HEP (2007) 178:129–147

Gene Trap Mutagenesis

A. Abuin (✉) · G. M. Hansen · B. Zambrowicz

Lexicon Genetics, 8800 Technology Forest Place, The Woodlands, TX, 77381, USA
aabuin@lexgen.com

Abstract Our ability to genetically manipulate the mouse has had a great impact on medical research over the last few decades. Mouse genetics has developed into a powerful tool for dissecting the genetic causes of human disease and identifying potential targets for pharmaceutical intervention. With the recent sequencing of the human and mouse genomes, a large number of novel genes have been identified whose function in normal and disease physiology remains largely unknown. Government-sponsored multinational efforts are underway to analyze the function of all mouse genes through mutagenesis and phenotyping, making the mouse the interpreter of the human genome. A number of technologies are available for the generation of mutant mice, including gene targeting, gene trapping and transposon, chemical or radiation-induced mutagenesis. In this chapter, we review the current status of gene trapping technology, including its applicability to conditional mutagenesis.

Keywords Gene trapping · Mouse · Knockout · Genome · Function

1 Introduction

The recent sequencing of the human and mouse genomes (Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002) underscores the need for the functional annotation of thousands of novel genes. For the most part, humans and mice share an almost identical set of genes and physiology. These similarities, along with the development of robust and precise mouse mutagenesis

techniques over the last 25 years, make the mouse the centerpiece model system for the interpretation of human gene function. Recognizing the central role of mouse genetics in the functional annotation of the human genome, government-sponsored initiatives, such as the Knockout Mouse Project, are being launched to knockout every gene in the mouse and analyze their function through the study of the resulting phenotype (Austin et al. 2004; Auwerx et al. 2004). These efforts call initially for the generation of null alleles in all genes through embryonic stem (ES) cell technologies, namely gene targeting and gene trapping. Gene targeting is a widely used technique in which a DNA construct, or targeting vector, is introduced into ES cells to generate the desired mutation through homologous recombination between sequences in the vector and the target locus (Bradley et al. 1992; Joyner 2000; Mansour et al. 1988). Gene targeting allows for the generation of virtually any desired mutation, such as deletions, insertions, point mutations (Hasty et al. 1991), and large chromosomal deletions and rearrangements (Ramirez-Solis et al. 1995), but requires significant upfront effort to generate a different targeting vector for each locus, as well as downstream efforts in the screening of multiple ES cell clones (typically hundreds) in order to identify the relatively rare homologous recombinants.

Gene trapping is a method of random mutagenesis in which the insertion of a DNA element into endogenous genes leads to their transcriptional disruption. Gene trapping elements can be endogenous in nature, such as transposable elements (Wilson et al. 1989), or exogenous recombinant DNA constructs (Gossler et al. 1989). Unlike gene targeting, the insertional nature of gene trapping does not allow for the generation of all types of mutations, such as deletions or point mutations. Its main advantage, however, is that a single gene trapping vector can be used to mutate and identify thousands of genes in ES cells (Zambrowicz et al. 1998; Wiles 2000). This scalability makes gene trapping very suitable for high-throughput, large-scale, and cost-effective mutagenesis programs like the ones being proposed by governmental agencies around the world. In fact, large-scale gene trapping efforts in mouse ES cells have already taken place, demonstrating the ability to mutate at least 60% of all mouse genes in ES cells, including potential pharmaceutical targets such as the *Wnk1* protein kinase (Zambrowicz et al. 2003).

2
Generating Null Alleles Through Gene Trap Mutagenesis

2.1
Types of Gene Trapping Vectors

Gene trapping vectors are typically designed to lack an essential transcriptional component, such as an enhancer (O'Kane and Gehring 1987), promoter (Hicks et al. 1997), or polyadenylation signal (polyA) (Niwa et al. 1993), ren-

