Nucleic Acids Research, 2005, Vol. 33, No. 2 ***e14***
doi:10.1093/nar/gni016

Gene trap and gene inversion methods for conditional gene inactivation in the mouse

Hong-Bo Xin, Ke-Yu Deng, Bo Shui, Shimian Qu[2], Qi Sun[1], Jane Lee, Kai Su Greene, Jason Wilson, Ying Yu, Morris Feldman and Michael I. Kotlikoff*

Department of Biomedical Sciences, College of Veterinary Medicine, [1]Computational Biology Service Unit, Cornell Theory Center, Cornell University Ithaca, NY 14853, USA and [2]Department of Cancer Biology, Vanderbilt University School of Medicine Nashville, TN 37235, USA

Received November 18, 2004; Revised and Accepted December 23, 2004

ABSTRACT

Conditional inactivation of individual genes in mice using site-specific recombinases is an extremely powerful method for determining the complex roles of mammalian genes in developmental and tissue-specific contexts, a major goal of post-genomic research. However, the process of generating mice with recombinase recognition sequences placed at specific locations within a gene, while maintaining a functional allele, is time consuming, expensive and technically challenging. We describe a system that combines gene trap and site-specific DNA inversion to generate mouse embryonic stem (ES) cell clones for the rapid production of conditional knockout mice, and the use of this system in an initial gene trap screen. Gene trapping should allow the selection of thousands of ES cell clones with defined insertions that can be used to generate conditional knockout mice, thereby providing extensive parallelism that eliminates the time-consuming steps of targeting vector construction and homologous recombination for each gene.

INTRODUCTION

Recombinase-mediated, conditional gene inactivation (1–3) has emerged as a powerful genetic method for the determination of the complex function of genes in mammals *in vivo*. While permanent germline gene inactivation in embryonic stem (ES) cells using insertional mutagenesis techniques has been developed in commercial and academic settings to provide large-scale genome coverage (4–6), similar approaches for conditional mutagenesis are not currently available. We have developed a system for conditional gene inactivation in the mouse based on recombinase-mediated, unidirectional DNA inversion that can provide substantial coverage of the mouse genome. In this system, a gene trap vector is randomly integrated into mouse ES cells in a manner that allows identification of the integration site, reactivation of the allele by removal of the trap cassette and subsequent conditional inactivation of the trapped gene. The basis of the conditional gene inactivation is a genetic element that is transcriptionally silent when inserted within an intron; when inverted by site-specific recombination, the element disrupts endogenous mRNA production from the locus of integration. Here, we describe the design of the high throughput, conditional gene inactivation system, provide *in vitro* and *in vivo* evidence in support of inversion-based gene inactivation and report the results of an initial ES cell screen resulting in ES cell clones that are available to make mice suitable for conditional gene inactivation by crosses with specific Cre-expressing transgenic mice.

MATERIALS AND METHODS

Plasmid constructs

Oppositely oriented mutant loxP (mLoxP) (7) and equivalently oriented Flp recombinase recognition sites (FRT) (8) were synthesized and subcloned into pGEM-T (pGEM-loxP2 and pGEM-FRT2). The GT cassette (pGEM-GT) was assembled by subcloning PGK-TK, the PGK promoter upstream of the puromycin phosphotransferase gene, and conserved splice donor sequences into pGEM-FRT2 (EvoRV). The GI cassette (pGEM-GI) was made by inserting the human bcl-2 gene intron 2/exon 3 splice acceptor (9), the internal ribosome entry site of the encephalomyocarditis virus (10), an enhanced

*To whom correspondence should be addressed at Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, T4018 VRT, Box 11, Ithaca, NY 14853-6401, USA. Tel: +1 607 253 3336; Fax: +1 607 253 3317; Email: mik7@cornell.edu
Correspondence may also be addressed to Hong-Bo Xin. Tel: +1 607 253 4321; Fax: +1 607 253 3317; Email: hx25@cornell.edu

green fluorescent protein (eGFP) reporter gene and the bovine growth hormone polyadenylation signal sequence (11) (pA) into pGEM-mloxP2 (EcoRV). To construct the electroporation vector, the GT cassette flanked by FRT sites was released from pGEM-GT using NotI and XhoI and subcloned into the same sites in the pGEM-GI vector in the opposite orientation of the GT cassette (referred to pGEM-rGT-GI), then released by XhoI and SalI digestion and circularly ligated with a human immunoglobulin G coding region used to protect critical sequences from nuclease digestion (12).

The pCAGGS-GI/rGI-nlsDsRED constructs for transfecting mammalian cells were generated by replacing the Cre/polyA coding sequence in the pCAGGS-nlsCre vector (13) with a DsRED (Clontech)/pA fragment and inserting the GI cassette into the intron between the promoter and nlsDsRED in both orientations.

Genome walking and bioinformatics

Selected ES cell clones were screened by Southern blotting for single or multiple insertions by probing EcoRI digested DNA with an eGFP fragment. The genome insertion site of clones with a single insertion was then determined by genome walking to obtain sequence on either side of the inserted cassette, using a modification of a method to amplify sequence between a single gene-specific primer and an unphosphorylated linker (14). Briefly, a double-stranded linker with an Tsp509I overhang was annealed from complementary single-strand oligonucleotides as shown below.

ES cell DNA was digested by Tsp509I, ligated to the unphosphorylated linker with T4 DNA ligase and column purified to remove excess linker. Oligonucleotide primers complimentary to the linker cassette and to either the 5′ or 3′ end of the gene trap (GT) cassette were used to amplify DNA between the cassette and the linker. The PCR products were diluted and amplified for the second time using a pair of nested primers complimentary to the trap cassette and linker (LCS1 and LCS2 above indicate the first and second round linker primers). The PCR products from the second round were sequenced directly after column purification. The tagged sequences were mapped to the mouse genome (NCBI Mouse build 32) by NCBI BLAST, using the Ensembl mouse 20.32b for gene annotation and a cutoff *E*-value for the BLAST analysis as 1×10^{-10} (15,16). The top 10 blast targets and the DNA transcripts in the targeted region were manually evaluated using a visualization tool that we developed. Only insertions of correct orientation within an annotated gene or *ab initio* gene prediction are reported (17).

Cell transfection

HEK293 cells were transfected with 0.8 μg of pCAGGS-revIC-nlsDsRED or pCAGGS-IC-nlsDsRED plasmid with and without 0.8 μg pCAGGSnls-cre plasmid DNA using 2 μl Lipofectamine 2000 (Invitrogen) for each well. Stable integrations were obtained by antibiotic selection and individual clones that had flipped to green were manually picked twice into 96 well plates for further expansion. Two of the eight colonies selected appeared 100% positive for eGFP and were transfected with Cre recombinase as described above and fluorescence recorded at 24 and 48 h after transfection. As a positive control, a stable HEK293 cell line carrying pCAGGS-IC-nlsDsRED was also transfected to insure Cre recombinase activity.

