Mutagenesis of diploid mammalian genes by gene entrapment

Qing Lin, Sarah L. Donahue, Tracy Moore-Jarrett, Shang Cao, Anna B. Osipovich and H. Earl Ruley*

Department of Microbiology and Immunology, Room AA4206, Medical Center North, Vanderbilt University, School of Medicine, 1161 21st Avenue South, AA4210, Nashville, TN 37232-2363, USA

ABSTRACT

The present study describes a genome-wide method for biallelic mutagenesis in mammalian cells. Novel poly(A) gene trap vectors, which contain features for direct cloning vector–cell fusion transcripts and for post-entrapment genome engineering, were used to generate a library of 979 mutant ES cells. The entrapment mutations generally disrupted gene expression and were readily transmitted through the germline, establishing the library as a resource for constructing mutant mice. Cells homozygous for most entrapment loci could be isolated by selecting for enhanced expression of an inserted neomycin-resistance gene that resulted from losses of heterozygosity (LOH). The frequencies of LOH measured at 37 sites in the genome ranged from 1.3×10^{-5} to 1.2×10^{-4} per cell and increased with increasing distance from the centromere, implicating mitotic recombination in the process. The ease and efficiency of obtaining homozygous mutations will (i) facilitate genetic studies of gene function in cultured cells, (ii) permit genome-wide studies of recombination events that result in LOH and mediate a type of chromosomal instability important in carcinogenesis, and (iii) provide new strategies for phenotype-driven mutagenesis screens in mammalian cells.

INTRODUCTION

The number and diversity of genes identified by the mammalian genome projects suggests that considerable biology remains to be characterized on a molecular level and has provided the impetus for developing genome-wide strategies to characterize gene functions important in normal and disease processes. Tagged sequence mutagenesis uses gene trap vectors to disrupt genes in cultured cells combined with rapid, DNA sequence-based screens to characterize the disrupted genes at the nucleotide level. The approach has been widely used to disrupt genes in mouse embryonic stem (ES) cells (1–3). Gene trapping has also been used, but to a far lesser extent, to identify genes responsible for recessive phenotypes in somatic cells (4–9).

Mutagenesis of mammalian cells is hindered by the fact that the autosomal genome is diploid and consequently, most entrapment mutations are recessive. The problem is circumvented by gene-based studies in ES cells where selected mutations can be transmitted through the mouse germline and subsequently bred to a homozygous state. However, gene inactivation in somatic cells requires pre-existing hemizygosity or spontaneous loss of heterozygosity; thus, even with strategies to enhance the recovery of loss-of-function mutations (4,5,7,10,11), gene trap mutagenesis has seen only limited use in phenotype-driven screens.

Mammalian cells heterozygous at a given locus undergo spontaneous conversion to a homozygous state by loss of heterozygosity (LOH) at frequencies of $\sim 10^{-5}$ per cell (12–15). Homozygous mutant cells can be selected based on phenotypes caused by gene dosage effects. For example, mutations involving the insertion of a neomycin-resistance gene (*Neo*) may be converted to a homozygous state simply by selecting for clones that survive in higher concentrations of G418 (16). Levels of neomycin resistance correlate with levels of *Neo* gene expression (17). Mitotic recombination, which appears to be the preferred mechanism for spontaneous LOH (13,14,18,19), generates cells homozygous for the inserted *Neo* genes, and the increase in *Neo* gene copy number allows moderately resistant cells to acquire higher levels of antibiotic resistance. However, unlike targeted mutations, LOH has not been reliably achieved with mutations induced by gene trapping. A major problem stems from variations in *Neo* gene expression that can result, e.g. when the entrapment cassette is expressed from different cellular promoters. In addition, some gene trap vectors use a 'corrected' version of the *Neo* gene that induces higher levels of antibiotic resistance than a widely used 'mutant' gene (20,21). Cells expressing the corrected *Neo* gene may be insufficiently G418-sensitive to select for LOH events (22). We reasoned that gene trap

---

*To whom correspondence should be addressed. Tel: +615 343 1379; Fax: +615 343 7392; Email: earl.ruley@vanderbilt.edu

mutations generated by a poly(A) trap (23–25) and that incorporate the less active *Neo* gene might be reproducibly converted to homozygosity, if the resulting mutant clones expressed similar, moderate levels of neomycin resistance. In the present study, new poly(A) trap vectors were developed in which gene trapping selects for inserted *Neo* sequences that splice to the 3′ ends of cellular genes. The vectors have additional features that permit vector–cell fusion transcripts to be directly cloned in *Escherichia coli* and that allow genes and chromosomes tagged by gene trapping to be engineered by DNA site-specific recombinases (26–31). The vectors are suitable for large-scale mutagenesis of mouse ES cells, and we show that most mutations selected from a stem cell library can be converted to a homozygous state following selection for higher levels of drug resistance. The ease and efficiency of obtaining homozygous entrapment mutations will (i) facilitate genetic studies of gene function in

cultured cells, (ii) permit genome-wide studies of recombination events that result in LOH and mediate a type of chromosomal instability important in carcinogenesis, and (iii) provide new strategies for phenotype-driven mutagenesis screens in mammalian cells.

RESULTS AND DISCUSSION

GTRx.x gene trap vectors (Figure 1a, Supplementary Data) function as 3′ gene [or poly(A)] traps (23–25,32). The virus inserts a *Neo* gene throughout the genome, and selection for G418 resistance generates clones in which *Neo* sequences splice to 3′ distal exons of cellular genes. The *Neo* gene in GTR1.x vectors was expressed from the promoter of the gene encoding RNA polymerase II (Pol2) (33), or in GTR2.x vectors from the PGK promoter (34). Expression of the
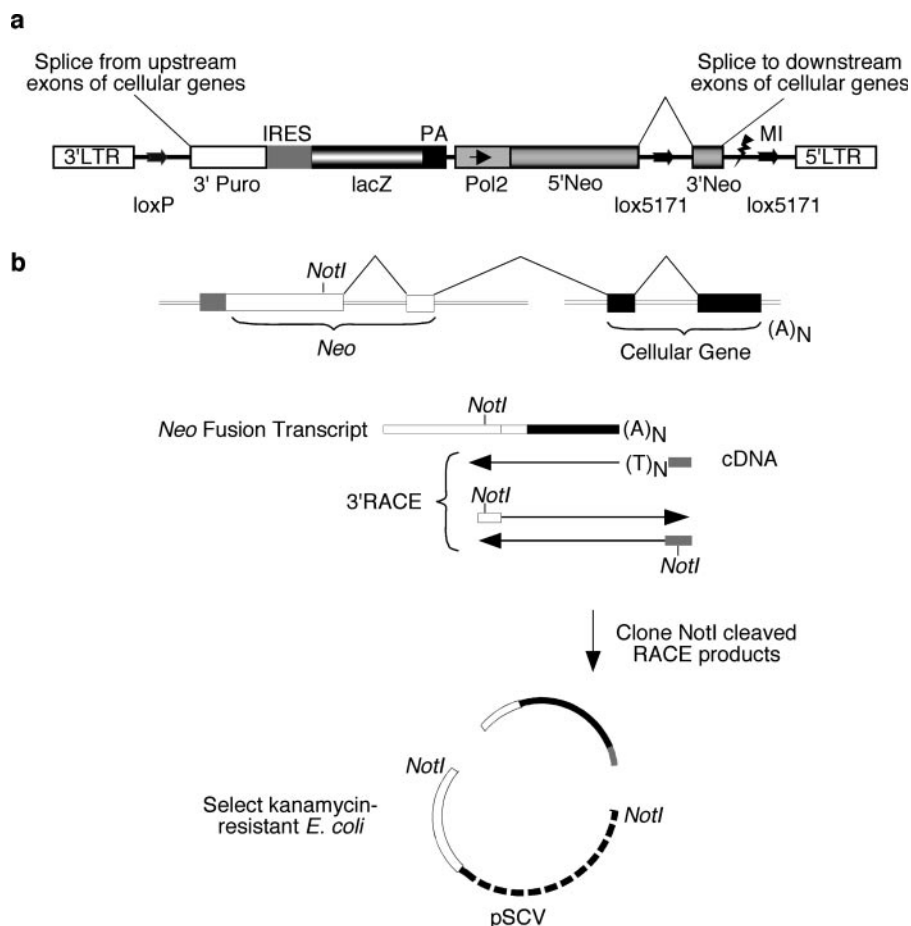