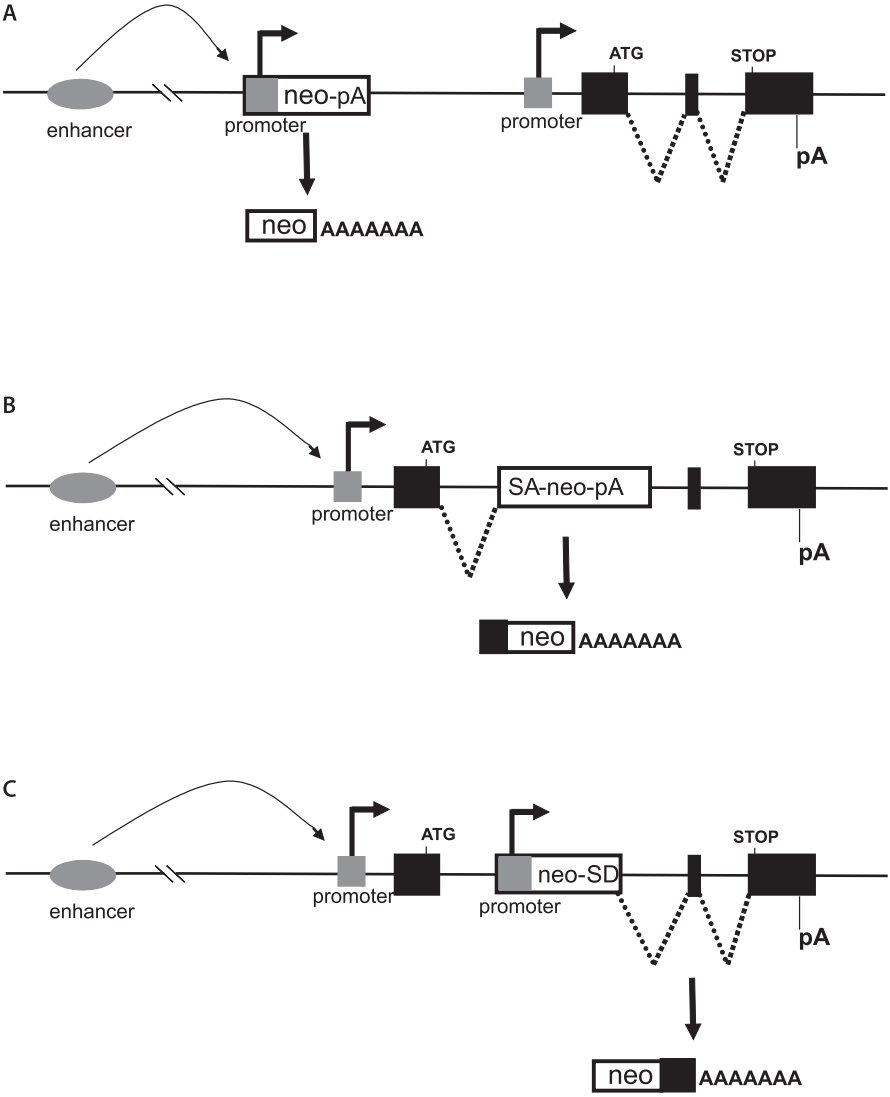


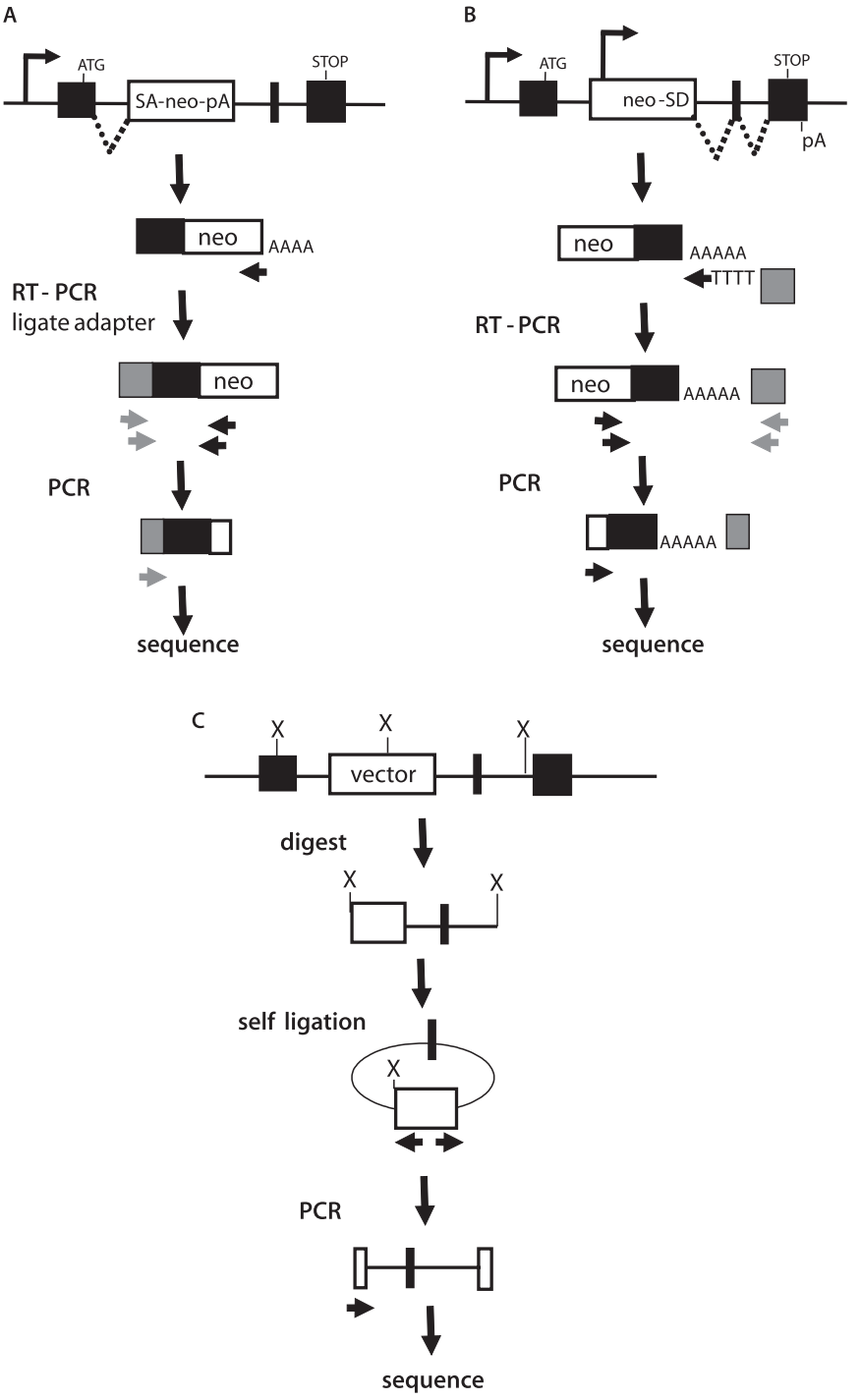
Fig. 1 A–C Different types of gene trapping strategies. **A** An enhancer trap, showing a gene trapping vector containing its own promoter, neomycin resistance (*neo*) gene and polyadenylation signal (*polyA*). The promoter in the vector requires the action of an enhancer to drive transcription of the *neo* mRNA. **B** A promoter trap, showing a vector containing a splice acceptor sequence (*SA*), followed by the *neo* gene and a polyA signal. Insertion in a gene, downstream of the endogenous promoter, leads to expression of the *neo* mRNA fused to the upstream exons (*black boxes*) of the trapped gene. **C** A polyA trap, showing a vector containing its own promoter, the *neo* gene and a splice donor sequence (*SD*). Insertion in a gene, upstream of the endogenous polyA, leads to expression of the *neo* mRNA fused to the downstream exons of the trapped gene

dering them transcriptionally active only when inserted into an endogenous gene (Fig. 1). Vectors containing a splice acceptor and a selectable marker but lacking a promoter element, also known as promoter traps, are the most commonly used type of gene trapping vector. Promoter trap vectors can be designed to contain a reporter gene for in vivo analysis of the expression of the trapped genes (Friedrich and Soriano 1991; Gossler et al. 1989). Constructs are introduced into the cells of interest, such as mouse ES cells, through retroviral infection, or by electroporation. Random insertion of the vector into a transcribed gene leads to expression of the selectable marker and resistance to antibiotics. Insertion of the vector into the gene also leads to its transcriptional disruption. Vectors lacking polyadenylation sequences, also known as polyA traps, can be used to trap untranscribed genes, since they contain their own promoter driving expression of the selectable marker.