In vitro recombination assay

GI cassette plasmid DNA (2 μg, pGEM-GI) was incubated with Cre recombinase protein (80 U) in recombination buffer at 37°C. Aliquots were removed at time points indicated and placed in 5 μl of 5× cold stop buffer (0.2% SDS, 40 mM EDTA and 2 mg/ml proteinase K). DNA was extracted and digested by BamH1 and Xho1 at 37°C for 2 h. The DNA digestion pattern was examined by electrophoresis in a 0.8% agarose gel.

Transgenic mice

The pCAGGS-rGI transgene was prepared by replacing nlsCre in pCAGGS-nlsCre with the opposite orientation of the GI cassette. DNA was purified using a QIAEX II kit and injected into C57BL/6J × DBA/2J F_2 zygotes by standard pronuclear microinjection techniques. Three founder lines were established on C57BL/6 backgrounds and crossed with Cre mice driven by the smMHC promoter (18) or an unpublished line developed by us that produces ubiquitous early embryonic Cre expression.

RESULTS

Description of the conditional gene inactivation system: the double switch system

The design of the gene trap/conditional gene inactivation vector is shown schematically in Figure 1A. The vector relies on two separate cassettes flanked by different site-specific DNA recombination recognition sequences, allowing for discrete control of each cassette. The GT cassette, consisting of a ubiquitous promoter (PGK), positive selection marker and splice donor sequence, is used to select for random integrations in mouse ES cells that have captured an endogenous gene polyadenylation signal (4,19–22). The GT cassette, flanked by FRT, is removed from ES cell clones by transfection with Flp recombinase after confirmation of the integration site, restoring the allele to functionality (Figure 1B, first switch). To facilitate selection for ES cell clones in which the GT cassette has been removed, the cassette includes a negative selection element.

The gene inactivation (GI) cassette provides for conditional inactivation of genes in which intronic integration has occurred (Figure 1B). The basis of gene inactivation is a recombinase-mediated, unidirectional DNA inversion from transcriptionally silent to transcriptionally active mode. The GI cassette is in the opposite orientation of the GT cassette to prevent recognition of the strong splice acceptor (SA) [human bcl-2 gene intron 2/exon 3 splice acceptor (9)] and polyadenylation (pA) signal sequences within the trapped locus; an IRES and eGFP reporter sequence serve to report the orientation of the GI cassette. These elements are flanked by

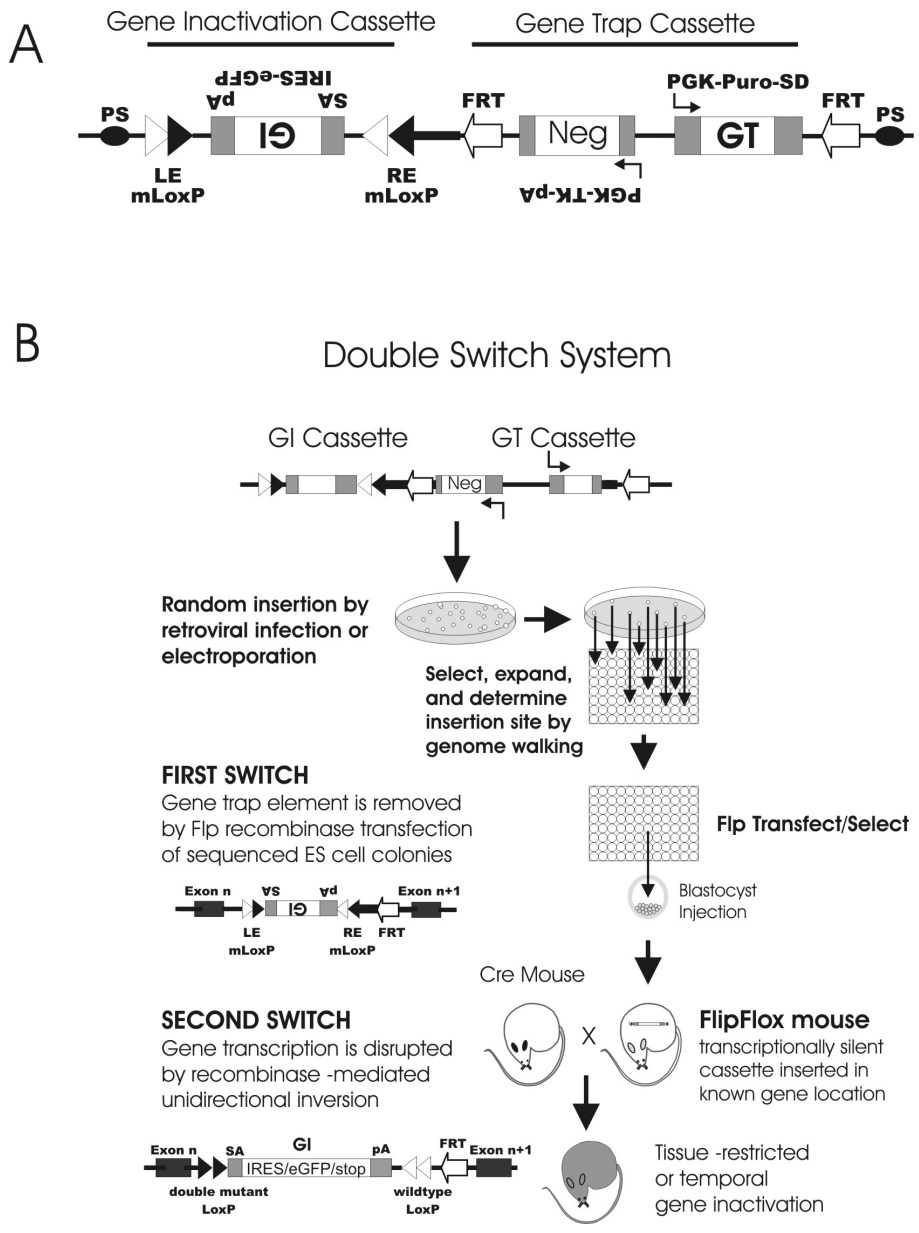

Figure 1. Vectors and conditional gene inactivation system. (**A**) Schematic representation of the gene trap vector. GI, gene inactivation; GT, gene trap cassettes; SA, splice acceptor; pA, polyadenylation signal; Neg, negative selection cassette; mLoxP, mutant LoxP sequence; LE, RE, left and right LoxP elements; PS, protection sequence to protect vector from exonuclease digestion; FRT, Flp recombinase recognition site; PGK-Puro-SD, phosphoglycerate kinase promoter directing expression of the puromycin phosphotransferase gene followed by a splice donor sequence; and PGK-TK-pA, the same promoter driving the thymidine kinase coding sequence for negative selection. (**B**) Schematic representation of double switch system. The GI cassette is transcriptionally inactive until inverted *in vivo*.