Figure 1. (**a**) Structures of the GTR retrovirus gene trap vectors. Expression of an intron-containing *Neo* gene (5′ Neo + 3′ Neo) carried by the GTR1.0 poly(A) trap vector selects for inserts in which the *Neo* gene, expressed from the promoter of the RNA polymerase 2 gene (Pol2), splices to downstream exons of cellular genes. Transcripts of occupied cellular genes splice to a 3′ exon [consisting of the 3′ end of a puromycin-resistance gene (3′ Puro), an internal ribosome entry site (IRES), a *lacZ* reporter and a polyadenylation site (PA)], disrupting their expression. A wild-type *loxP* site (*loxP*, left of 3′ Puro) and mutant *loxP* sites (*lox 5171*, on either side of the 3′ *Neo* exon) allow the body of the provirus to be replaced by other sequences by Cre-mediated cassette exchange. An RNA instability sequence (MI, flash symbol) increases the specificity of gene entrapment by reducing the levels of unspliced *Neo* transcripts. The positions of the provirus long terminal repeats (3′ and 5′ LTRs) are also indicated. Viruses lacking either the message instability sequence (pGTR1.3) or the *lox 5171* in the *Neo* intron (GTR1.2) or both elements (GTR1.1) have also been constructed. GTR1.4–1.7 are identical to GTR1.0–1.3 except they contain an enhanced green fluorescence (EGFP) reporter instead of *lacZ*. GTR2.0–2.3 are identical to GTR1.0–1.3 except *Neo* is expressed from the *PGK* promoter. Some elements are not drawn to scale to enhance clarity. (**b**) Direct cloning of 3′ RACE products. Gene entrapment by GTR vectors generates clones in which the *Neo* gene (white boxes) is expressed from transcripts that splice to downstream exons of cellular genes (black boxes). The intron and NotI endonuclease cleavage site in the *Neo* coding sequence ensure that only recombinant plasmids that contain cDNA inserts amplified from spliced *Neo*-cell fusion transcripts can give rise to kanamycin-resistant *E.coli*.

occupied cellular genes is disrupted by a 3′ exon consisting of sequences from the 3′ end of a puromycin-resistance gene, an internal ribosome entry site (IRES) and a reporter protein [either a nuclear β-galactosidase (*lacZ*; GTR1.0–GTR1.3) or enhanced green fluorescent protein (EGFP; GTR1.4–GTR1.7)]. Wild-type and mutant [*lox5171*, (35)] *loxP* sites allow provirus inserts to be engineered by recombinase-mediated cassette exchange (RMCE) (27–29). GTR1.0 and GTR1.3 contain an additional *loxP5171* site located in a synthetic intron inserted into the *Neo* gene (Supplementary Figure 1c). The 3′ *Puro* segment provides the 3′ end of a split puromycin-resistance gene and, when used in combination with the 5′ end of the gene, is designed to select for Cre-mediated inter- and intra-chromosomal recombination events, as has been described using a split *Hprt* gene (30,31). The split *Neo* gene in GTR1.0 and GTR1.3 can also be used for this purpose if the 3′ *Neo* exon is first deleted via recombination at the *lox5171* sites.

The *Neo* gene was engineered by inserting an intron and a NotI cleavage site by site-directed mutagenesis; neither modification affected the protein coding sequence (Supplementary Data and Supplementary Figure 1c). These features allow 3′ RACE products from spliced fusion transcripts to be cloned directly in *E.coli* (Figure 1b). Briefly, fusion transcripts amplified by 3′ RACE are cleaved with NotI and ligated to a plasmid (pSCV) containing *Neo* sequences upstream of the NotI site under the control of a strong bacterial promoter. Bacterial clones containing the desired RACE products are then selected on kanamycin plates. All steps in the process (expansion and cryopreservation of neomycin-resistant ES clones, RNA extraction, 3′ RACE, and DNA sequencing) were performed in a 96-well format. Fusion transcripts were cloned from 70–80% of ES cells grown in a single well without using nested PCR or highly competent *E.coli*, equivalent to the efficiency of cloning 3′ RACE products manually (28).