2.2 Identification of Trapped Genes

Unlike other random mutagenesis methods, such as chemical (Brown and Peters 1996) or radiation-induced mutagenesis (You et al. 1997), gene trapping vectors can be used to readily identify the mutated gene (Fig. 2). This is typically accomplished through 5′ or 3′ RACE (rapid amplification of cDNA ends) of fusion transcripts generated by the splicing of endogenous exons into promoter traps and out of polyA traps, respectively (Fig. 2A, B). It is important to note that 5′ or 3′ RACE allows for the identification of the mutated gene by virtue of its cDNA sequence, but does not pinpoint the exact location of the insertion within the genome. RACE was virtually the only tool available for identifying a trapped gene prior to the availability of the mouse genome sequence. With the sequencing of the mouse genome, however, a short sequence tag of the genomic DNA immediately flanking the gene trapping vector is sufficient, in most cases, to determine its precise chromosomal location. Flanking genomic sequence can be obtained through a variety of methods, the most common of which is

Fig. 2 A–C Identification of the trapped gene. **A** 5′ rapid amplification of cDNA ends (5′ RACE). After RT-PCR, an adapter oligonucleotide is ligated to the 5′ end of the cDNA for the transcript produced by a promoter trap construct. Primers complementary to the vector and the adapter and then used to amplify the fusion transcript and identify the trapped gene by sequencing. A nested PCR strategy is typically used. **B** 3′ RACE. A polyT primer with a unique sequence tail is used for RT-PCR. Nested PCR with primers complementary to the vector and the unique tail is then used to amplify the fusion transcript generated by a polyA trap construct. **C** Inverse Genomic PCR (IPCR). Genomic DNA from individual gene trapped clones is cleaved with one or more restriction enzymes (X) to produce a vector–genomic junction fragment that can be ligated to produce a circular template. Two vector-specific primers can be used to amplify the flanking genomic DNA. Sequence of the flanking genomic DNA allows for the precise mapping of the insertion within the genome


A
ATG
SA-neo-pA
STOP
neo
AAAA
RT - PCR
ligate adapter
neo
PCR
sequence
B
ATG
neo-SD
STOP
pA
neo
AAAAA
TTTT
RT - PCR
neo
AAAAA
PCR
AAAAA
sequence
C
X
X
X
vector
digest
X
X
self ligation
X
PCR
sequence

inverse genomic PCR (IPCR) (Silver and Keerikatte 1989). In this technique, genomic DNA from gene-trapped ES cells is cleaved with a restriction enzyme that cuts within the vector, producing a vector–genomic junction fragment that is then circularized by ligation at low DNA concentrations that favor intramolecular ligation. The mouse genomic DNA adjacent to the vector can then be amplified by PCR using outward-facing primers complementary to the vector. The use of multiple enzymes or combinations of enzymes increases the probability of generating a circular template of appropriate size for efficient PCR amplification (Fig. 2C). IPCR is ideally suited for large-scale retroviral gene trap insertion site identification because it requires no prior knowledge of the disrupted gene sequence and can be used repeatedly for the same gene trapping vector. Furthermore, since retroviral insertion mutations are discrete in that they are not associated with deletions or alterations of the inserted vector or the genomic sequence at the site of insertion (Varmus 1988), IPCR can be successfully applied to all gene trap mutations. There are several advantages to this approach for gene identification. First, the technique is more robust since it uses DNA rather than RNA as a template. Second, it provides more accurate data with which to predict gene disruption. This is particularly important when the gene of interest contains multiple promoters or transcriptional start sites, in which case particular attention must be paid to the exact location of the gene trapping vector with respect to each transcript. In these instances, RACE data often fails to provide adequate confirmation of the transcript classes affected by the gene trap. Third, the precise genomic insertion site of the vector is essential for PCR-based genotyping of mice generated through gene trapping (Fig. 3). For large-scale, publicly available gene trap libraries, it is imperative that such data be available.

2.3 Mutagenicity of Gene Trapping Vectors

Although the number of publications describing mutant mice generated through gene trapping lags significantly behind the number of gene-targeting publications, an increasing volume of evidence shows that gene trap insertions in genes are mutagenic and tend to result in null alleles (Skarnes et al. 1992; Voss et al. 1998) (also, see Table 1). A number of groups around the world (Hicks et al. 1997; Zambrowicz et al. 1998; Wiles et al. 2000; Hansen et al. 2003) have carried out efforts to mutate large numbers of genes in mouse ES cells through gene trapping, and initiatives like the Knockout Mouse Project (Austin et al. 2004) will only increase the number of gene-trapped mouse lines characterized in the future. Thus, a detailed assessment of the mutagenicity of gene trap mutations is an important step at this juncture. We have used gene trapping to generate OmniBank, a library of more than 270,000 frozen mouse ES cell clones uniquely identified by their corresponding OmniBank Sequence Tag (OST). The insertion mutations contained in this gene trap library represent

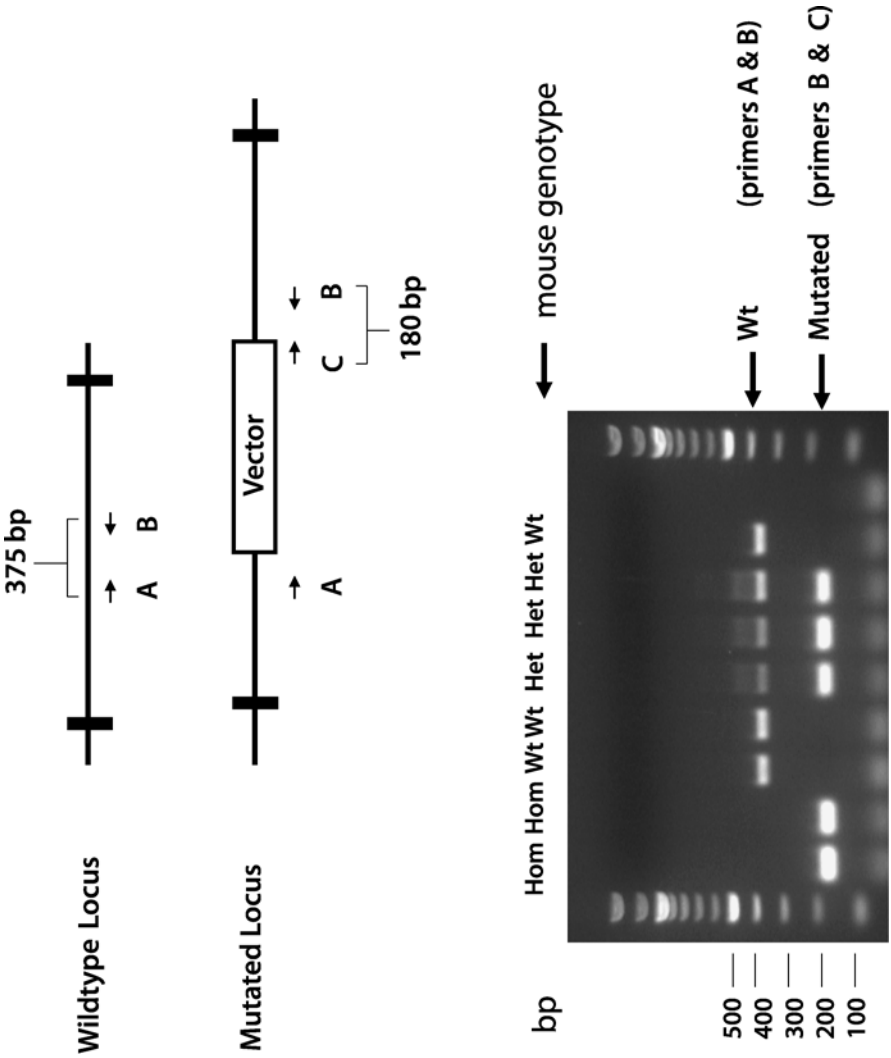


Fig. 3 Genotyping strategy for gene trap mutations. Primers A and B flank the genomic insertion site and amplify a genomic DNA product for the wild-type allele. Primer C, complementary to the gene trapping vector, and primer B amplify the mutated allele. *Black boxes* represent exons of the trapped gene. *bp* base pairs

