (genome wide)</td>
<td>gene function; development</td>
<td>196</td>
</tr>
<tr>
<td>Mouse Functional Genomics Research Group (RIKEN, Japan) (<a href="http://www.gsc.riken.go.jp/Mouse/">http://www.gsc.riken.go.jp/Mouse/</a>)</td>
<td>recessive (genome wide)</td>
<td>immunity; aging</td>
<td></td>
</tr>
<tr>
<td>Centre for Modelling Human Disease (Ontario, Canada) (<a href="http://www.cmhd.ca/sub/mutagenesis.htm">http://www.cmhd.ca/sub/mutagenesis.htm</a>)</td>
<td>dominant (genome wide)</td>
<td>late-onset (tumorigenesis; and senescence-related phenotypes)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>recessive (genome wide)</td>
<td>hematological disorders; kidney disease; limb patterning defects; vasculogenesis and tumorigenesis; learning, memory, and behavior; autoimmune disease; inflammation; bone defects; diabetes and glucose intolerance; cardiopulmonary function; cerebral defects</td>
<td></td>
</tr>
</tbody>
</table>

ENU, N-ethyl-N-nitrosourea.

trait is confirmed by test breeding to ensure that it is dependent upon an ENU-induced genetic lesion.

Most of the ENU screens carried out in recent years have been dominant screens, due to the simplicity of the logistics of the mutagenesis and breeding protocols and the success in the recovery of mutations through this approach. However, dominant alleles are a unique type of mutation that covers the gamut from haploinsufficiency to neomorphs, and a dominant phenotype sometimes makes it difficult to relate the biological function of the mutated gene to its true physiological role. Furthermore, as many human genetic disorders are caused by recessive mutations, it is clear that dominant ENU screens will be unable to uncover all the important disease-causing genes. However, unlike dominant screens, recessive screens are very technically and logistically demanding. The genetic reagents generated by chromosomal engineering, such as deletions [see Section 2.3.2, above (Ref. 287)], can aid in mutation screening by exploiting the concept of segmental haploidy (197). For example, ENU-treated male mice (potentially carrying a recessive mutation) are mated to wild-type females to generate G1 males, who are then mated to females carrying deletions. The ENU-induced mutation can be identified in the second generation if it lies within the deletion interval. In addition, if the G1 males are mated to multiple females carrying contiguous deletions, then a large chromosomal region can be scanned for mutations (103). For regions of the genome where it is impossible to use deletions in region-specific screens because haploid insufficiency has a deleterious effect on cell liability or the fertility of the mice, balancer chromosomes (engineered inversions) can be used (see Section 2.3.5, above) (197). Balancer chromosomes carry both dominant and recessive markers, such that mice homozygous for the balancer will die, whereas heterozygous mice show visible characteristics such as coat coloration (284). When used in combination with chemically mutagenized mice, the balancer chromosome can be used to isolate and maintain a mutation, thus significantly reducing the time and effort involved in performing ENU screens (103, 104).

An interval of mouse chromosome 7 surrounding the albino (Tyr; c) locus, corresponding to a 6- to 11-cM Tyr deletion, has been the target of a large-scale ENU-induced mutagenesis screen (193, 194). A segment of chromosome 7, from a mutagenized genome bred from ENU-treated males, was made hemizygous opposite the deletion to allow for recognition and recovery of new recessive mutations that map within the albino deletion complex. This mutagenesis experiment has now been completed, the results of finalized complementation and deletion-mapping studies provide evidence for recovery of a total of 31 nonclustered muta-
tions (including the 9 previously reported mutations (193)), representing 10 loci, and for allelic series of mutations, with differences in severity, at several loci (196). Another mutagenesis effort, designed to isolate recessive mutations of many phenotypic classes, is targeting mouse chromosome 11 (104). The goal of this effort is to saturate this chromosome with mutations to define gene function, then use linkage conservation between the mouse and human to predict gene function in the human. Two mutagenesis schemes have been undertaken, including a two-generation deletion pedigree (Fig. 13A) and a three-generation inversion pedigree (Fig. 13B). The deletion scheme can only be used for deletions (Df) whereby the heterozygous mouse (Df/H11002) is still viable. Although this limits the screen to certain regions, it allows mutations to be isolated in

---

*Physiol Genomics* • VOL. 11 • www.physiolgenomics.org
only two generations. In contrast, whereas the inversion scheme allows a larger portion of the chromosome to be screened for mutations, it requires three generations of breeding (103, 104).