3′ RACE products from 979 ES cell clones were cloned, sequenced and compared against the mouse genome, EST and RefSeq databases using MultiBlaster, a relational database for performing BLAST searches on large numbers of DNA sequences (36). A total of 141 sequences contained repetitive DNA or were derived from probable sister clones (colonies from the same dish with identical 3′ RACE products). Of the remaining 838 sequences, 822 returned significant matches with sequences in the mouse genome sequence (MM). Most were derived from transcribed regions of the genome, as 568 matched sequences represented in the Expressed Sequence Tag (EST) database; 505, 371 and 337 matched entries for genes in the UniGene, RefSeq and Mouse Genome Informatics (MGI) databases, respectively (Figure 2). The UniGene, RefSeq and MGI matches correspond to genes with increasing degrees of characterization/annotation. Nearly all matches had *P*-values $<10^{-50}$, and none was $>10^{-20}$. Only 31% of the 3′ RACE products matched mouse genomic sequences (MGSCv3) for which there was no information to suggest that the provirus had inserted into a previously identified gene or transcription unit. Some of these inserts may reflect the presence of either new genes or additional exons that would extend the boundaries of adjacent transcription units. However, many probably correspond to cryptic or orphan 3′ exons, as suggested by
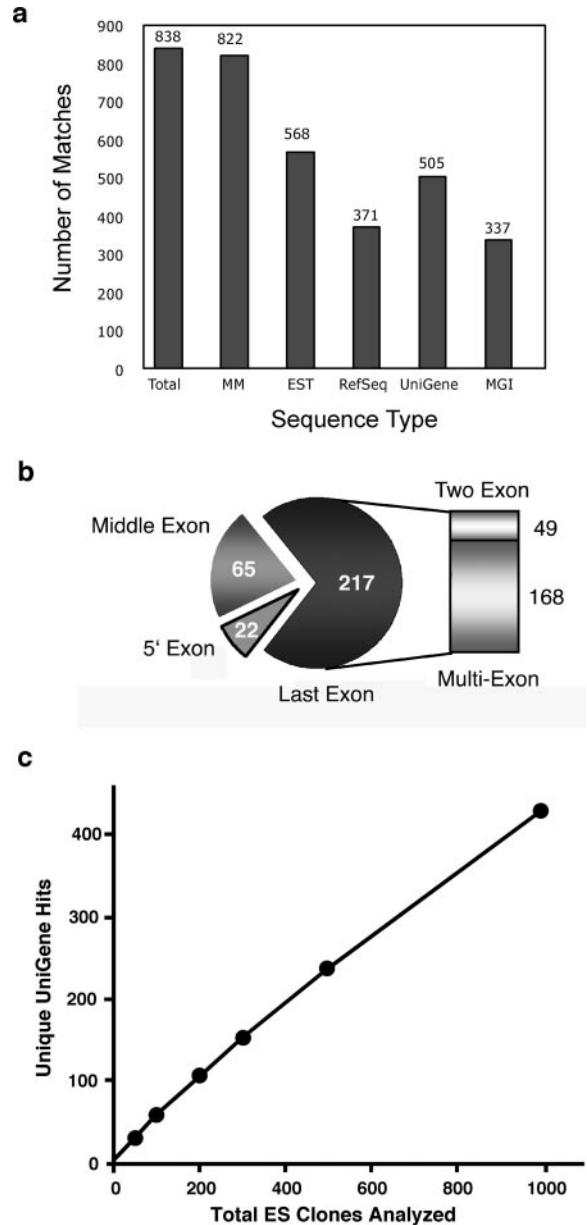

Figure 2. Tagged sequence mutagenesis with the GTR gene trap vectors. Composition of 3′ RACE sequences (**a**) 838 vector–fusion transcripts cloned by 3′ RACE matched sequences in the mouse genome (MM), expressed sequence tag (EST) and RefSeq databases, as shown. Matches corresponding to cellular genes as determined by UniGene and Mouse Genome Informatics entries (MGI) are also indicated. The positions of GTR inserts in cellular genes as deduced from the sequence of 3′ RACE products. (**b**) The majority (217) of inserts in well-characterized genes (304) spliced to the last exon of cellular genes; of these, 168 contained multiple exons and 49 contained two exons. (**c**) Disruption of novel gene sequences by GTR gene trap vectors. The number of inserts into unique genes (based on UniGene designations) is plotted against the total number of ES cell clones analyzed. All matches plotted on the *y*-axis involved non-repetitive sequences with *P*-values $< 10^{-20}$. ES cell clones plotted on the *x*-axis included sister clones from the same dish and clones from which 3′ RACE products contained repetitive (non-informative) DNA sequences.

31 RACE products in which the sequence is derived from a well-characterized gene but in the opposite transcriptional orientation from the occupied gene. The task of distinguishing between genes and cryptic/orphan 3′ exons is complicated

in some cases by widespread examples of anti-sense transcripts in genes, spliced ESTs that are not incorporated into the structural models of nearby genes and abundant cytoplasmic poly(A)-containing RNAs that do not match annotated gene sequences. For comparison, 56% of cytoplasmic polyadenylated RNAs do not contain annotated exon or intron sequences (37).

The GTR vectors preferentially targeted the last intron and expressed fusion transcripts that spliced to a single downstream exon (Figure 2b). This provides further evidence that poly(A) traps select for fusion transcripts that evade nonsense-mediated decay (28,38). However, the preference was less pronounced, as 29% of the inserts in well-characterized genes were in upstream introns.

As with other gene trap vectors, gene targeting was not random (2,39): 15% of the unique UniGene sequences were targeted multiple times in clones derived from independent culture dishes, representing potential preferred sites for virus integration and/or gene trap selection. These include 7 inserts in 160010M07Rik by different GTR vectors and 11 inserts in *CDC57* (*4933434G05Rik*), all by GTR2.3. The overall accumulation of gene trap inserts into unique UniGene clusters (Figure 2c) as a function of the total ES cell clones analyzed (including sister clones and non-informative 3′ RACE products with repetitive sequences) is similar to that reported for other gene trap vectors (39); however, the number of mutations analyzed is too small to predict eventual genome coverage (2,40).

Clones from the entrapment library were highly germline competent as all 10 entrapment loci that have been tested to date were readily transmitted through the germline (data not shown). While detailed analysis of the mutant mice is outside the scope of the present study, the GTR vectors appeared to be effective mutagens: 3 of 9 inserts into annotated genes induced obvious phenotypes when bred to a homozygous state. Specifically, an insert into *Hesx1* produced similar defects in eye development to those described for a targeted null mutation (41); the *Dymeclin* mutation caused defects in bone growth similar to defects observed in humans (42); and animals homozygous for the *Pfdn1* mutation die within 5 weeks of age. In all cases examined (*Cradd*, *Dymeclin* and *Pfdn1*), entrapment mutagenesis significantly ablated expression of the occupied allele (Figure 3) even though all occurred in the last intron of the gene. Thus, the entrapment library provides a resource from which mutations in genes of interest can be selected for transmission into the germline. The mutations have been contributed to the International Gene Trap Consortium [IGTC, (2)], and the 3′ RACE sequences have been submitted to the GenBank GSS (Genome Survey Sequence, http://www.ncbi.nlm.nih.gov/dbGSS/index.html) database (Accession numbers: CZ169539–CZ170518). Specific mutations can be identified by searching the GSS database with cDNA sequences of the genes of interest. Matching database entries provide information about the mutation, gene trap vector and contact information for obtaining the ES cell clone. As with any gene trap mutation, the occupied allele should first be evaluated with regard to the position and orientation of the vector within the occupied gene, e.g. to avoid inserts where the vector is in the wrong transcriptional orientation or is located downstream (3′) of protein coding sequences.
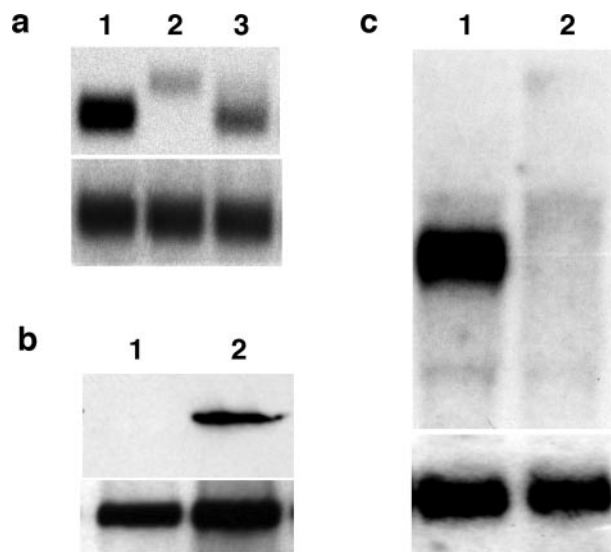

Figure 3. Loss of occupied gene expression in homozygous mutant cells and tissues. *Pfdn1* expression (**a**, upper panel) in embryonic fibroblast cells from wild-type (lane 1), homozygous mutant (lane 2) and heterozygous (lane 3) fibroblasts was analyzed by northern blot analysis. The blot was stripped and probed with a GAPDH sequence as a loading control (lower panel). Cradd protein expression (**b**, upper panel) in primary spleen cells from homozygous mutant mice (lane 1) and wild type (lane 2) was assessed by western blot analysis. As a loading control, the blots were stripped and analyzed using an anti-β-actin antibody (lower panel). *Dymeclin* expression (**c**, upper panel) in liver tissue from wild type (lane 1) and homozygous mutant mice (lane 2) was analyzed by northern blot analysis. Hybridization to a GAPDH probe (lower panel) provides a loading control.