mutations in approximately 60% of all mouse genes (Zambrowicz et al. 2003). OSTs were generated by 3′ RACE (Fig. 2B). As mentioned above, the cDNA sequence data provided by the OST is useful for predicting the relative location of the gene trap event within a gene of interest; however, it does not pinpoint the exact insertion site in the genome nor facilitate the design of specific PCR genotyping assays. Therefore, OmniBank gene trap mutations are further

Table 1 OmniBank gene trapped mouse lines: efficient generation of null alleles

OmniBank trapped gene	Published confirmation of null allele	Reference
Thioredoxin 2 (*Trx-2*)	- No protein detected by Western blot	Nonn et al. 2003
Mdm4 (*Mdmx*)	- No protein detected by Western blot	Migliorini et al. 2002
Insulin-degrading enzyme (*IDE*)	- No protein detected by Western blot	Farris et al. 2003
Insulin-degrading enzyme (*IDE*)	- No message detected by RT-PCR - No protein detected by Western blot - No IDE enzymatic activity detected	Miller et al. 2003
tRNA synthetase p38 subunit	- No message detected by Northern blot - No protein detected by Western blot	Kim et al. 2002
Wave1	- No protein detected by Western blot	Dahl et al. 2003
Testis–brain RNA-binding protein (*TB-RBP*)	- No protein detected by Western blot	Chennathukuzhi et al. 2003
Insulin-like growth	- No message detected by RT-PCR	Hansen et al. 2004
factor II mRNA-binding protein 1 (*IMP1*)	- No message detected after extended exposures of whole-mount in situ hybridizations	
Calpain 3 (p94)	- No message detected by RT-PCR - No protein detected by Western blot	Kramerova et al. 2004
GYS1 glycogen synthase	- No protein detected by Western blot - No GYS1 enzymatic activity detected - No glycogen detected	Pederson et al. 2004
Tektin-t	- No message detected by Northern blot - No protein detected by Western blot	Tanaka et al. 2004
Phosphatidylinositol 5-phosphate 4-kinase	- No message detected by RT-PCR - No protein detected by Western blot	Lamia et al. 2004

characterized prior to mouse knockout production using IPCR (Fig. 2C). By identifying the precise genomic insertion site of the vector, it is possible to eliminate ES clones in which the gene trapping vector inserts within the promoter region or upstream of the gene, but for which an OST was still generated through splicing into the second exon of the gene. These gene trap insertions are unlikely to be mutagenic.

To evaluate the mutagenicity of OmniBank gene trap mutations in vivo, all nonembryonic lethal OmniBank mouse lines are subjected to a sensitive RT-PCR quality control step. We check for the loss of the endogenous transcript in homozygous mutant animals relative to wild-type controls. RNA is isolated from selected tissues known to normally express the transcript and subjected to RT-PCR using primers complementary to exons flanking the insertion site of the vector (Fig. 4).

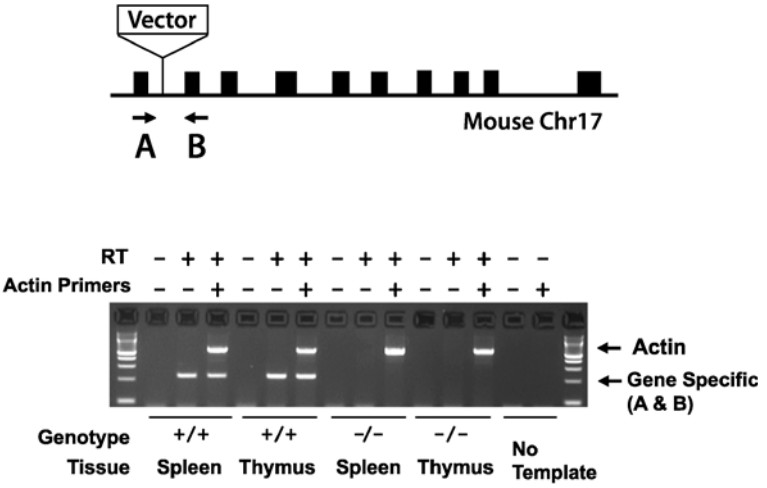


Fig. 4 Assessing the mutagenicity of gene trap insertions by RT-PCR. Primers D and F are complementary to exons flanking the insertion site in the *PolH* gene in mouse chromosome 17 (accession number NM_030715). RT-PCR using primers D and F shows absence of endogenous message in the spleen and thymus of homozygous animals. Control primers to the murine *beta actin* gene were used (accession number M12481)

To date, we have generated and analyzed 1,155 OmniBank mouse lines in which the vector insertion was confirmed by IPCR to be within the gene of interest. Of these, 370 (32%) were not analyzed by RT-PCR due to embryonic lethality (315 lines) or perinatal lethality or reduced viability (55 lines). Of the remaining 785 lines, 706 (90%) showed no detectable wild-type transcript by RT-PCR and 79 (10%) showed drastically reduced levels of wild-type transcript. For selected lines showing reduced but detectable expression, quantitative PCR (QPCR) analysis has generally shown message levels of less than 5% of wild-type controls. Further evidence that gene trap mutations produce null alleles is shown in Table 1. In this table, we have gathered independently published, recent analyses of OmniBank mouse lines where the authors used methods other than RT-PCR to confirm that the allele resulted in a null mutation. Taken together, this data demonstrates that intragenic insertions efficiently disrupt gene transcription in vivo.