3.1.2.4 ALTERNATIVE SCREENING STRATEGIES. Obtaining mutants will not be possible for all genes, due to the redundancy of gene function and the wide range of compensatory mechanisms that exist in cellular pathways. This is important, as phenotype-based screens only allow for the identification of genes whose functions are not covered by redundancy (13). In an attempt to overcome this issue, “sensitized” screens can be performed, where “sensitization” occurs by changes in genetic background (such as the use of “modifiers”) or environmental challenges (such as a salt challenge to detect susceptibility to hypertension). “Modifiers” are ENU-mutated mice that, when mated to a mouse carrying a dominant mutation with a specific phenotype, produce F1 offspring showing an increase or decrease (modification) in the severity of the phenotype. For example, Mom1 (“modifier of Min-1”) mice (50), when mated to Min-1 mice (which carry a dominant mutation in the mouse homolog of the APC gene (155), responsible for intestinal adenoma formation in humans), produce offspring showing elevated levels of tumor formation. Another screening strategy is allele-specific screens, where ENU mutagenesis is used to produce point mutations in specific genes to generate additional alleles, and noncomplementation in compound heterozygotes is used as the screening endpoint (13). Point mutations leading to hypomorphic, hypermorphic, or neomorphic alleles can provide valuable information about the function of that gene [for an example, see the 5 ENU-induced alleles of the quaking (qk) locus mentioned above]. The generation of ENU-induced allelic series can be enhanced by using ES cells treated with an inhibitor of O6-alkylguanine-alkyltransferase [a DNA repair enzyme involved in removing various alkyl adducts from the O6 position of guanine bases (178)] prior to exposure to ENU, with the treated cells still retaining their germ-line competency (39).

3.1.2.5 GENERATING RESOURCES FOR MUTANT MICE. Although ENU mutagenesis was pioneered nearly 25 years ago (204), only recently have major centers directed entire programs toward the purpose of generating hundreds of new ENU mouse mutants (see Table 2) (87, 167). Large amounts of phenotype data are accumulating on large numbers of these mutant mouse strains, creating the need for phenotype databases that can be linked with mapping and mutagenesis databases. The primary mouse database is the Mouse Genome Database housed at the Jackson Laboratory (http://www.informatics.jax.org), which includes a Mouse Locus Catalog describing existing mouse mutants and provides extensive mapping information on their locations. In addition, the International Mouse Strain Resource (IMSR) has been developed (54) (http://www.jax.org/pub-cgi/imsrlist and http://imsr.har.mrc.ac.uk/), which simplifies the search for mutant mice. However, to provide a genetic resource to the scientific community, the mutant mice themselves must also be readily available. Many mutant stocks are available from commercial vendors and the Jackson Laboratory (plus several NIH-funded Mutant Mouse Resource Centers in various regions of the US). Oak Ridge National Laboratory (ORNL) has a large collection of mouse stocks, most of which propagate mutations (single base-pair changes to rearrangements) induced by radiations or chemicals in various stages of male or female gametogenesis (http://bio.lsd.ornl.gov/mouse/). The European Mutant Mouse Archive (EMMA), with centers in Monterotondo (Roma, Italy), the Centre National de la Recherche Scientifique (CNRS; Orleans, France), the Gulbenkian Institute (Lisbon, Portugal), and the Karolinska Institute (Huddings, Sweden) is a resource for the European efforts. Nevertheless, it is important to point out that, despite the important role played by each of these centers, chemical mutagenesis is equally suitable as a research tool for smaller laboratories. Smaller laboratories us-
ing chemical mutagenesis as a genetic approach to identify new genes involved in embryonic development include groups who screen for recessive, ENU-induced mutations that cause morphological abnormalities at midgestation (Kathryn Anderson’s group at the Sloan-Kettering Institute; http://www.ski.edu/project_summary.cfm?Lab=62&Project=161) and embryonic day 18.5 (David Beier’s group at Brigham and Women’s Hospital; http://www.hms.harvard.edu/dms/bbs/fac/beier.html).

3.1.3 Transposon-mediated mutagenesis. Transposons are naturally occurring genetic elements that can move from one genomic locus to another. Transposons are flanked by terminal inverted repeats (IRs) that contain binding sites for the transposable, with transposition mediated by the transposable protein encoded by the transposon (118, 257). In contrast to large-scale ENU mutagenesis schemes where identification of the causative point mutation can be difficult and time-consuming, transposon-tagged mutagenesis has proven to be one of the most effective means in establishing a genotype-phenotype relationship in model organisms such as yeast (201, 251), C. elegans (184), Drosophila (6, 20, 268), and mice (in a few special situations) (3, 93, 169). The limited application of this approach in mice has primarily been due to the lack of an appropriate (“cut-and-paste” type) transposon.