We tested whether homozygous mutants could be selected from heterozygous gene trap clones and examined the influence of chromosome location on the frequencies of LOH. A total of 37 randomly selected clones with mutations (Supplementary Figure 2 and Supplementary Table 1) induced by the GTR1.3 vector were placed in media containing 2.0 mg/ml G418 to select potential clones that had undergone spontaneous LOH. The frequencies of resistance to 2.0 mg/ml G418 ranged from 1.3×10^{-5} to 1.2×10^{-4}. Thus, all 37 clones with inserted GTR1.3 proviruses expressed degrees of neomycin sensitivity that are required in order to select for homozygous mutations generated by gene targeting (16,19). However, clones containing the GTR2.3 vector (in which *Neo* is expressed from the PGK promoter) displayed far greater resistance to G418 (Supplementary Table 1), even at 3.0 mg/ml G418 (data not shown). All of the GTR vectors were constructed using a 'mutant' *Neo* gene known to induce lower levels of antibiotic resistance than a 'corrected' variant (20). Nevertheless, the intrinsic neomycin sensitivity of cells with GTR2.3 proviruses was insufficient to select for potential LOH events that occur at per-cell frequencies of $\sim 10^{-5}$ (12–16,19). Levels of neomycin resistance correlate with levels of *Neo* gene expression (17), and relatively low levels of resistance are necessary in order to select for LOH events (22). Therefore, use of the Pol2 promoter, which is four times weaker than the PGK promoter (28), appears to be an important variable with regard to the reliable selection of potential homozygous mutant entrapment clones.
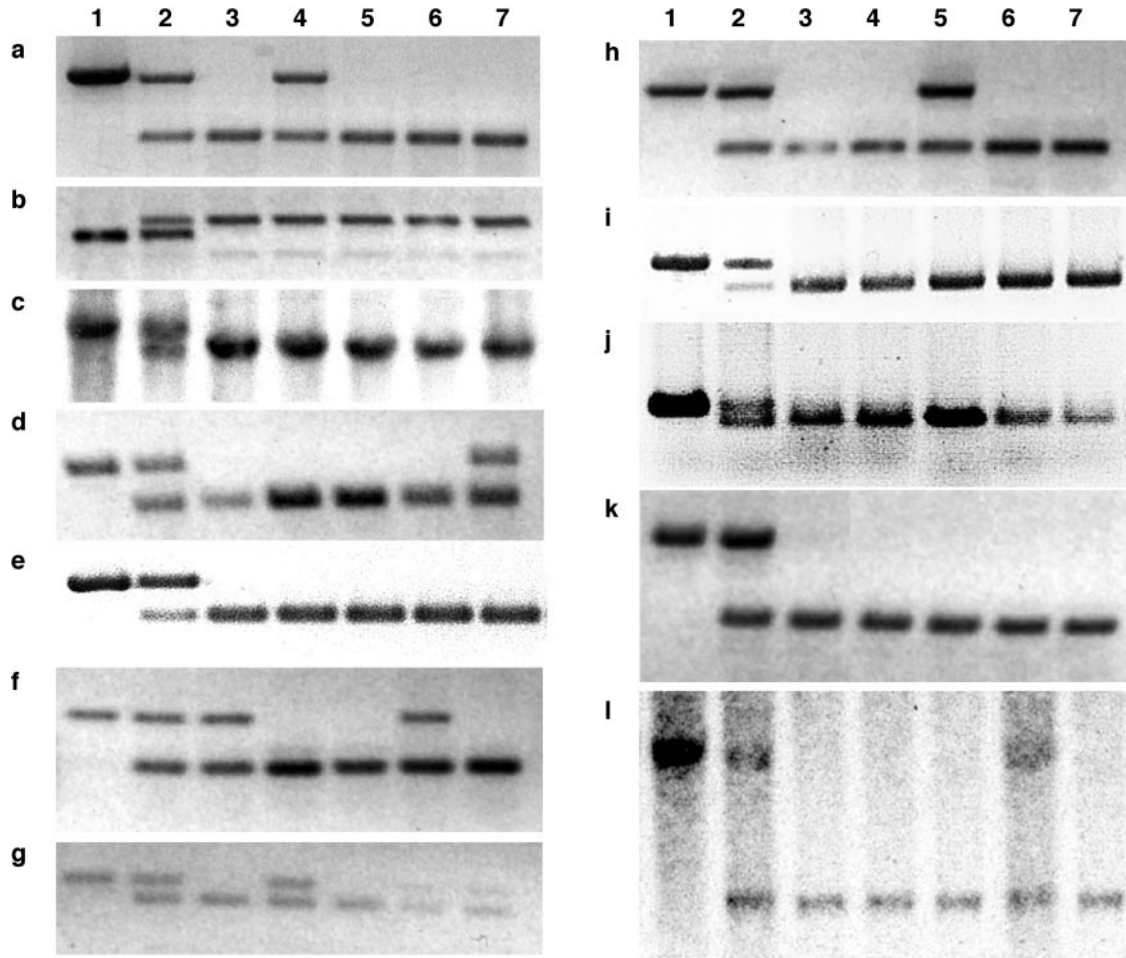
Figure 4. LOH at entrapment loci following selection in 2.0 mg/ml G418. DNAs from the parental ES cells (lane 1), heterozygous mutant entrapment clones (lane 2) and clones isolated following selection in 2.0 mg/ml G418 (lanes 3–7) were genotyped by either Southern blot hybridization (j, l) or PCR (a–i, k) using gene-specific probes and primers. The mutant clones contained entrapment vectors inserted in or near (*) the following genes: *Ttc18* (**a**) *Rbm4* (**b**), *IL8Ra** (**c**), *M11t10* (**d**), *Cradd* (**e**), *1700012G11Rik* (**f**), *Rnf7* (**g**), *LOC240669* (**h**), *Hesx1* (**i**), *1810030N24Rik* (**j**), *Dscr1l1* (**k**) and *Xrcc5* (**l**). Bands corresponding to the wild-type allele migrate more slowly in all cases except (b) than the allele disrupted by gene entrapment. LOH is indicated by the absence of the wild-type allele in cells selected in high G418 media.