mLoxP sites, which have previously been shown to promote irreversible deletion of a flanked sequence (23,24). The mLoxP sequences are oriented in opposite directions to promote unidirectional inversion of the intervening sequence upon exposure to Cre recombinase (25). Thus, mice produced from ES cells containing the GI cassette in its anti-promoter orientation should have a functional allele, but when crossed with transgenic mice expressing Cre recombinase in a tissue-specific or temporally controlled manner, the transcriptionally silent GI cassette is irreversibly inverted and a fusion transcript of the truncated endogenous gene and the GI cassette formed by splicing, resulting in the expression of the reporter gene and premature termination of the endogenous gene mRNA (Figure 1B, second switch).

Transcriptional control by the recombinase-mediated unidirectional DNA inversion

The effectiveness of a DNA inversion strategy to inactivate mouse genes in a conditional manner relies on several factors. The inserted GI cassette must not alter the expression of the trapped or targeted gene in its initial (inverted with respect to the endogenous promoter or anti-promoter) orientation, thereby insuring that the allele is functional in the absence of recombinase expression; once inverted by the recombinase

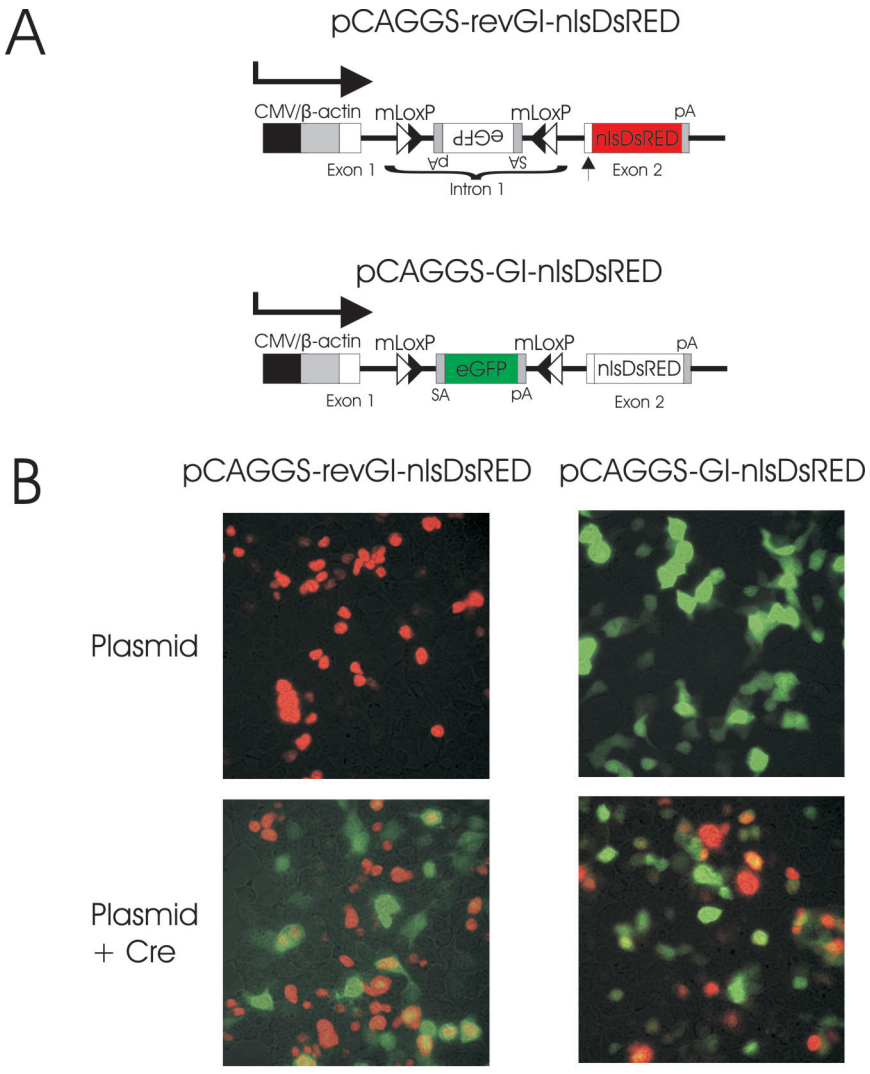

Figure 2. Orientation-dependent transcriptional activity of the gene inactivation cassette. (**A**) Schematic representation of reporter vectors. The nlsDsRED sequence was inserted into the second exon of the β-actin gene (arrow denotes remaining partial second exon) and the GI cassette (containing IRES-eGFP) inserted in anti-promoter (rev) and promoter orientations. (**B**) Vectors were transfected into HEK293 cells. In the anti-promoter orientation, no eGFP activity was observed and the construct had no effect on dsRED expression (top left), whereas in the promoter orientation, the GI cassette ablated expression DsRED (top right) and the expression of eGFP was equivalent to control plasmid transfections. Subsequent transfection of these cultures with Cre recombinase resulted in conversion to the opposite transcriptional state at 24 h. Note transient transfection and protein lifetime prevents evaluation of extent of conversion.

(promoter orientation), the cassette must completely disrupt the expression of the endogenous gene so that full gene inactivation is achieved; and inversion of the DNA element *in vivo* must be efficient and unidirectional for the switch to be effective.