Even for the severe hypomorphs in which RT-PCR is able to detect greatly reduced levels of wild-type transcript, the mutation is likely to be useful for the elucidation of gene function, as exemplified in the case of the *Slit3* mutation. In an independently published analysis of a *Slit3* OmniBank gene trap mutation (Liu et al. 2003), the authors were unable to detect any message by in situ hybridization. When the more sensitive RT-PCR technique was used, greatly reduced expression was detected. QPCR revealed 1.48% and 1.27% of wild-type message levels in the homozygotes, depending on the QPCR primers used. Importantly, the phenotype of this OmniBank line (congenital diaphragmatic

hernia, kidney agenesis, and cardiac defects) is the same as the one described for a published *Slit3* knockout generated through gene targeting and predicted to lead to a null mutation (Yuan et al. 2003). Taken together, sequential QC processes such as IPCR of the ES cells in vitro and RT-PCR of the mice in vivo provide a robust method for production of null alleles through gene trapping.

Despite the ability of gene trap events to disrupt the transcript in which they are inserted, it is important to note that, just as with gene targeting, careful attention must be paid to any potential alternative splicing of the gene of interest. Depending on the location of the insertion, mRNAs transcribed from alternate promoters or containing multiple polyadenylation signals may not be affected by the gene trap (Li et al. 2003). The sequencing of the human and mouse genomes, along with the mapping of mRNA and expressed sequence tags (EST) data to public genome-viewing software, can greatly aid in the choice of a gene trap that will disrupt all transcripts of the gene of interest. As the number of publicly available gene trap events increases, the chance of finding an insertion at the desired location will also increase. Alternatively, there may be cases when it may be of interest to disrupt only a subset of transcripts. An example that illustrates the potential benefit of transcript-specific gene traps is that of the *Nogo* gene. *Nogo* codes for a putative inhibitor of axonal regeneration. Alternative splicing and an alternate promoter lead to the production of three *Nogo* transcripts encoding three distinct proteins. A mutation generated through gene targeting that disrupts all transcripts resulted in embryonic lethality (Zheng et al. 2003). Interestingly, mice carrying a gene trap mutation that selectively disrupts two of the three transcripts were viable, allowing for the study of axonal regeneration in adult mice (Kim et al. 2003).

3 Conditional Gene Trapping

The ability of gene trapping to mutate genes on a genome-wide scale has also prompted investigations into the use of this approach to generate more custom mutations such as conditional alleles. Recombinase sites, the essential elements of conditional allele construction, can be incorporated in various configurations within a gene trapping vector to allow for removal or inversion of its functional components. This provides some level of control over the transcriptional disruption of the target gene; however, it hinges upon the assumption that intronic gene trap insertions can be transcriptionally silent when an inverted or partially excised gene trapping vector is present within an intron. At present there is scant data available to properly address this issue; therefore it remains to be seen whether the gene trap approach will be generally useful for conditional allele creation. Regardless, the current demand for mutant alleles and the ease of producing these alleles through gene trapping will continue to push such efforts forward.

Several approaches for generation of conditional gene trap alleles have been reported to date (Li et al. 2005; Schnutgen et al. 2005; Xin et al. 2005). The com-

plexity and flexibility of each approach varies widely, from the most simplistic, in which the splice acceptor element is flanked by *lox* P sites, to elaborate approaches using nested recombinase sites to allow multi-step modifications. The most versatile of these vectors was recently reported by Schnutgen et al. (2005). Their vectors contain multiple nested site-specific recombinase sites allowing gene trapping, allele repair, and allele reinduction through successive recombination steps. Their vectors contain a standard splice acceptor and selection cassette flanked by four nested sets of heterotypic target sequences for the Cre and Flp recombinases. This design allows conditional induction of the mutant allele or conditional reversion of the wild-type allele. In this report, the authors also show that, at least in vitro, their gene trap elements are silent when inverted. Using RT-PCR, they evaluated the transcription of two gene trap mutations in endogenous loci and found no detectable disruption of the wild-type allele when the gene trap elements were inverted. This data provides a first step toward the production of a useful conditional resource, but more data is needed, particularly in vivo studies. As a general resource for either null or conditional alleles, this library also lacks data for germline transmission rates for these ES cell clones and does not currently include genomic insertion site sequence data to confirm the location of these mutations and enable genotyping of the mouse lines.

A single report is currently available describing the function of conditional gene trapping vectors within an endogenous gene in vivo. This report describes an approach that allows conditional allele repair through the excision of the splice acceptor element within an otherwise standard gene trapping vector (Li et al. 2005). An insertion of a gene trapping vector containing *lox* P sites flanking the splice acceptor element within the Myocardin-related transcription factor b (*Mrtf-b*) gene was used to study the requirements for this gene within a specific cell lineage. *Mtrf-b* is essential for embryonic development. Homozygous null *Mtrf-b* embryos die late in gestation or shortly after birth due to defects in cardiac outflow tracts. It was hypothesized that these defects were cell autonomous and therefore might be rescued by the presence of wild-type transcript within this cell lineage. This was confirmed by crossing this line with a mouse line that expresses the Cre recombinase protein specifically in the neural crest lineage (Jiang et al. 2000). Homozygous mutants carrying the *Wnt-1 Cre* transgene survived to birth and showed no obvious phenotypic defects several weeks after birth. This suggests that Cre excision of the splice acceptor element of the gene trap vector was successful in restoring wild-type *Mtrf-b* splicing, and that the presence of the partially excised vector in this instance did not significantly impede the production of wild-type transcript. This is a promising first step in the application of this approach; however, additional studies and much more detailed analysis of the transcriptional effects of these insertions need to be done in order to conclude that conditional gene trap resources will provide all of the function desired from these types of alleles.

4
Other Applications of Gene Trapping

In addition to its ability to generate null or conditional mutations in mice for the study of gene function, gene trapping has been successfully utilized for a number of other purposes. The high-throughput nature of gene trapping allows for the rapid generation of large numbers of mutants. With the proper assay in place, the effects of these mutations on the cellular process of interest can be probed. In general, however, these screens are limited by the fact that most gene trapping mutations will be heterozygous. This hurdle can be overcome in certain hypodiploid cells containing large regions of functional haploidy, such as certain Chinese hamster ovary (CHO) cells. Hubbard et al. used a promoter trap screen and wheat germ agglutinin selection to identify CHO cells deficient in glycosylation (Hubbard et al. 1994), including cells nullizygous for GlcNAc transferase I (Chang et al. 1993). In another attempt to overcome the heterozygosity of gene trap mutations, a gene trapping system was designed to express antisense RNA to the trapped gene from a transactivated promoter. In principle, successful knockdown of the remaining wild-type allele would result in functional homozygosity and the ability to directly screen for recessive phenotypes in vitro. This system was used to screen for gene trap mutants with a transformed phenotype and the identification of TSG101, a putative tumor suppressor gene (Li and Cohen 1996). Finally, a promising development toward the use of gene trapping for in vitro screens has been the development of ES cells deficient for the Bloom syndrome (*Blm*) gene. These cells display a greatly elevated rate of mitotic recombination, which increases the chance that any given gene trap insertion will become homozygous (Guo et al. 2004; Yusa et al. 2004).