Genotypic analysis of 12 different mutants induced by GTR1.3 confirmed that a significant proportion (82% overall) of the colonies surviving in 2.0 mg/ml G418 had undergone LOH (Figure 4). Cells homozygous for all 12 entrapment loci tested were recovered at frequencies ranging from 40% to 100% of the high G418-resistant colonies selected from each clone. The 12 entrapment clones were randomly selected based on the availability of flanking sequence probes and primers capable of distinguishing the wild-type and occupied alleles. Since entrapment clones generated by GTR1.3 produced colonies resistant to 2.0 mg/ml G418 at similar frequencies (Supplementary Table 1), and a high proportion of the high G418-resistant colonies analyzed from each tested clone had undergone LOH, we conclude that cells homozygous for most mutations in the gene trap library can be selected in 2.0 mg/ml G418.

In contrast, we have encountered only one GTR1.3 entrapment clone for which homozygous mutant ES cells could not be isolated. These experiments (Cao *et al*., unpublished) were performed as part of a separate study to characterize a mutation in *Pfdn1* (prefoldin), a chaperone that assists in the folding of cytoskeletal proteins (43). Briefly, cells heterozygous for the *Pfdn1* mutation formed colonies in 2.0 mg/ml G418 at frequencies (5.6×10^{-6}) more than 10-fold lower than observed with other GTR1.3 entrapment clones and similar to the frequencies observed with ES cells heterozygous for a targeted mutation in *Ssrp1* (data not shown), which encodes a chromatin remodeling protein essential for ES cell viability (44). Moreover, the colonies arising in high G418 were small, and the cells could not be propagated further. Therefore, prefoldin appears to be required for the clonal outgrowth of ES cells, accounting for the failure to isolate homozygous mutant cells.

The frequency of colony formation in high G418 increased with increasing distance from the centromere (Figure 5). The R^2 by linear regression analysis was 0.54 for all genes in aggregate (Figure 5a) and 0.78 for the eight loci on chromosome 4 considered separately (Figure 5b). The influence of chromosome position was observed despite many potential variables that could influence the induction or recovery of clones with LOH [e.g. levels of *Neo* expression, DNA sequence effects on recombination, or clonal variation in
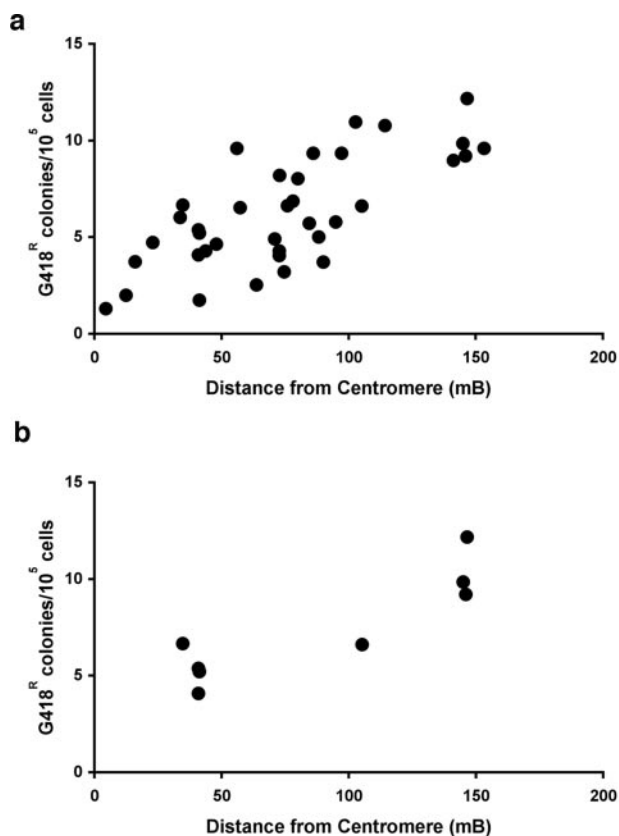
Figure 5. The frequency of presumptive LOH at sites throughout the genome increases with distance from the centromere. A total of 37 ES cell clones, each containing a single GTR1.3 gene trap vector, were placed in media containing 2.0 and 0.3 mg/ml G418. The frequency of colony formation (Presumptive LOH) in the higher concentration of G418 (normalized to the number of colonies at the lower concentration) is plotted against the distance of each mutation from the centromere. The average values for three independent experiments are plotted for all 37 entrapment loci (**a**) and for all 8 inserts located on chromosome 4 (**b**). Linear regression analyses of the two groups produced R-squared values of 0.54 and 0.78, respectively. The SD values, which were 1060% of the average values, have been omitted for clarity.

Figure 6. Loss of *Xrcc5* expression in homozygous entrapment clones. (**a**) RNA was isolated from the parental AC1 ES cells (lane 1), the heterozygous *Xrcc5* entrapment mutant before selection in 2.0 mg/ml G418 (lane 2) and clones isolated by selection in 2.0 G418 (lanes 3–6) and northern blot analysis was performed using *Xrcc5* (downstream of exon 1 cloned by 3′RACE) and β-actin specific probes (upper and lower panels, respectively). *Neo–Xrcc5* fusion transcripts in heterozygous and homozygous mutant cells are generated by splicing of the *Neo* sequences to exon 2 of the *Xrcc5* gene. (**b**) Radiation sensitivity of *Xrcc5* heterozygous and homozygous mutant cells. Parental ES cells (1), heterozygous *Xrcc5* entrapment mutant (2) homozygous *Xrcc5* entrapment mutants homozygous *Xrcc5* entrapment mutants (3, 5, 6) a control *Xrcc6*-deficient Chinese hamster ovary (CHO) cell line (4) and a radiation-sensitive CHO cell line (Xrs5) defective in Ku86 were exposed to increasing doses of γ-irradiation, and cell survival was measured in a clonogenic assay.

plating efficiencies ranging from 30% to 50% (data not shown)]. The chromosome position effect suggests that mitotic recombination plays a significant role in spontaneous LOH, consistent with previous observations involving *Neo* genes inserted by gene targeting in ES cells (19) and the *APRT* gene in other cell types (13,14,18).