To determine the extent to which the GI cassette satisfies these conditions, we constructed reporter vectors incorporating the cassette in a manner that allowed determination of its transcriptional activity. The GI cassette, containing the flanking mLoxP sequences and the eGFP reporter, was placed in the first intron of the β-actin gene, downstream of the pCAGGS promoter (13), in either the anti-promoter or promoter orientation. The DsRED coding sequence with nuclear localization signal (nlsDsRED) was inserted downstream of the GI cassette in the second exon of the β-actin construct (Figure 2A). Optical separation of the fluorescence signals, as well as the nuclear localization of DsRED compared with the cytosolic expression of eGFP, allowed us to independently assess the orientation and transcriptional activity (silent or truncating) of the GI cassette, as well as the efficiency of recombinase-mediated inversion following transfection with Cre recombinase. HEK293 cells were transiently transfected with these constructs and examined at 24 and 72 h following transfection. As shown in Figure 2B (left panel), the expression of the GI cassette in its anti-promoter orientation resulted in the robust expression of the downstream reporter gene, DsRED. In two such experiments, the percentage of cells expressing DsRED was 24.5 and 30.8% 24 h after transfection, values roughly equivalent to the transfection efficiency of the experiment, indicating no detectable disruption of the downstream allele. However, when the GI cassette was inserted in the promoter orientation (Figure 2B, right panel), virtually all

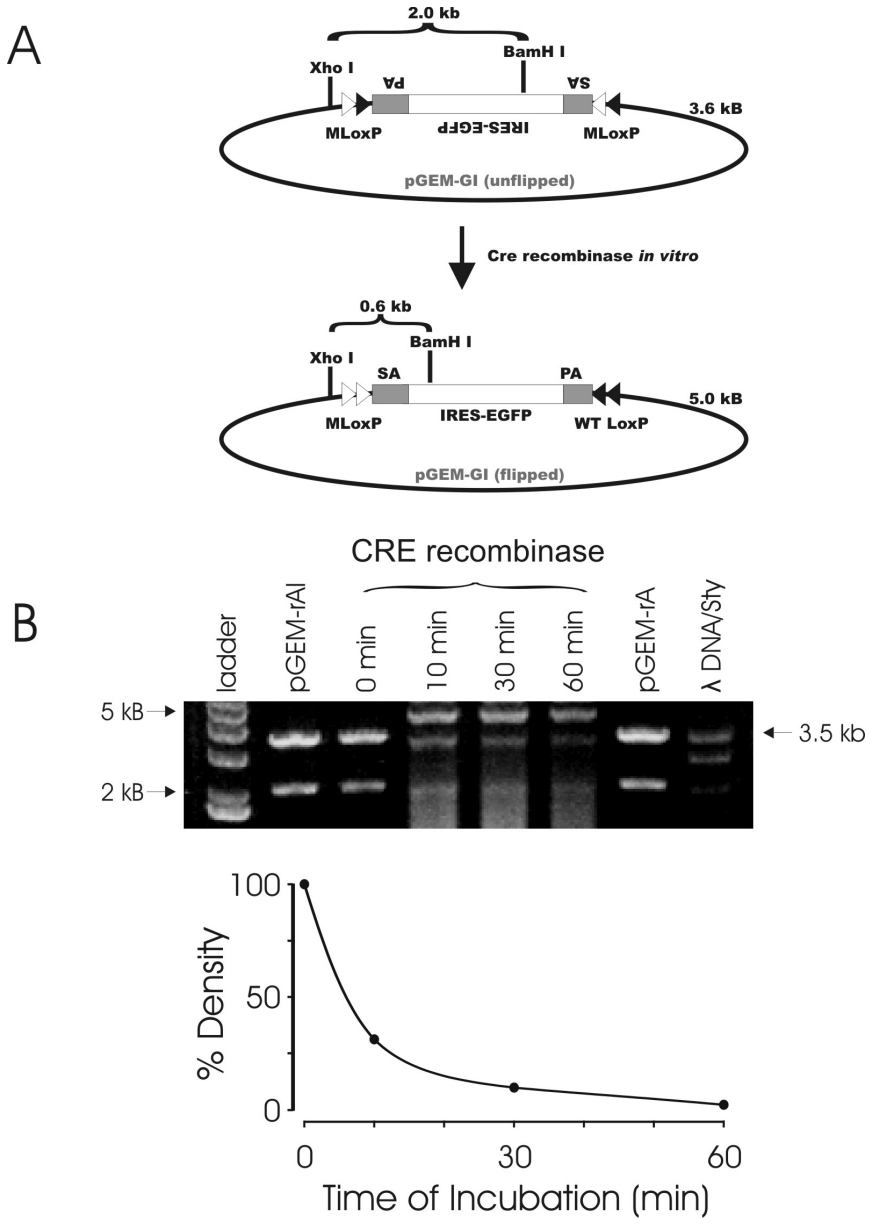

Figure 3. Unidirectional DNA inversion by mLoxP sequences *in vitro*. (**A**) Plasmid components with Xho1 and BamH1 sites. The insert and plasmid backbone are 2.6 and 3.0 kb, respectively. Digestion of the vector in the constructed orientation results in 2.0 and 3.6 kb (3.0 plasmid + 0.6 insert) products. Inversion of the insert reduces the size of the insert fragment to 0.6 kb, resulting in a 5 kb product. (**B**) Plasmid at top was exposed to Cre recombinase protein for the times shown, digested with BamH1 and Xho1 and separated on agarose gel (above). The 2.0 and 3.6 kb bands are eliminated with increasing time of exposure to the enzyme and the predicted 5 kb band is seen (0.6 kb band not shown). The graph (below) shows densitometric quantification of the original 3.6 kb band.

transcription of the downstream DsRED reporter gene was eliminated, whereas 32 and 48% were eGFP positive in two separate experiments. In these experiments, the percentage of eGFP positive cells was within the range of transfection efficiencies observed in our experiments (varied between 10 and 73% in all experiments), and the brightness of individual cells was equivalent to that observed with transient transfection of eGFP expression plasmids, indicating that the GI cassette pseudo exon was incorporated into RNA by the transcriptional machinery in a highly efficient manner and produced a robust transcriptional block. Experiments in which a Cre recombinase expression plasmid was co-transfected with the inverted GI constructs resulted in the expected manipulation of the GI cassette; 90% of transfected cells (expressing either fluorophore) were eGFP positive. Similarly, 79% of cells co-transfected with Cre and the GI cassette in the promoter orientation (green) flipped to red (Figure 2). Because the transient transfection of the constructs and the stability of the translated proteins precluded a clear assessment of the efficiency of DNA inversion, we constructed stable cell lines with the same constructs. In four such experiments, 27.6% of cells were flipped from one orientation to the other (range of 15–44%), a figure close to the transfection efficiency for these experiments.

We next investigated the efficiency of mLoxP sequences (23), which have previously been shown to promote irreversible gene deletion (24), in catalyzing the required unidirectional DNA inversion. LoxP recognition sequences placed in opposite orientations lead to Cre recombinase catalyzed inversion of the intervening sequence, such that at equilibrium 50% of the intervening sequence is in each orientation (23,26). As shown in Figure 3, restriction digest analysis of plasmid DNA containing the GI cassette flanked by oppositely oriented mLoxP sites indicated that Cre recombinase protein resulted in a rapid and unidirectional inversion of the cassette *in vitro*, resulting in a predictable alteration in DNA fragment sizes. Greater than 90% of the DNA was inverted within 1 h of exposure to 80 U Cre recombinase at 37°C, indicating efficient recognition of the mutant sequence by Cre recombinase, and a markedly reduced binding affinity for the recombined sequence.