Another use of gene trapping is the identification of genes that are differentially regulated under certain conditions, such as differentiation of specific cell types, or expressed in a specific tissue or organ of interest. Promoter traps with selectable markers and/or reporter genes can be used to select for and follow the expression pattern of the trapped gene in the desired context, without the need for homozygosity. Positive or negative selection for a resistance marker placed under the control of endogenous promoters by the gene trap events can be used to identify genes that are up- or down-regulated in a process of interest. This strategy has been used to screen for genes that are differentially regulated during myoblast differentiation in vitro, such as lysosomal cysteine protease cathepsin B (Gogos et al. 1996), and genes that are either up- or down-regulated by germ-cell signaling in Sertoli cells (Vidal et al. 2001). Gene trapping screens have also been designed to identify genes encoding membrane and secreted proteins (Skarnes et al. 1995), genes displaying specific patterns of expression during development (Gossler et al. 1989; Friedrich and Soriano 1991), genes that control neuronal connections (Leighton et al. 2001), and genes expressed in cardiovascular lineages (Kuhnert and Stuhlmann 2004) or specific tissues, such as the hippocampus (Steel et al. 1998) and inner ear (Yang et al. 1997).

Thus, although the current emphasis is in the use of gene trapping as a tool for the large-scale generation and phenotypic analysis of mice carrying null mutations, it is important to recognize the versatility of gene trapping as a powerful in vitro and in vivo screening tool.

5 Targeted Trapping

A recent variation of the gene trapping approach has combined the directed approach of gene targeting through homologous recombination with the efficiency of mutating genes in ES cells through gene trapping (Friedel et al. 2005). This technique, dubbed targeted trapping, uses gene trap constructs for homologous recombination by flanking these cassettes with genomic sequences, thereby targeting them precisely to a chosen intron (Fig. 5). Standard homologous recombination approaches, as mentioned previously, are relatively inefficient at producing correctly targeted ES cells. The advantage of this approach is that the authors report surprisingly high frequencies of correctly targeted events provided the gene of interest is expressed at sufficient levels in ES cells (>50%). In addition, the authors define a threshold of expression in ES cells that appears to be necessary for successful gene trapping as well as targeted trapping. This suggests that although targeted trapping could dramatically reduce at least one of the hurdles associated with homologous recombination (targeting efficiency), it will likely only be useful within the subset of genes that have already been disrupted through large-scale gene trap efforts. As such, this approach will likely not be significant with respect to the Knockout Mouse Project. A limitation to this approach relative to traditional gene targeting is that the high targeting efficiency is apparently lost when a genomic deletion is engineered along with the insertion of the gene trapping components. Furthermore, this approach requires that the arms of homology not include elements of the promoter, which limits the ability for placement of the gene trap cassette near the first exon of many genes. It could prove useful, however, in creating additional allele variety for specific genes, particularly in cases where few or no desirable gene trap alleles are available. It may also be of use in combination with, and extension of, the most recent advances in conditional gene trap approaches allowing the development of a more tractable method for creating conditional alleles.

6 Concluding Remarks

Functional annotation of the mammalian genome is the next essential step in our effort to better understand human physiology and identify new treatments for human disease. The most widely accepted experimental approach

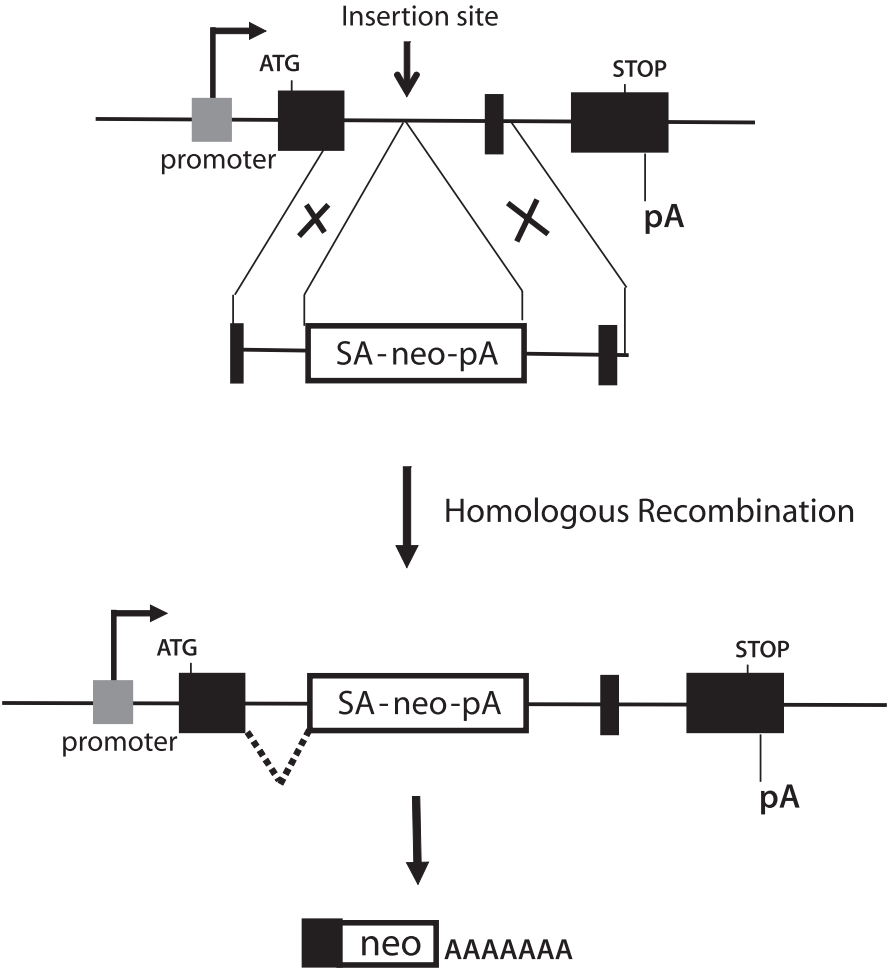


Fig. 5 Targeted gene trapping. A promoterless gene trapping construct containing a splice acceptor sequence (*SA*) followed by the *neo* gene and a polyadenylation signal (*pA*) is flanked by uninterrupted arms of homology to facilitate insertion into the first intron of a hypothetical target gene through homologous recombination. Correctly targeted insertion of the gene trapping cassette leads to expression of the *neo* mRNA fused to the upstream exons (*black boxes*) of the trapped gene

for functional annotation is the gene knockout in a whole mammalian model, the mouse. Gene trapping has gained prominence in recent years as the most efficient high-throughput approach for generating gene knockouts in ES cells, and sufficient data is now available to show that gene trapping vectors are effective at generating null alleles. Phenotypic annotation of the entire complement of mammalian genes will require a combination of multiple gene knockout techniques, and the most efficient combination of these will speed

our progress toward discovery of novel targets for pharmaceutical development and ultimately, disease treatment.