GTR1.3 represents the first insertional mutagen capable of biallelic mutagenesis of mammalian genes. However, like other early-generation poly(A) traps, the vector preferentially targets 3′ introns, reflecting selection for fusion transcripts that escape nonsense-mediated decay (28,38). This interferes with the mutagenesis of a subset of genes that lack significant protein coding sequence in the last exon and may limit genome coverage (40). Inclusion of an IRES in the poly(A) trap cassette greatly enhances the recovery of inserts positioned throughout genes and increases the expression of fusion transcripts that append multiple, downstream exons (38). The IRES also appears to decrease the proportion of trapping events that splice to cryptic 3′ exons, probably by increasing the size of the region within genes compatible with gene
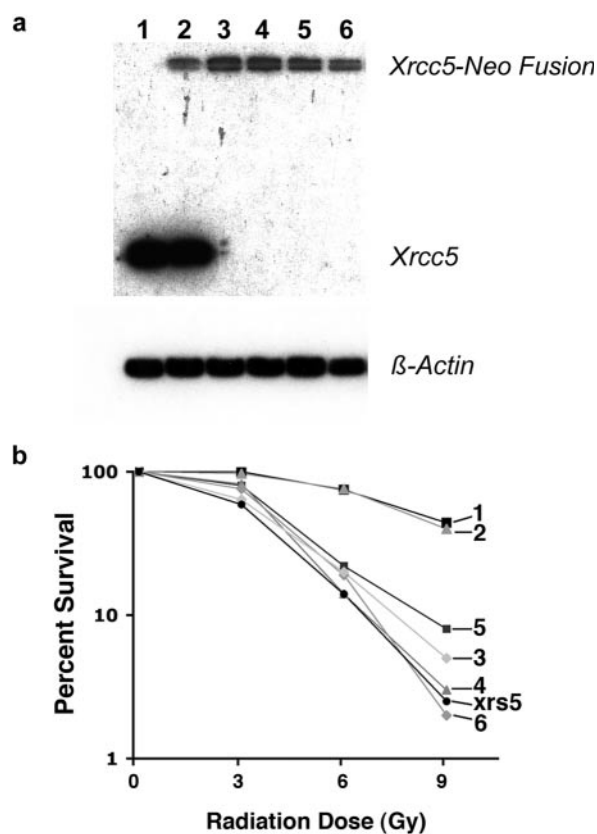
trap selection. Alternatively, small, non-coding poly(A) trap cassettes may be incorporated into a 5′ gene trap vector (40).

The ease and efficiency of obtaining homozygous entrapment mutations will enhance the utility of mutant ES cell libraries in several ways. First, cells deficient in any gene of interest (assuming the gene is expressed in ES cells and is not required for cell viability) can be readily obtained for biochemical and metabolic studies of gene function, without the time or expense of introducing the mutation into the germline. For example, LOH involving a mutation in the *Xrcc5* gene (Figure 4l) resulted in a complete loss of *Xrcc5* expression in the homozygous mutant cells (Figure 6a). As expected the *Xrcc5*-deficient cells were also hypersensitive to γ-irradiation as compared with the parental or heterozygous mutant ES cells or cells defective in either *Xrcc6* or *Ku86* that display radiation-sensitive phenotypes (Figure 6b). Second, analysis of homozygous mutants will be useful in assessing whether gene entrapment has induced a null mutation as illustrated by the *Xrcc5* mutation

(Figure 6a). Third, the ability to select homozygous mutant cells will provide an early assessment of whether the disrupted genes or chromosome deletions engineered post-entrapment are required for cell viability. Fourth, allelic imbalance is a common manifestation of chromosome instability in human cancers, which may harbor >10 000 regions of LOH per cell (45). The source of this genome-wide LOH is unknown and unfortunately, frequencies of LOH at specific sites are typically measured at relatively few loci (e.g. *Tk*, *Aprt*, *Hprt* or cell surface antigens) where gene inactivation confers a selectable phenotype. Entrapment ES cell clones provide resources to study factors, such as carcinogens (46), localized elements in the genome or mutations in genes required for genome maintenance, that influence the frequencies of LOH at many sites throughout the genome.

Finally, LOH involving GTR poly(A) traps will assist phenotype-driven mutagenesis screens in mammalian cells (6,7,47). Mutagens incorporating Pol2Neo as an LOH selection cassette, should facilitate the recovery of homozygous mutants by combining selection for high G418 resistance together with strategies that enhance the frequencies of LOH (7,46,47). Alternatively, since LOH involving inserted *Neo* resistance genes extend across large chromosome regions (19), stem cell clones containing inserted GTR1.3 vectors can be used to enhance the recovery of homozygous recessive mutants located on the same chromosome as the gene trap vector. This provides an alternative to the use of site-specific recombinases to induce mitotic recombination (10,11), eliminating the need to insert recombinase target sequences at allelic sites in the genome.

METHODS

Gene trapping

AC1 mouse ES cells were derived from an explanted 129svJ blastocyst cultured on feeder layers of irradiated mouse embryo fibroblasts and were cultured as described previously (48).

Construction of GTR poly(A) trap vectors is described in the Supplementary Data. Retroviruses were prepared by transfecting GTR plasmids into Phoenix Eco cells by calcium phosphate coprecipitation. Virus production by individual clones was titered as Neo^R c.f.u. in NIH3T3 cells (28). Supernatants from producer lines with titers of 200 Neo^R c.f.u./ml were used to infect AC1 ES cells: 24 h post-infection, the cells were placed in selective media containing 250 μg/ml of G418 (Invitrogen) and cultured for 7 days during which the media was changed every day. Individual G418-resistant colonies were transferred to a single well of a 96-well plate. After 3–5 days, the plates were passaged to three 96-well plates and grown for an additional 2–3 days. One plate was used to prepare RNA for 3′ RACE and two plates were cryopreserved in liquid nitrogen.

Identification of genes disrupted by gene entrapment

Disrupted genes were identified by sequencing cloned *Neo* fusion transcripts amplified by 3′ RACE. Total RNA was extracted using the RNeasy96 system (Qiagen Inc, Valencia, CA) according to the manufacturer's instructions. cDNA was

synthesized using the superscript II reverse transcrtiptase (Invitrogen) in a 20-μl reaction containing 1 μg of total RNA and a NotI-adaptor-oligo-dT primer (5′-GACTAAC-CCGGCTCGAGCGGCCGCTTTTTTTTTTTTTTTTTTT-3′). The cDNA was then amplified by two rounds of PCR in a 50-μl reaction using the hotstart *Taq* polymerase kit (Qiagen). The PCR contained 2 μl of the above cDNA product with a *Neo*-specific primer (5′-ATGGCCGCTTTTCTGGATTCAT-CG-3′) and NotI adaptor primer (5′-GACTAACCCGGCT-CGAGCGGCCGCT-3′). All the reactions were performed in 96-well plates. PCR products were purified using QIAquick 96 cartridges (Qiagen) and digested with NotI for 1 h and then purified again over QIAquick 96.

NotI-digested 3′ RACE products were ligated together with the selective cloning vector (pSCV, see Supplementary Data) and transferred into chemically competent DH5α *E.coli*. The transformed cells were cultured for 2 h in Luria–Bertani (LB) media and plated on LB agar plates containing 25 μg/ml kanamycin. Individual colonies were picked and grown in 96-well ml LB media overnight. Plasmids were prepared with QIAprep 96 Miniprep cartridges (Qiagen) and sequenced (28), using a *Neo*-specific primer: TCCCGATTC-GCAGCGCATCGCC. It should be noted that pSCV incorporates the 'corrected' *Neo* sequence (20). Selective cloning vectors incorporating the less active variant formed far fewer colonies on 25 μg/ml kanamycin plates (Q.L., unpublished). When entrapment clones were grown on *Neo*-resistant feeder layer cells, 3′ RACE also generated small inserts, of 110 nt in size, generated by recombination between the fusion transcripts (which contain a NotI site) and *Neo* transcripts expressed by the feeder cells. To eliminate this background, plasmids were pre-screened to identify clones with larger RACE products.