The unidirectional nature of this inversion was further confirmed in cells containing stable insertions of the GI cassette. In these experiments, the reporter construct described above, containing an antibiotic resistance cassette, was transfected in its anti-promoter (red) orientation and stable integrants were selected with G418. Surviving cells displayed no eGFP fluorescence and were virtually all positive for DsRED. These cells were transfected with Cre recombinase and eGFP positive clones manually picked twice to obtain homogenous eGFP-expressing cultures, in which the GI cassette had been flipped into the same orientation as the upstream promoter. As shown in Figure 4, subsequent transfections of these cells with Cre recombinase failed to produce any DsRED-positive cells, indicating that once inverted, the GI cassette could no longer flip to the anti-promoter orientation, whereas transfection of control cells stably expressing the cassette in the promoter (green) orientation with Cre recombinase resulted in the marked expression of DsRED. These results indicate that once inverted, the GI cassette is insensitive to the recombinase.

Transfection of cell lines may result in a far greater expression of Cre recombinase than can be achieved *in vivo* in the mouse. To confirm the effectiveness of DNA inversion in a more practical context, we created lines of transgenic mice, in which the transgene consisted of the GI cassette, flanked by mLoxP sites (Figure 5A), was inserted in the pCAGGS plasmid (13) in the opposite orientation of the CMV promoter/β-actin enhancer. Hemizygous GI^+ mice were crossed with mice expressing Cre recombinase under control of the smooth muscle-specific myosin heavy chain promoter (18), or with mice that direct Cre expression to all tissues of the early embryo (K.-Y. Deng, H.-B. Xin and M.I. Kotlikoff, manuscript in preparation), and tissues analyzed for fluorescence and construct orientation to determine the extent to which the cassette is inverted *in vivo*. As shown in Figure 5B, PCR analysis of the tissues of offspring from these crosses using primers specific for the promoter orientation of the GI cassette indicated that only double transgenic (Cre^+/GI^+) mice contained the flipped cassette, and that this orientation of the cassette was specific to tissues in which Cre recombinase expression was directed. In litters from two separate founders, three-week-old double transgenic ($GI^+/smCre^+$) mice displayed robust eGFP fluorescence confined to vascular and nonvascular smooth muscle tissues (Figure 5C), consistent with the cellular expression of

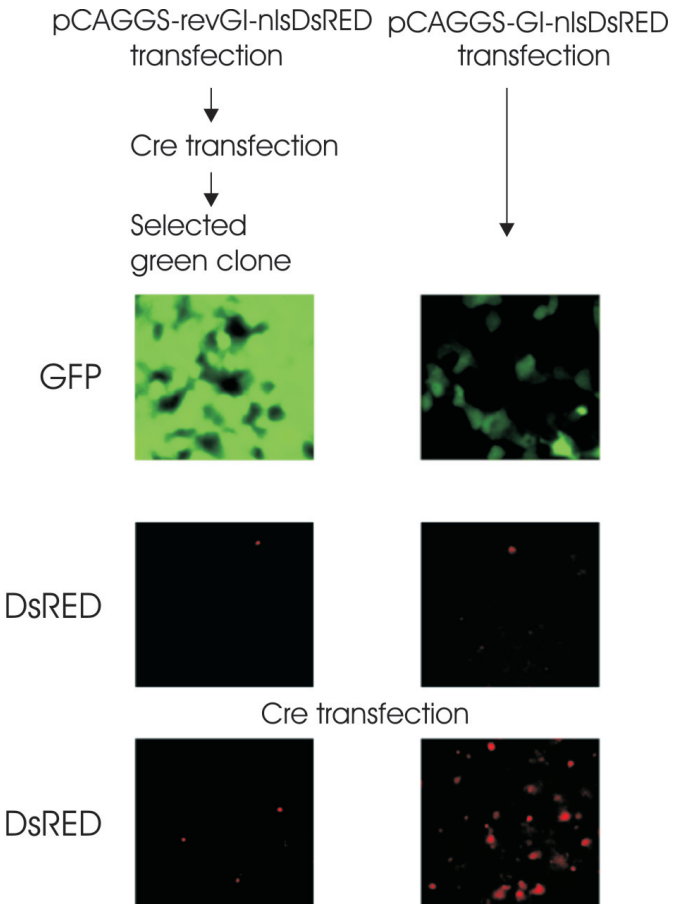

Figure 4. Unidirectional DNA inversion *in vivo*. Left, cells were selected for stable expression of the GI cassette in its transcriptionally silent (anti-promoter, red) orientation, transfected with Cre recombinase expression plasmid to invert the cassette, and green clones selected. Top and middle are eGFP and DsRED images of same fields. Following an additional transfection with Cre recombinase, no inversion of the previously flipped cassette occurred (bottom, DsRED image). Right, control cells stably transfected with the GI cassette in its transcriptionally active (promoter, green) orientation demonstrated robust gene inversion when transiently transfected with Cre recombinase. Bottom images taken 48 h post-Cre transfection.

Cre recombinase. Mice carrying the GI cassette but not the Cre transgene never displayed eGFP fluorescence. Moreover, analysis of offspring from crosses producing early embryonic expression of Cre recombinase resulted in a generalized inversion of the GI cassette, as indicated by robust fluorescence of all cells in 8 and 18.5 day mice (Figure 5D).

Gene trapping and determination of insertion site

Unlike insertional mutagenesis in which 3′ RACE can be used to identify a disrupted exon sequence, conditional gene inactivation by DNA inversion and transcriptional block requires intronic insertion, and unambiguous determination of the site of vector integration is therefore necessitated. To test the described vectors and establish the feasibility of identification strategies, a gene trap screen was conducted in R1 ES cells by electroporating cells in the presence of gene trap vector DNA (Figure 1A). Cells were selected with puromycin for productive integrations and resistant clones transferred to triplicate 96 well plates, one of which was used to make DNA for