The Tc1/mariner superfamily of transposons transpose by a “cut-and-paste” mechanism, whereby the transposable protein and the flanking IR sequences engage in a series of molecular events leading to excision of the element from its genomic locus and reintegration into a different locus (185). One such transposon system is Sleeping Beauty (SB), a synthetic transposable 1.6-kb element made from defective copies of an ancestral Tc1-like fish element (91), flanked by 250 bp terminal IRs and encoding a single protein, the Sleeping Beauty transposase, that catalyzes its transposition from one genomic locus to the next (reviewed in Ref. 92). Recently, SB has been shown to be capable of inserting foreign genes into the chromosomes of cultured mouse ES cells (139), as well as somatic integration of DNA into mouse chromosomes, resulting in long-term transgene expression in adult mice (273). Furthermore, the SB element, like many DNA transposons, has been shown to exhibit preferential integration (transposition) into linked loci (139). This is a useful feature for a mutagenic agent when used in combination with mice containing specific chromosomal deletions (segmental hemizygosity). Thus the SB element looks to be a promising basis for establishing a general transposon-tagged mutagenesis scheme in mice.

3.2 Gene-Trap Mutagenesis

Although ENU mutagenesis has many advantages (105), it provides no molecular landmarks with which to recover mutated genes. One mutagenesis system that does is gene trapping. Gene-trap vectors have evolved from enhancer-trap vectors, a molecular tool used to identify and characterize mammalian enhancer sequences from cell lines (263). Gene-trap mutagenesis is a technique that randomly generates loss-of-function mutations and reports the expression of many mouse genes (reviewed in Refs. 37, 230). Gene-trap strategies in mouse ES cells are increasingly being used for detecting patterns of gene expression (26, 64, 72, 219, 222, 258, 270, 272), identifying targets of signaling molecules and transcription factors (222, 232), and isolating and mutating endogenous genes (83, 223, 279).

3.2.1 Vectors. 3.2.1.1 Basic trapping vectors. The basic types of trap vector are promoter-trap and gene-trap vectors (enhancer-trap also exist, but are rarely used). As shown in Fig. 14A, promoter-trap vectors contain a promoterless reporter gene and selectable marker (64, 191, 254). Reporter expression and mutagenesis occur when the vector inserts into an exon to generate a fusion transcript that comprises upstream endogenous exonic sequence and the reporter gene. Promoter-trap vectors generally contain a selectable marker for drug resistance or for β-gal or GFP. Several trap vectors have also been developed that are fusion proteins; β-geo, a fusion between neo and LacZ, is the best known of these reporters (64). One disadvantage of promoter trapping is that only genes that are transcriptionally active in ES cells can be selected (although this is at least 50% of all genes). As shown in Fig. 14B, gene-trap vectors are characterized by splice acceptors proceeded by a promoterless reporter gene and function by generating a fusion transcript with the upstream coding sequence and the reporter gene, to generate a null allele. The fusion transcript generated by the trapping of a gene serves as a template for 5’-RACE, allowing the trapped gene to be cloned (222). However, a disadvantage of gene-trap vectors is that because the insertion occurs in an intron, alternative splicing can take place, leading to lower levels of wild-type transcripts and resulting in hypomorphic alleles (145).

3.2.1.2 Specific trapping vectors. It is potentially possible to mutate every gene in the mouse genome by a combination of trapping vectors to reduce any bias (55). However, the bottleneck in gene trapping is isolation of the trapped genes. These vectors have been further developed to provide a tagged site in the genome in the form of loxP site, so that cassettes can be reintroduced into the cell line at the trapping site by recombinase-mediated cassette exchange (RMCE) (60). These vectors are known as postinsertional modification vectors. By combining homologous recombination and gene-trapping strategies, there will be the possibility to modify the trap site. Several groups have included recombination sites (loxP, PRT) in their vectors, allowing recombinase-mediated, postinsertional modifications of the gene-trap locus (9, 78). For example, it is possible to use these vectors to knock-in a transgene into the trapped site to ablate or immortalize cells expressing the trapped gene using a diph-
theria toxin cDNA or an SV40 T-antigen, respectively (28, 265).

To trap genes not expressed in undifferentiated ES cells, polyadenylation (poly-A)-trap vectors have been developed, in which a constitutive promoter drives the expression of a selectable marker (such as neomycin) that lacks a poly-A signal but contains a splice donor signal (166, 207, 279). Therefore, a spliced poly-A signal from an endogenous gene is needed to generate neomycin-resistant (Neo\textsuperscript{r}) clones; only gene insertions will generate Neo\textsuperscript{r} clones, whereas intergenic insertions will be lost.