Isolation of flanking genomic DNA

Flanking genomic DNA sequences were cloned by inverse PCR (49) or by ligation-mediated PCR (50) modified by the addition of a C3 spacer (Integrated DNA Technologies) to the NlaIII minus adaptor to block the amplification of fragments via adaptor primers alone. Briefly, genomic DNA was (i) digested with NlaIII, ligated to a 1:1 mixture of NlaIII plus (5′-GTAATACGACTCACTATAGGGCTCCGCTTAAGGG-ACCATG-3′) and minus (5′-Phos-GTCCCTTAAGCGGAG-C3 spacer) strand adaptors, (ii) digested with PstI to prevent amplification of sequences from the 5′ LTR and (iii) subjected to two rounds (30 cycles each) of PCR using nested primers to the LTR and adaptor sequences.

First round PCR, LTR: 5′-GCTAGCTTGCCAAACCTA-CAGGTGG-3′

Adaptor: 5′-GTAATACGACTCACTATAGGGCTCCG-3′
Second round PCR, LTR: 5′-CCAAACCTACAGGTG-GGGTCTTTC-3′

Adaptor: 5′-AGGGCTCCGCTTAAGGGAC-3′
Regions of flanking genomic DNA downstream of the provirus in the *Ttc18*, *Rbm4*, *M11t10*, *Cradd*, *1700012G11Rik*, *Rnf7*, *LOC240669*, *Hesx1* and *Dscr1l1* genes or near the *IL8Ra* gene were sequenced, and the sequences have been submitted to the GenBank GSS database (accession numbers, DX977440–DX977449, respectively). The sequences of primers used for genotyping are available on request.

Selection and analysis of LOH

Serially diluted cells were plated in triplicate onto 150 mm plates and allowed to attach overnight. Subsequently, unattached cells were removed and selection media containing either 0.0, 0.3 or 2.0 mg/ml G418 was added to each dish. After 12 days, the number of colonies surviving in each dish was counted, and the frequency of colony formation at 2.0 mg/ml G418 was determined by dividing the number of colonies obtained from 0.3 mg/ml G418 selection to that obtained from 2.0 mg/ml G418 selection.

The genotypes of clones surviving in 2.0 mg/ml G418 were determined by Southern blot and PCR analysis. For Southern blot analysis, 5 μg of endonuclease-cleaved DNA was fractionated on 0.9% (w/v) agarose gels and hybridized to probe genomic DNA sequences adjacent to the entrapment vector. For PCR analysis, 200 ng of genomic DNA was amplified using two primers complementary to genomic DNA located on either side of the site of provirus insertion and one primer specific for the entrapment vector.

Loss of Cradd expression in homozygous mutant cells was determined by western blot analysis using a rabbit polyclonal antibody against a region containing Leu-97 of the human CRADD/RAIDD protein (catalog number 4899, Cell Signaling Technology, Inc.)

Analysis of *Xrcc5* entrapment clones

Serially diluted cells were plated in triplicate onto 150 mm plates and allowed to attach overnight. Subsequently, cells were irradiated in culture medium at a dose rate of 3 Gy/min (200 kV, 4 mA, 0.78 mm A1). Colonies were counted at 12 days after irradiation, and the percentage of surviving cells was determined relative to the numbers of colonies from untreated cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Stryke,D., Kawamoto,M., Huang,C.C., Johns,S.J., King,L.A., Harper,C.A., Meng,E.C., Lee,R.E., Yee,A., L'Italien,L. *et al.* (2003) BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. *Nucleic Acids Res.*, **31**, 278–281.

2. Skarnes,W.C., von Melchner,H., Wurst,W., Hicks,G., Nord,A.S., Cox,T., Young,S.G., Ruiz,P., Soriano,P., Tessier-Lavigne,M. *et al.* (2004) A public gene trap resource for mouse functional genomics. *Nat. Gen.*, **36**, 543–544.

3. Austin,C.P., Battey,J.F., Bradley,A., Bucan,M., Capecchi,M., Collins,F.S., Dove,W.F., Duyk,G., Dymecki,S., Eppig,J.T. *et al.* (2004) The knockout mouse project. *Nat. Genet.*, **36**, 921–924.

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

5. Hubbard,S.C., Walls,L., Ruley,H.E. and Muchmore,E.A. (1994) Generation of Chinese hamster ovary cell glycosylation mutants by retroviral insertional mutagenesis: integration into a discrete locus generates mutants expressing high levels of N-glycolyneuraminic acid. *J. Biol. Chem.*, **269**, 3717–3724.

6. Organ,E.L., Sheng,J., Ruley,H.E. and Rubin,D.H. (2004) Discovery of mammalian genes that participate in virus infection. *BMC Cell Biol.*, **5**, 41.

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

8. Sheng,J., Organ,E.L., Hao,C., Wells,K.S., Ruley,H.E. and Rubin,D.H. (2004) Mutations in the IGF-II pathway that confer resistance to lytic reovirus infection. *BMC Cell Biol.*, **5**, 32.

9. Murray,J.L., Mavrakis,M., McDonald,N.J., Yilla,M., Sheng,J., Bellini,W.J., Zhao,L., Le Doux,J.M., Shaw,M.W., Luo,C.C. *et al.* (2005) Rab9 GTPase is required for replication of human immunodeficiency virus type 1, filoviruses, and measles virus. *J. Virol.*, **79**, 11742–11751.

10. Koike,H., Horie,K., Fukuyama,H., Kondoh,G., Nagata,S. and Takeda,J. (2002) Efficient biallelic mutagenesis with Cre/loxP-mediated inter-chromosomal recombination. *EMBO Rep.*, **3**, 433–437.

11. Liu,P., Jenkins,N.A. and Copeland,N.G. (2002) Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells. *Nat. Genet.*, **30**, 66–72.

12. Morley,A.A. (1991) Mitotic recombination in mammalian cells in vivo. *Mutat. Res.*, **250**, 345–349.

13. Shao,C., Deng,L., Henegariu,O., Liang,L., Raikwar,N., Sahota,A., Stambrook,P.J. and Tischfield,J.A. (1999) Mitotic recombination produces the majority of recessive fibroblast variants in heterozygous mice. *Proc. Natl Acad. Sci. USA*, **96**, 9230–9235.

14. Gupta,P.K., Sahota,A., Boyadjiev,S.A., Bye,S., Shao,C., O'Neill,J.P., Hunter,T.C., Albertini,R.J., Stambrook,P.J. and Tischfield,J.A. (1997) High frequency in vivo loss of heterozygosity is primarily a consequence of mitotic recombination. *Cancer Res.*, **57**, 1188–1193.

15. Wijnhoven,S.W., Kool,H.J., van Teijlingen,C.M., van Zeeland,A.A. and Vrieling,H. (2001) Loss of heterozygosity in somatic cells of the mouse. An important step in cancer initiation? *Mutat. Res.*, **473**, 23–36.