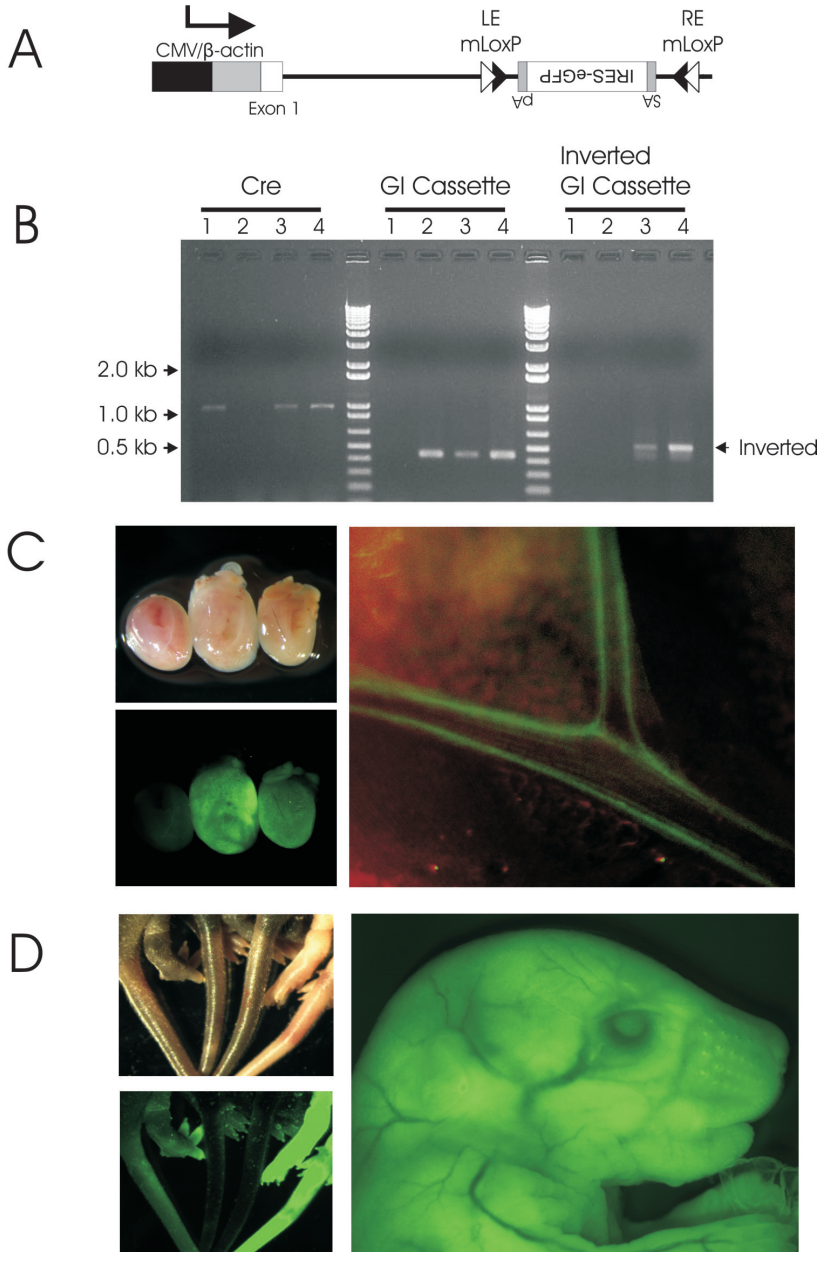

Figure 5. Recombinase-mediated DNA inversion in transgenic mice. (**A**) Transgene containing the GI cassette (with eGFP reporter) driven by the pCAGGS. (**B**) Genotyping of F1 offspring of GI^+ and Cre recombinase transgenic mice expressed in smooth muscle (lanes 1–3) or whole embryo (lane 4). Only mice in lanes 3 and 4 expressed both the GI and Cre recombinase transgenes. Using primers specific for the inverted form of the cassette, a PCR product was detected in DNA prepared from smooth muscle tissue (lane 3) or whole embryo (lane 4) only in mice expressing both transgenes. (**C**) Tissue from double transgenic mice with smooth muscle restricted Cre expression. Left, light (above) and fluorescent (below) images of urinary bladders from Cre^-,GI^+ (left) and Cre^+,GI^+ mice (middle and right); 0.8×. Right, superimposed light and fluorescent image of artery from Cre^-,GI^+ mouse shows fluorescence restricted to smooth muscle; 10×. (**D**) Left, light (above) and fluorescent (below) images of tails from 18.5 day mice. Tails at left and right: Cre^+,GI^+; tails in middle: Cre^-,GI^+. Fluorescence is muted by pigment in tail at left. Right, fluorescent image of 8 day Cre^+,GI^+ mouse upper body tissues.

sequencing. Resistant colonies were first screened by Southern blotting with an eGFP fragment after digestion with EcoR1. Of 239 resistant clones, a single band was identified in 136 clones, with multiple bands in 64 clones and no evidence of insertion in the remaining clones.

Several sequencing strategies were pursued to obtain genomic sequence flanking the integrated vector. The most successful method was a modification of a genome walking procedure (14), which allowed unambiguous determination of the exact site of vector integration in a high percentage of cases. Amplification of flanking DNA by genome walking resulted in useable sequence from one end or both ends of the cassette in 81 (60%) of the single integrants. As shown in Table 1, our bioinformatics strategy resulted in the identification of the locus of insertion in 39 clones, or 48% of the clones in which flanking sequence was obtained. This is a conservative estimate, as only known or *ab initio* predicted genes were considered in the analysis, whereas in many cases,