A secretory-trap vector has been developed, taking advantage of protein sorting and the fact that β-gal activity is abolished in the endoplasmic reticulum, to specifically trap genes that encode secreted and transmembrane proteins and are expressed in ES cells (223). Containing a transmembrane (TM) domain immediately downstream of a splice acceptor (SA) site, this vector, pGT1.8TM, is able to prevent the β-gal fusion protein produced when a secretory gene is trapped from entering the endoplasmic reticulum. Thus secreted proteins can be identified based on their β-gal activity (223). With the idea being that β-gal and PLAP would stain both the cell body and axon, respectively, thus providing markers to study axon guidance and growth (121).

3.2.2) Method of delivery. The most commonly used methods of delivery of the trapping vector into the ES cells are retroviral infection and electroporation. Although retroviruses are reported to have a propensity to insert themselves at the 5’ end of genes [particularly in the first intron and 5’ untranslated region (77, 111, 227, 252)], plasmid-based vectors (electroporated into ES cells) show a more random spectrum of integration (41). The preference for inserting at the 5’ end of genes makes retroviral vectors more likely to produce a null allele (230). One advantage of retroviral vectors is that they are known to insert an intact single copy of the trapping vector into a locus, whereas plasmid vectors are often multiple copy insertions. Plus, the infection efficiency of viruses can be 100%, thereby reducing the number of cells needing to be transfected (271). Achieving saturation mutagenesis will most probably require the use of both strategies (230).

3.2.3) Large-scale trapping screens. 3.2.3.1) Phenotypic screens. The first application of gene trapping to isolate lethal mutations generated 24 strains; heterozygous intercrosses showed that 9 of the 24 strains carried embryonic-lethal mutations (64). More recently, phenotypic analysis of 60 mouse ES cell lines with secretory gene-trap insertions reported that one-third of the generated mouse strains had recessive
lethal phenotypes, with five additional strains showing visible adult phenotypes (151). To date, >100 gene-trap insertions have been described in the literature, with 60% of the insertions showing an obvious phenotype and with 40% being recessive lethal mutations, making the frequency of recessive lethal mutations and obvious phenotypes generated by gene-trap mutagenesis close to that generated by gene targeted mutagenesis (230). Phenotypic screens of gene-trap strains have not really begun, but will probably replace ENU in the years to come.

3.2.3.2) EXPRESSION SCREENS. The in vitro differentiation of ES cells has been used to study the effects of targeted mutations on the hematopoietic, vascular, myoblast, and other early lineages (99, 165, 215, 264). Expression trapping and induction trapping exploit this use of ES cells to identify and mutate genes that are expressed in specific cell lineages or that respond to specific signals. Expression-trap screens have been performed to identify and mutate genes that are expressed in hematopoietic (84, 157, 229) and endothelial lineages (229), cardiomyocytes (12), chondrocytes (12), and neurons (219). Induction screens have identified and mutated genes regulated by retinoic acid (63, 66, 112, 208, 145), engrailed homeobox proteins (142), and irradiation (248). In addition, many groups use reporter gene expression as a way to assess whether a specific developmental pathway has been disrupted (by their trapping) (232, 270). More recently, a new method for fast and efficient trapping of genes whose transcription is regulated by exogenous stimuli has been developed; composed of a promoterless retroviral vector transducing a green fluorescent protein-nitroreductase fusion protein (GFP-NR) downstream from a splice acceptor site (147). The GFP expression allows for flow cytometric identification and sorting of cells in which the trap is integrated downstream of an active promoter, whereas the NR allows pharmacological selection against constitutive GFP-NR expression (147).

3.2.3.3) GENOTYPIC SCREENS. Sequence-based (genotypic) screens have been the most recent type of gene-trap screen to emerge, largely because 5′-RACE has only recently become amenable to high-throughput analysis (243). Using poly-A gene trapping (which permits 3′-RACE to identify the trapped gene), >100,000 trapped insertions in ES cells have been deposited into “OmniBank” (a branch of Lexicon Genetics) (279). With the use of the OmniBank database, clones harboring an insertion in a particular gene can be identified, and mice derived from the trapped cell lines can be purchased (at a minimum cost of US $25,000).