References

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, Grieder FB, Heintz N, Hicks G, Insel TR, Joyner A, Koller BH, Lloyd KC, Magnuson T, Moore MW, Nagy A, Pollock JD, Roses AD, Sands AT, Seed B, Skarnes WC, Snoddy J, Soriano P, Stewart DJ, Stewart F, Stillman B, Varmus H, Varticovski L, Verma IM, Vogt TF, von Melchner H, Witkowski J, Woychik RP, Wurst W, Yancopoulos GD, Young SG, Zambrowicz B (2004) The knockout mouse project. Nat Genet 36:921–924

Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, Berns A, Bradley A, Brown S, Carmeliet P, Chambon P, Cox R, Davidson D, Davies K, Duboule D, Forejt J, Granucci F, Hastie N, de Angelis MH, Jackson I, Kioussis D, Kollias G, Lathrop M, Lendahl U, Malumbres M, von Melchner H, Muller W, Partanen J, Ricciardi-Castagnoli P, Rigby P, Rosen B, Rosenthal N, Skarnes B, Stewart AF, Thornton J, Tocchini-Valentini G, Wagner E, Wahli W, Wurst W (2004) The European dimension for the mouse genome mutagenesis program. Nat Genet 36:925–927

Bradley A, Hasty P, Davis A, Ramirez-Solis R (1992) Modifying the mouse: design and desire. Biotechnology (N Y) 10:534–539

Brown SD, Peters J (1996) Combining mutagenesis and genomics in the mouse – closing the phenotype gap. Trends Genet 12:433–435

Chang W, Hubbard SC, Friedel C, Ruley HE (1993) Enrichment of insertional mutants following retrovirus gene trap selection. Virology 193:737–747

Chennathukuzhi V, Stein JM, Abel T, Donlon S, Yang S, Miller JP, Allman DM, Simmons RA, Hecht NB (2003) Mice deficient for testis-brain RNA-binding protein exhibit a coordinate loss of TRAX, reduced fertility, altered gene expression in the brain, and behavioral changes. Mol Cell Biol 23:6419–6434

Dahl JP, Wang-Dunlop J, Gonzales C, Goad ME, Mark RJ, Kwak SP (2003) Characterization of the WAVE1 knock-out mouse: implications for CNS development. J Neurosci 23:3343–3352

Farris W, Mansourian S, Chang Y, Lindsley L, Eckman EA, Frosch MP, Eckman CB, Tanzi RE, Selkoe DJ, Guenette S (2003) Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain in vivo. Proc Natl Acad Sci U S A 100:4162–4167

Friedel RH, Plump A, Lu X, Spilker K, Jolicoeur C, Wong K, Venkatesh TR, Yaron A, Hynes M, Chen B, Okada A, McConnell SK, Rayburn H, Tessier-Lavigne M (2005) Gene targeting using a promoterless gene trap vector ("targeted trapping") is an efficient method to mutate a large fraction of genes. Proc Natl Acad Sci U S A 102:13188–13193

Friedrich G SP (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev 5:1513–1523

Gogos JA TR, Lowry W, Sloane BF, Weintraub H, Horwitz MJ (1996) Gene trapping in differentiating cell lines: regulation of the lysosomal protease cathepsin B in skeletal myoblast growth and fusion. J Cell Biol 134:837–847

Gossler AJA, Rossant J, Skarnes WC (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244:463–465

Guo G, Wang W, Bradley A (2004) Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. Nature 429:891–895

Hansen J, Floss T, Van Sloun P, Fuchtbauer EM, Vauti F, Arnold HH, Schnutgen F, Wurst W, von Melchner H, Ruiz P (2003) A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A 100:9918–9922

Hansen TV, Hammer NA, Nielsen J, Madsen M, Dalbaeck C, Wewer UM, Christiansen J, Nielsen FC (2004) Dwarfism and impaired gut development in insulin-like growth factor II mRNA-binding protein 1-deficient mice. Mol Cell Biol 24:4448–4464

Hasty P, Ramirez-Solis R, Krumlauf R, Bradley A (1991) Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. Nature 350:243–246

Hicks GG SE, Li XM, Li CH, Pawlak M, Ruley HE (1997) Functional genomics in mice by tagged sequence mutagenesis. Nat Genet 16:338–344

Hubbard SC WL, Ruley HE, Muchmore EA (1994) Generation of Chinese hamster ovary cell glycosylation mutants by retroviral insertional mutagenesis. J Biol Chem 269:3717–3724

Jiang X, Rowitch DH, Soriano P, McMahon AP, Sucov HM (2000) Fate of the mammalian cardiac neural crest. Development 127:1607–1616

Joyner AL (2000) Gene targeting: A practical approach. Oxford University Press, Oxford, England

Kim JE, Li S, GrandPre T, Qiu D, Strittmatter SM (2003) Axon regeneration in young adult mice lacking Nogo-A/B. Neuron 38:187–199

Kim JY, Kang YS, Lee JW, Kim HJ, Ahn YH, Park H, Ko YG, Kim S (2002) p38 is essential for the assembly and stability of macromolecular tRNA synthetase complex: implications for its physiological significance. Proc Natl Acad Sci U S A 99:7912–7916

Kramerova I, Kudryashova E, Tidball JG, Spencer MJ (2004) Null mutation of calpain 3 (p94) in mice causes abnormal sarcomere formation in vivo and in vitro. Hum Mol Genet 13:1373–1388

Kuhnert F, Stuhlmann H (2004) Identifying early vascular genes through gene trapping in mouse embryonic stem cells. Curr Top Dev Biol 62:261–281

Lamia KA, Peroni OD, Kim YB, Rameh LE, Kahn BB, Cantley LC (2004) Increased insulin sensitivity and reduced adiposity in phosphatidylinositol 5-phosphate 4-kinase beta-/- mice. Mol Cell Biol 24:5080–5087