Table 1. List of trapped genes and insertion sites

Gene trap clones Clone	Location	Intron/exon	Gene	Annotation
AB4	11:46999933	Intron 2 (4)	GENSCAN00000030349	Unknown
AB6	3:20959528	5′ UTR (1.2 kb)	Genscan prediction	Unknown
AB7	19:4600746	5′ UTR (147 bp)	ENSMUSG00000006456	Rbm14 Unknown function, RNA-binding region RNP-1 (RNA recognition motif)
AB10	19:9205621	5′ UTR (73)	ENSMUSG00000044504	Zbtb3, zinc finger and BTB domain containing 3
AB23	Y:3064117	Intron 3 (9)	ENSMUSG00000053211	Zfy1, zinc finger protein 1, Y linked
AB24	1:24787681	Intron 1 (16)	ENSMUSG00000026139	LMBR membrane protein (lipocalin like motif)
AB34	5:129049765	3′ UTR (1.2 kb)	ENSMUSESTG00000024736	Unknown
AB35	8:97091210	5′ UTR (48 bp)	ENSMUSESTG00000002583	Unknown
AB41	11:14526302	Intron 2 (3)	GENSCAN00000063363	Nuclear envelope pore membrane protein POM 121
AB45	3:33612037	5′ UTR (5)	ENSMUSG00000027671	53 kDa BRG1-associated factor A (actin-related protein BAF53A)
AB52	15:83418824	5′ UTR (108 bp)	ENSMUSG00000022478	Phf5a (PHD finger protein 5a)
AB57	13:97799446	Exon 1(1)	GENSCAN00000141083	Unknown
AB60	7:37794879	5′ UTR (119 bp)	ENSMUSG00000025389	Lens fiber major intrinsic protein (mip26)
AB61	9:46147000	5′ UTR (307)	ENSMUSG00000032115	Hyou1 Cab140) hypoxia up-regulated 1
AB72	18:84776833	Intron 1(5)	GENSCAN00000002917	Unknown
AB82	15:77123042	Intron 4(7)	ENSMUSG00000053168	Ly6c lymphocytes antigen LY-6C precursor
AB112	15:91187795	5′ UTR (72 bp)	ENSMUSG00000022618	Chk1, choline/ethanolamine kinase
AB122	14:110179384	Intron 1(2)	ENSMUSESTG00000021020	Unknown
AB162	9:63143799	Intron 3(5)	GENSCAN00000046398	Unknown
AB166	1:108891482	Exon 2(2)	ENSMUSG00000048549	Bcl2. apoptosis regulator
AB167	6:104002195	Intron 2(21)	ENSMUSG00000030075	Cntn3, contactin 3; plasmacytoma-associated neuronal glycoprotein
AB172	7:NT_081205	Exon 1(9)	GENSCAN00000100015	Unknown
AB179	15:77109835	Intron 3(5)	ENSMUSG00000053168	Ly6c, similar to lymphocyte antigen 6 complex, locus I
AB191	X:81394369	Intron 2(3)	GENSCAN00000021466	Unknown
AB203	8:95320457	5′ UTR (191 bp)	ENSMUSG00000033009	Prolyl 4-hydroxylase, alpha subunit
AB208	3:30902840	Intron 1(15)	ENSMUSG00000027684	Evi1, ecotropic viral integration site 1 (embryonic lethal)
AA3	7:66130092	5′ UTR (74)	ENSMUSESTT00000000727	Unknown
AA4	9:125575183	Intron 1(5)	ENSMUSG00000035498	Unknown
B6	15:9404595	Intron 4(6)	ENSMUSG00000055463	Unknown
B2	19:35685964	Intron 1(3)	GENSCAN00000029475	Unknown
AC2	X:125314423	Intron 1(3)	ENSMUSESTT00000005143	Morf4l2, transcription factor-like protein MORF-related gene X protein. Splice variant
AE1	4:117345976	Intron 11(21)	GENSCAN00000056387	Cdc20, cell division cycle protein 20 homolog. Splice variant
AD2	14:16382207	Intron 2(3)	GENSCAN00000011182	Unknown
AD8	1:134403527	Intron 1(2)	GENSCAN00000079343	Unknown
AE2	11:118712586	Intron 3(12)	ENSMUSG00000039976	Unknown, RabGAP/TBC domain
AF2	11:70076743	Exon 1 (5)	ENSMUSG00000018919	Tm4sf5, transmembrane 4 superfamily, member 5, unknown function
AG2	1:63421977	Intron 1(3)	ENSMUSESTG00000010455	Unknown
AG9	2:32724693	Exon 1(2)	ENSMUSG00000039164	Unknown
AH2	17:8732662	Intron 1(2)	GENSCAN00000120409	Unknown

expressed sequence tag (EST) data or separate analysis of the local sequence by Genscan suggested the presence of a potential gene locus, within which the vector was appropriately oriented to capture the polyadenylation sequence. In addition, although flanking sequences allowed determination of the site of integration in over 90% of the clones examined, in many cases the trapped polyadenylation sequence could not be unambiguously determined and was therefore not listed. These failures could be attributed to incomplete annotation or inadequate sequence (a substantial number of insertions in which the gene locus could not be identified occurred in areas of the genome that contain substantial sequence gaps), but could also result from the trapping of pseudo polyadenylation sequences that are not associated with an endogenous gene. Of the unambiguous integrants shown in Table 1, 56% were intronic, whereas 28% occurred in the 5′-untranslated region (5′-UTR), an average of 114 bp from the start of exon 1. Thus, in this screen ~10% of all resistant clones could be identified as intronic insertions in known genes.

DISCUSSION

Our experiments demonstrate the practical basis of the unidirectional DNA inversion for the production of conditional knockout mice. We have shown that (i) the inserted DNA vector is transcriptionally silent before activation; (ii) following unidirectional inversion the vector powerfully disrupts transcription; (iii) unidirectional inversion of the cassette can be efficiently achieved in cells and in transgenic mice; (iv) gene trapping results in numerous single intronic insertions within endogenous genes; (v) the GT cassette can be readily removed leaving the transcriptionally silent inverted element; and (vi) the site of insertion is easily determined. Taken together, these experiments indicate that unidirectional

DNA inversion is an effective strategy for control of gene expression *in vivo*. The combination of unidirectional gene inversion with gene trap methods provides a high-throughput method for the generation of knockout mice.

We used a gene trap vector designed for electroporation to conduct a small-scale screen for ES cells with single, intronic insertions. The overall throughput for unambiguous insertions of this nature was 10%, which when scaled should yield many hundreds of useful clones. Throughput would likely be markedly enhanced by incorporation of the cassettes within a retroviral vector, providing for more efficient transfer of the vector and increasing the percentage of single insertions (27). Additionally, replacement of the GT cassette (Figure 1A) with an FRT-flanked selection cassette oriented inverse to the inactivation cassette, and flanking both cassettes with homologous sequences, enables the system's use as a targeted replacement vector. In this application, the vector would be targeted to an intron, which simplifies vector construction as minor PCR errors and frame shifts are not a concern; alternatively, the vector could be targeted to an intronic locus in a mouse bacterial artificial chromosome to produce highly efficient targeting vectors (28–31). These application would extend the utility of inversion-based gene inactivation, as gene trap approaches are limited by several factors including poor accessibility of some genes, a relatively high percentage of trapped sequences that are not expressed as mRNA and recombinations at the site of insertion (32,33).

A key requirement in the implementation of a gene trapping/conditional knockout strategy is the determination of the location of insertion of the vector, insuring that insertion is intronic and allowing a determination of the likely functional result of transcriptional termination at the insertion site. In cases of gene overlap, knowledge of the insertion site is particularly important. The genome walking method described here allows rapid and efficient determination of flanking sequence, and the attendant employment of bioinformatics strategies to rapidly determine the location of the insertion and the identity of the gene locus from sequence databases. Determination of the insertion site was achieved in 60% of the single insertion clones, and repeated analysis of the other clones would probably improve this percentage. A significant limitation in the use of the sequence information was ambiguity as to the trapped gene locus; as mammalian gene annotation continues, we anticipate attendant improvements in the determination of functionally relevant insertions.