16. Mortensen,R.M., Conner,D.A., Chao,S., Geisterfer-Lowrance,A.A. and Seidman,J.G. (1992) Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell Biol.*, **12**, 2391–2395.

17. Paludan,K., Duch,M., Jorgensen,P., Kjeldgaard,N.O. and Pedersen,F.S. (1989) Graduated resistance to G418 leads to differential selection of cultured mammalian cells expressing the neo gene. *Gene*, **85**, 421–426.

18. Van Sloun,P.P., Wijnhoven,S.W., Kool,H.J., Slater,R., Weeda,G., van Zeeland,A.A., Lohman,P.H. and Vrieling,H. (1998) Determination of spontaneous loss of heterozygosity mutations in Aprt heterozygous mice. *Nucleic Acids Res.*, **26**, 4888–4894.

19. Lefebvre,L., Dionne,N., Karaskova,J., Squire,J.A. and Nagy,A. (2001) Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nat. Genet.*, **27**, 257–258.

20. Yenofsky,R.L., Fine,M. and Pellow,J.W. (1990) A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proc. Natl Acad. Sci. USA*, **87**, 3435–3439.

21. Skarnes,W.C., Moss,J.E., Hurtley,S.M. and Beddington,R.S. (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA*, **92**, 6592–6596.

22. Mortensen,R.M. (2000) Production of a homozygous mutant embryonic stem cell line (Double Knockout). In Ausubel,F.M., Kingston,R.E., Brent,R., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*. John Wiley and Sons Inc., New York, p. Supplement 52.

23. Ishida,Y. and Leder,P. (1999) RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucleic Acids Res.*, **27**, e35.

24. Yoshida,M., Yagi,T., Furuta,Y., Takayanagi,K., Kominami,R., Takeda,N., Tokunaga,T., Chiba,J., Ikawa,Y. and Aizawa,S. (1995) A new strategy of gene trapping in ES cells using 3'RACE. *Transgenic Res.*, **4**, 277–287.

25. Zambrowicz,B., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L., Person,C. and Sands,A.T. (1998) Disruption and sequence identification of 2000 genes in mouse embryonic stem cells. *Nature*, **392**, 608–611.

26. Hardouin,N. and Nagy,A. (2000) Gene-trap-based target site for cre-mediated transgenic insertion. *Genesis*, **26**, 245–252.

27. Araki,K., Imaizumi,T., Sekimoto,T., Yoshinobu,K., Yoshimuta,J., Akizuki,M., Miura,K., Araki,M. and Yamamura,K. (1999) Exchangeable gene trap using the Cre/mutated lox system. *Cell Mol. Biol. (Noisy-le-grand)*, **45**, 737–750.

28. Osipovich,A.B., Singh,A. and Ruley,H.E. (2005) Post-entrapment genome engineering: first exon size does not affect the expression of fusion transcripts generated by gene entrapment. *Genome Res.*, **15**, 428–435.

29. Cobellis,G., Nicolaus,G., Iovino,M., Romito,A., Marra,E., Barbarisi,M., Sardiello,M., Di Giorgio,F.P., Iovino,N., Zollo,M. *et al.* (2005) Tagging genes with cassette-exchange sites. *Nucleic Acids Res.*, **33**, e44.

30. Zheng,B., Sage,M., Cai,W.W., Thompson,D.M., Tavsanli,B.C., Cheah,Y.C. and Bradley,A. (1999) Engineering a mouse balancer chromosome. *Nat. Genet.*, **22**, 375–378.

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

32. Salminen,M., Meyer,B.I. and Gruss,P. (1998) Efficient poly A trap approach allows the capture of genes specifically active in differentiated embryonic stem cells and in mouse embryos. *Dev. Dyn.*, **212**, 326–333.

33. Matsuda,E., Shigeoka,T., Iida,R., Yamanaka,S., Kawaichi,M. and Ishida,Y. (2004) Expression profiling with arrays of randomly disrupted genes in mouse embryonic stem cells leads to in vivo functional analysis. *Proc. Natl Acad. Sci. USA*, **101**, 4170–4174.

34. Adra,C.N., Boer,P.H. and McBurney,M.W. (1987) Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter. *Gene*, **60**, 65–74.

35. Lee,G. and Saito,I. (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene*, **216**, 55–65.

36. Osipovich,A.B., White-Grindley,E.K., Hicks,G.G., Roshon,M.J., Shaffer,C., Moore,J.H. and Ruley,H.E. (2004) Activation of cryptic 3′ splice sites within introns of cellular genes following gene entrapment. *Nucleic Acids Res.*, **32**, 2912–2924.

37. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.

38. Shigeoka,T., Kawaichi,M. and Ishida,Y. (2005) Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells. *Nucleic Acids Res.*, **33**, e20.

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

40. Zambrowicz,B.P., Abuin,A., Ramirez-Solis,R., Richter,L.J., Piggott,J., BeltrandelRio,H., Buxton,E.C., Edwards,J., Finch,R.A., Friddle,C.J. *et al.* (2003) Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl Acad. Sci. USA*, **100**, 14109–14114.

41. Dattani,M.T., Martinez-Barbera,J.P., Thomas,P.Q., Brickman,J.M., Gupta,R., Martensson,I.L., Toresson,H., Fox,M., Wales,J.K., Hindmarsh,P.C. *et al.* (1998) Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat. Genet.*, **19**, 125–133.

42. Paupe,V., Gilbert,T., Le Merrer,M., Munnich,A., Cormier-Daire,V. and El Ghouzzi,V. (2004) Recent advances in Dyggve-Melchior-Clausen syndrome. *Mol. Genet. Metab.*, **83**, 51–59.

43. Vainberg,I.E., Lewis,S.A., Rommelaere,H., Ampe,C., Vandekerckhove,J., Klein,H.L. and Cowan,N.J. (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell*, **93**, 863–873.

44. Cao,S., Bendall,H., Hicks,G.G., Nashabi,A., Sakano,H., Shinkai,Y., Gariglio,M., Oltz,E.M. and Ruley,H.E. (2003) The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. *Mol. Cell Biol.*, **23**, 5301–5307.

45. Stoler,D.L., Chen,N., Basik,M., Kahlenberg,M.S., Rodriguez-Bigas,M.A., Petrelli,N.J. and Anderson,G.R. (1999) The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proc. Natl Acad. Sci. USA*, **96**, 15121–15126.

46. Donahue,S.L., Lin,Q., Cao,S. and Ruley,H.E. (2006) Carcinogens induce genome-wide loss of heterozygosity in normal stem cells without persistent chromosomal instability. *Proc. Natl Acad. Sci. USA*, **103**, 11642–11646.

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

48. Hicks,G.G., Shi,E.G., Li,X.M., Li,C.H., Pawlak,M. and Ruley,H.E. (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nat. Genet.*, **16**, 338–344.

49. von Melchner,H., DeGregori,J.V., Rayburn,H., Reddy,S., Friedel,C. and Ruley,H.E. (1992) Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev.*, **6**, 919–927.

50. Wu,X., Li,Y., Crise,B. and Burgess,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.