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Leighton PA, Mitchell KJ, Goodrich LV, Lu X, Pinson K, Scherz P, Skarnes WC, Tessier-Lavigne M (2001) Defining brain wiring patterns and mechanisms through gene trapping in mice. Nature 410:174–179

Li B, Dedman JR, Kaetzel MA (2003) Intron disruption of the annexin IV gene reveals novel transcripts. J Biol Chem 278:43276–43283

Li J, Zhu X, Chen M, Cheng L, Zhou D, Lu MM, Du K, Epstein JA, Parmacek MS (2005) Myocardin-related transcription factor B is required in cardiac neural crest for smooth muscle differentiation and cardiovascular development. Proc Natl Acad Sci U S A 102:8916–8921

Li L, Cohen SN (1996) Tsg101: a novel tumor susceptibility gene isolated by controlled homozygous functional knockout of allelic loci in mammalian cells. Cell 85:319–329

Liu J, Zhang L, Wang D, Shen H, Jiang M, Mei P, Hayden PS, Sedor JR, Hu H (2003) Congenital diaphragmatic hernia, kidney agenesis and cardiac defects associated with Slit3-deficiency in mice. Mech Dev 120:1059–1070

Mansour SL, Thomas KR, Capecchi MR (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature 336:348–532

Migliorini D, Denchi EL, Danovi D, Jochemsen A, Capillo M, Gobbi A, Helin K, Pelicci PG, Marine JC (2002) Mdm4 (Mdmx) regulates p53-induced growth arrest and neuronal cell death during early embryonic mouse development. Mol Cell Biol 22:5527–5538

Miller BC, Eckman EA, Sambamurti K, Dobbs N, Chow KM, Eckman CB, Hersh LB, Thiele DL (2003) Amyloid-beta peptide levels in brain are inversely correlated with insulysin activity levels in vivo. Proc Natl Acad Sci U S A 100:6221–6226

Niwa H AK, Kimura S, Taniguchi S, Wakasugi S, Yamamura K. J (1993) An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. Biochemistry (Tokyo) 113:343–349

Nonn L, Williams RR, Erickson RP, Powis G (2003) The absence of mitochondrial thioredoxin 2 causes massive apoptosis, exencephaly, and early embryonic lethality in homozygous mice. Mol Cell Biol 23:916–922

O'Kane CJGW (1987) Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A 84:9123–9127

Pederson BA, Chen H, Schroeder JM, Shou W, DePaoli-Roach AA, Roach PJ (2004) Abnormal cardiac development in the absence of heart glycogen. Mol Cell Biol 24:7179–7187

Ramirez-Solis R, Liu P, Bradley A (1995) Chromosome engineering in mice. Nature 378:720–724

Schnutgen F, de Zolt S, van Sloun P, Hollatz M, Floss T, Hansen J, Altschmied J, Seisenberger C, Ghyselinck NB, Ruiz P, Chambon P, Wurst W, von Melchner H (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A 102:7221–7226

Silver J, Keerikatte V (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J Virol 63:1924–1928

Skarnes WCAB, Joyner AL (1992) A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev 6:903–918

Skarnes WCMJ, Hurtley SM, Beddington RS (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. Proc Natl Acad Sci U S A 92:6592–6596

Steel M MJ, Clark KA, Kearns IR, Davies CH, Morris RG, Skarnes WC, Lathe R (1998) Gene-trapping to identify and analyze genes expressed in the mouse hippocampus. Hippocampus 8:444–457

Tanaka H, Iguchi N, Toyama Y, Kitamura K, Takahashi T, Kaseda K, Maekawa M, Nishimune Y (2004) Mice deficient in the axonemal protein Tektin-t exhibit male infertility and immotile-cilium syndrome due to impaired inner arm dynein function. Mol Cell Biol 24:7958–7964

Varmus H (1988) Retroviruses. Science 240:1427–1435

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C et al (2001) The sequence of the human genome. Science 291:1304–1351

Vidal F, Lopez P, Lopez-Fernandez LA, Ranc F, Scimeca JC, Cuzin F, Rassoulzadegan M (2001) Gene trap analysis of germ cell signaling to Sertoli cells: NGF-TrkA mediated induction of Fra1 and Fos by post-meiotic germ cells. J Cell Sci 114:435–443

Voss AK, Thomas T, Gruss P (1998) Efficiency assessment of the gene trap approach. Dev Dyn 212:171–180

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Wiles MV VF, Otte J, Fuchtbauer EM, Ruiz P, Fuchtbauer A, Arnold HH, Lehrach H, Metz T, von Melchner H, Wurst W (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat Genet 24:13–14

Wilson C, Pearson RK, Bellen HJ, O'Kane CJ, Grossniklaus U, Gehring WJ (1989) P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. Genes Dev 3:1301–1313

Xin HB, Deng KY, Shui B, Qu S, Sun Q, Lee J, Greene KS, Wilson J, Yu Y, Feldman M, Kotlikoff MI (2005) Gene trap and gene inversion methods for conditional gene inactivation in the mouse. Nucleic Acids Res 33:e14

Yang W, Musci TS, Mansour SL (1997) Trapping genes expressed in the developing mouse inner ear. Hear Res 114:53–61

You Y, Bergstrom R, Klemm M, Lederman B, Nelson H, Ticknor C, Jaenisch R, Schimenti J (1997) Chromosomal deletion complexes in mice by radiation of embryonic stem cells. Nat Genet 15:285–288

Yuan W, Rao Y, Babiuk RP, Greer JJ, Wu JY, Ornitz DM (2003) A genetic model for a central (septum transversum) congenital diaphragmatic hernia in mice lacking Slit3. Proc Natl Acad Sci U S A 100:5217–22

Yusa K, Horie K, Kondoh G, Kouno M, Maeda Y, Kinoshita T, Takeda J (2004) Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. Nature 429:896–899

Zambrowicz BP, Abuin A, Ramirez-Solis R, Richter LJ, Piggott J, BeltrandelRio H, Buxton EC, Edwards J, Finch RA, Friddle CJ, Gupta A, Hansen G, Hu Y, Huang W, Jaing C, Key BW Jr, Kipp P, Kohlhauff B, Ma ZQ, Markesich D, Payne R, Potter DG, Qian N, Shaw J, Schrick J, Shi ZZ, Sparks MJ, Van Sligtenhorst I, Vogel P, Walke W, Xu N, Zhu Q, Person C, Sands AT (2003) Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. Proc Natl Acad Sci U S A 100:14109–14114

Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392:608–611

Zheng B, Ho C, Li S, Keirstead H, Steward O, Tessier-Lavigne M (2003) Lack of enhanced spinal regeneration in Nogo-deficient mice. Neuron 38:213–224