One limitation of the described transcriptional arrest strategy should be noted: the assumption of conservation of function of a trapped allele following removal of the GT cassette assumes minimal transcriptional effects associated with the intronic insertion of the GI cassette. As *cis* transcriptional effects associated with intronic insertions are well documented and can result in inefficient or aberrant RNA splicing (34,35), a careful evaluation of the transcriptional function of the trapped allele seems prudent. We have not evaluated this effect as its importance will depend on the nature of the insertion (which intron and where within the intron), as well as the extent of alternative splicing of the specific gene. We also note that the described ES cell collection can also be used to create mice with a constitutively inactivated allele by inverting the GI cassette *in vitro* and retaining the GT cassette.

While still emerging, conditional gene expression strategies in the mouse have already had major impacts on cancer biology and immunobiology. It is likely that temporal and tissue-specific control of the expression of virtually all clinically relevant genes will be achieved in the mouse over the next two decades. The above described system should provide important advantages in the achievement of this goal, and should also be useful for the functional annotation of genes of other eukaryotic species.

ACKNOWLEDGEMENTS

We thank Drs J. Miyasaki for the pCAGGs vector, Dr S. Dymecki for the pFRT2 plasmid, Dr P. Leder for the bcl-2 splice acceptor, and Drs P. Soloway and J. Schimenti for their critical reading of a draft of the manuscript. Supported by NIH grants DK065992 and HL45239, and the NYS CAT fund. Funding to pay the Open Access publication charges for this article was provided by DK065992.

REFERENCES

1. Sauer,B. and Henderson,N. (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl Acad. Sci. USA*, **85**, 5166–5170.
2. Gu,H., Marth,J.D., Orban,P.C., Mossmann,H. and Rajewsky,K. (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science*, **265**, 103–106.
3. Sauer,B. (1998) Inducible gene targeting in mice using the Cre/lox system. *Methods*, **14**, 381–392.
4. Skarnes,W.C., Auerbach,B.A. and Joyner,A.L. (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.
5. Zambrowicz,B.P., 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.
6. Wiles,M.V., Vauti,F., Otte,J., Fuchtbauer,E.M., Ruiz,P., Fuchtbauer,A., Arnold,H.H., Lehrach,H., Metz,T., von Melchner,H. *et al.* (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nature Genet.*, **24**, 13–14.
7. Albert,H., Dale,E.C., Lee,E. and Ow,D.W. (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.*, **7**, 649–659.
8. McLeod,M., Craft,S. and Broach,J.R. (1986) Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Mol. Cell. Biol.*, **6**, 3357–3367.
9. Tsujimoto,Y. and Croce,C.M. (1986) Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl Acad. Sci. USA*, **83**, 5214–5218.
10. Duke,G.M., Hoffman,M.A. and Palmenberg,A.C. (1992) Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J. Virol.*, **66**, 1602–1609.
11. Soriano,P., Montgomery,C., Geske,R. and Bradley,A. (1991) Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell*, **64**, 693–702.
12. Neilan,E.G. and Barsh,G.S. (1999) Gene trap insertional mutagenesis in mice: new vectors and germ line mutations in two novel genes. *Transgenic Res.*, **8**, 451–458.
13. Niwa,H., Yamamura,K. and Miyazaki,J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, **108**, 193–199.
14. Kilstrup,M. and Kristiansen,K.N. (2000) Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res.*, **28**, E55.
15. Birney,E., Andrews,T.D., Bevan,P., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cuff,J., Curwen,V., Cutts,T. *et al.* (2004) An overview of Ensembl. *Genome Res.*, **14**, 925–928.

16. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.
17. Burge,C. and Karlin,S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol*., **268**, 78–94.
18. Regan,C.P., Manabe,I. and Owens,G.K. (2000) Development of a smooth muscle-targeted cre recombinase mouse reveals novel insights regarding smooth muscle myosin heavy chain promoter regulation. *Circ. Res*., **87**, 363–369.
19. Gossler,A., Joyner,A.L., Rossant,J. and Skarnes,W.C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science*, **244**, 463–465.
20. Friedrich,G. and Soriano,P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev*., **5**, 1513–1523.
21. Niwa,H., Araki,K., Kimura,S., Taniguchi,S., Wakasugi,S. and Yamamura,K. (1993) An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J. Biochem*., **113**, 343–349.
22. 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.
23. Hoess,R.H., Wierzbicki,A. and Abremski,K. (1986) The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res*., **14**, 2287–2300.
24. Araki,K., Araki,M. and Yamamura,K. (1997) Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Res*., **25**, 868–872.
25. Schnutgen,F., Doerflinger,N., Calleja,C., Wendling,O., Chambon,P. and Ghyselinck,N.B. (2003) A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat. Biotechnol*., **21**, 562–565.
26. Aranda,M., Kanellopoulou,C., Christ,N., Peitz,M., Rajewsky,K. and Droge,P. (2001) Altered directionality in the Cre-LoxP site-specific recombination pathway. *J. Mol. Biol*., **311**, 453–459.
27. Chang,W., Hubbard,S.C., Friedel,C. and Ruley,H.E. (1993) Enrichment of insertional mutants following retrovirus gene trap selection. *Virology*, **193**, 737–747.
28. Yang,X.D.W., Model,P. and Heintz,N. (1997) Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol*., **15**, 859–865.
29. Zhang,Y.M., Buchholz,F., Muyrers,J.P.P. and Stewart,A.F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genet*., **20**, 123–128.
30. Muyrers,J.P., Zhang,Y., Testa,G. and Stewart,A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res*., **27**, 1555–1557.
31. Lee,E.C., Yu,D.G., de Velasco,J.M., Tessarollo,L., Swing,D.A., Court,D.L., Jenkins,N.A. and Copeland,N.G. (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*, **73**, 56–65.
32. Voss,A.K., Thomas,T. and Gruss,P. (1998) Efficiency assessment of the gene trap approach. *Dev. Dyn*., **212**, 171–180.
33. Stanford,W.L., Cohn,J.B. and Cordes,S.P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nature Rev. Genet*., **2**, 756–768.
34. Klinz,F.J. and Gallwitz,D. (1985) Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **13**, 3791–3804.
35. Rijli,F.M., Dolle,P., Fraulob,V., LeMeur,M. and Chambon,P. (1994) Insertion of a targeting construct in a Hoxd-10 allele can influence the control of Hoxd-9 expression. *Dev. Dyn*., **201**, 366–377.