3.2.3.4) PUBLIC RESOURCES. The current direction of gene-trap mutagenesis is similar to that of chemical mutagenesis: a combination of large-scale mutagenesis centers carrying out high-throughput screens to generate a worldwide mutant resource and smaller, investigator-driven focused screens. Details of established large-scale screens that possess frozen libraries of mutagenized ES cells are given in Table 3. Each of these laboratories is allowing full access to the libraries of frozen ES cell clones. In a recent international workshop on gene-trap mutagenesis, a consortium of insertion mutagenesis laboratories was formed in association with the International Mutant Mouse Con-

Table 3. Large-scale gene trapping projects

<table>
<thead>
<tr>
<th>Location</th>
<th>Project Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Manitoba Institute of Cell Biology (Canada) (<a href="http://www.escells.ca">http://www.escells.ca</a>)</td>
<td>Aim to generate an allelic series of promoter-trap mutations for genes expressed in undifferentiated ES cells. A searchable database provided (with ~300 clones deposited/month).</td>
</tr>
<tr>
<td>German Human Genome Project Gene Trap Project (<a href="http://tikus.gsf.de/">http://tikus.gsf.de/</a>)</td>
<td>Carrying out a genotype-driven gene-trap mutagenesis programme (267). Delivering β-geo gene-trap and promoter-trap vectors into ES cells by electroporation and retroviral infection. More than 4,500 trapped genes can be searched by gene name or sequence.</td>
</tr>
<tr>
<td>BayGenomics Gene Trap Project (USA) (<a href="http://baygenomics.ucsf.edu/">http://baygenomics.ucsf.edu/</a>)</td>
<td>More than 750 insertions generated from the combined screens by the Skarnes and Tessier-Lavigne laboratories. Insertion sequence, expression, and phenotypic data on the &gt;340 genes can be searched, and clones requested. A second genotypic screen recently been initiated to generate an insertion resource for all classes of gene.</td>
</tr>
<tr>
<td>University of California Resource of Gene Trap Insertions (<a href="http://socrates.berkeley.edu/~skarnes/resource.html">http://socrates.berkeley.edu/~skarnes/resource.html</a>)</td>
<td>A resource of insertional mutations in mouse ES cells and mice. Carrying out mutational analysis of secreted and transmembrane proteins in mice. Sequence tags identifying the target gene been generated from &gt;500 secretory-trap cell lines. More than 80 insertions been transmitted to the germ line of mice for phenotype analysis.</td>
</tr>
</tbody>
</table>
sortium (162) to promote the accessibility of gene-trap resources (for more details see Ref. 230).

4) CONCLUSIONS AND FUTURE DIRECTIONS

As a model organism, the mouse has much to offer, not the least of which is the similarity of its genome to ours. In this review we have discussed the many techniques and technologies that can be used to modify this genome and highlight the great advance that the mouse genome sequence will afford mouse geneticists. In an era where biologists are asking the question “how is this gene involved in this disease” rather than “which gene is involved in this disease” the mouse is more than ever the organism able to provide the answers. While much mechanistic information can be learned from lower order organisms such as yeast and Drosophila, these lack the anatomy and physiology to be frontline tools for drug discovery. Nothing could reinforce this view more than the massive investments being made in mouse facilities and infrastructure by major pharmaceutical firms, despite turbulent times in biotechnology.

As the future for mouse genetics seems bright, it is prudent to comment on how the field will progress in the years ahead. The most dramatic change to mouse genetics will be the increase in phenotype-driven screens. It is not inconceivable to expect that this mode of inquiry will overtake gene-driven approaches in the next decade. At this point the bottleneck in these studies is our ability to identify the specific mutation causing the phenotype of interest. Recently developed methods for finding mutations in human cancers, involving heteroduplex analysis, are likely to be employed in finding mutations in mutant mouse strains, further accelerating progress. Within the next decade is likely that a catalog of mouse mutations will become available. It will be possible to order a mouse harboring a mutation in your gene of interest and thus circumvent the time-consuming process of developing a knockout line. Despite this, people will continue to perform gene targeting in mouse ES cells as they seek to examine the role of subtle mutations and regulatory elements such as enhancers. It will be an era when the pace of discovery will be fierce, and the rewards to humankind immense.

We thank all of those whose tireless efforts have provided valuable insights into the ever-burgeoning field of mouse genetics, and we apologize to those whose work was not cited.

REFERENCES


Geng Y, Whoriskey W, Park MY, Bronson RT, Medema Gajovic S, Chowdhury K, and Gruss P.


Vijaya S, Steffen DL, and Robinson HL. Genet. 249.


Von Melchner H and Ruley HE. Genet. 250.

Von Melchner H and Ruley HE. Genet. 251.

Vooijs M, Jonkers J, and Berns A. Genet. 252.


