Genomewide production of multipurpose alleles for the functional analysis of the mouse genome

Frank Schnütgen*†, Silke De-Zolt*†, Petra Van Sloun*†, Melanie Hollatz†‡, Thomas Floss†‡, Jens Hansen†‡, Joachim Altschmied*, Claudia Seisenberger†‡, Norbert B. Ghyselinck§, Patricia Ruiz†¶, Pierre Chambon§, Wolfgang Wurst†‡, and Harald von Melchner*†‖

*Department of Molecular Hematology, University of Frankfurt Medical School, 60590 Frankfurt am Main, Germany; ‡Institute of Developmental Genetics, GSF-National Research Center for Environment and Health, 85764 Neuherberg, Germany; §Institut Clinique de la Souris and Institut de Génétique et de Biologie Moléculaire et Cellulaire, CU de Strasbourg, 67404 Illkirch Cedex, France; and ¶Center for Cardiovascular Research, Charité Universitätsmedizin, and Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

Contributed by Pierre Chambon, March 22, 2005

A type of retroviral gene trap vectors has been developed that can induce conditional mutations in most genes expressed in mouse embryonic stem (ES) cells. The vectors rely on directional site-specific recombination systems that can repair and reinduce gene trap mutations when activated in succession. After the gene traps are inserted into the mouse genome, genetic mutations can be produced at a particular time and place in somatic cells. In addition to their conditional features, the vectors create multipurpose alleles amenable to a wide range of postinsertional modifications. Here we have used these directional recombination vectors to assemble the largest library of ES cell lines with conditional mutations in single genes yet assembled, presently totaling 1,000 unique genes. The trapped ES cell lines, which can be ordered from the German Gene Trap Consortium, are freely available to the scientific community.

conditional mutagenesis | ES cells | gene trapping | site-specific recombination | insertional mutagenesis

With the complete sequencing of the human and mouse genomes, attention has shifted toward comprehensive functional annotation of mammalian genes (1, 2). Among the various approaches for addressing gene function, mutagenesis in the mouse is the most relevant for extrapolation to human genetic disease. Although several model organisms have been used in a variety of mutagenesis approaches, the mouse offers particular advantages because its genome structure and organization are closely related to those of the human genome. More importantly, mouse embryonic stem (ES) cells, which grow indefinitely in tissue culture, allow the generation of mice with defined mutations in single genes for functional analysis and studies of human disease.

Several mutagenesis strategies have been deployed in mice, ranging from random chemical (*N*-ethyl-*N*-nitrosourea) mutagenesis coupled with phenotype-driven screens (3, 4) to sequence-based approaches using ES cell technology, such as gene trapping and gene targeting (5, 6).

Gene trapping is a high-throughput approach that is used to introduce insertional mutations across the mouse genome. It is performed with gene trap vectors where the principal element is a gene trapping cassette consisting of a promoterless reporter gene and/or selectable marker gene, flanked by an upstream 3′ splice site (splice acceptor; SA) and a downstream transcriptional termination sequence (polyadenylation sequence; pA). When inserted into an intron of an expressed gene, the gene trap cassette is transcribed from the endogenous promoter in the form of a fusion transcript in which the exon(s) upstream of the insertion site is spliced in frame to the reporter/selectable marker gene. Because transcription is terminated prematurely at the inserted polyadenylation site, the processed fusion transcript encodes a truncated and nonfunctional version of the cellular protein and the reporter/selectable marker (7). Thus, gene traps simultaneously inactivate and report the expression of the trapped gene at the insertion site and provide a DNA tag (gene trap sequence tag, GTST) for the rapid identification of the disrupted gene. Because gene trap vectors insert randomly across the genome, a large number of mutations can be generated in ES cells within a limited number of experiments. Gene trap approaches have been used successfully in the past by both academic and private organizations to create libraries of ES cell lines harboring mutations in single genes (8–11). Collectively, the existing resources cover ≈66% of all protein-coding genes within the mouse genome (12). However, the gene trap vectors that have been used to generate the currently available resources induce only null mutations; mouse mutants generated from these libraries can show only the earliest and nonredundant developmental function of the trapped gene. Therefore, for most of the mutant strains, the significance of the trapped gene for human disease remains uncertain, because most human disorders result from late-onset gene dysfunction. In addition, between 20% and 30% of the genes targeted in ES cells are required for development and cause embryonic lethal phenotypes when transferred to the germ line, precluding functional analysis in the adult (9, 13).

To circumvent the limitations posed by germ-line mutations, conditional gene targeting strategies use site-specific recombination to spatially and temporally restrict the mutation to somatic cells (14). However, targeted mutagenesis in ES cells requires a detailed knowledge of gene structure and organization to physically isolate a gene in a targeting vector. Although the completed sequencing of the mouse genome greatly assists targeted mutagenesis, the generation of mutant mouse strains by this procedure that can handle only one target at a time is time consuming, labor intensive, expensive, and relatively inefficient.

To address this problem, we have developed a generation of gene trap vectors based on a strategy for directional site-specific recombination termed flip-excision (FlEx) (15). The vectors employ two directional site-specific recombination systems that, when activated in succession, invert the gene trap from its mutagenic orientation on the sense, coding strand to a nonmutagenic orientation on the antisense, noncoding strand. We show that mutations induced by these vectors in ES cells can be both repaired and reinduced by site-specific recombination and introduce a new resource of ES cell lines primed for conditional mutagenesis.

Materials and Methods

Plasmids. pFlipROSAβgeo was assembled in pBabeSrf, a modified pBabepuro retroviral vector lacking the promoter and enhancer elements from the 3′ LTR (16). Pairs of the heterotypic frt/F3 and

Abbreviations: GTST, gene trap sequence tag; SA, splice acceptor; RTs, recombinase target sequences; pA, polyadenylation sequence; FlEx, flip-excision; RBBP7, retinoblastoma binding protein 7; Glt28d1, glycosyltransferase 28 domain containing 1 gene.

†F.S., S.D.-Z., P.V.S., M.H., T.F., J.H., C.S., P.R., W.W., and H.v.M. are members of The German Gene Trap Consortium.

‖To whom correspondence should be addressed at: Department of Molecular Hematology, University of Frankfurt Medical School, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany. E-mail: melchner@em.uni-frankfurt.de.

lox511/loxP recombinase target sequences (RTs) were cloned in the illustrated orientation (Fig. 1*A*) into the unique BamHI and EcoRI sites of pBabeSrf, yielding the intermediate plasmid pBLF. RTs were obtained by synthetic oligonucleotide annealing and extension overlap PCR. To enable efficient recombination, 86- and 46-bp spacers were inserted between frt/F3 and loxP/lox511 sites, respectively. To obtain pFlipRosaβgeo, a SAβgeopA cassette derived from the gene trap vector ROSAβgeo (17) was inserted into the SnaBI site of pBLF between the inversely oriented RT pairs. The final pFlipRosaβgeo vector was verified by sequencing. The pFlipRosaCeo vector was obtained from pFlipRosaβgeo by replacing the SAβgeo cassette with the Ceo fusion gene derived from pU3Ceo (16). The final pFlipRosaCeo plasmid was verified by sequencing. Oligonucleotide and primer sequences used in the various cloning steps are available on request.

The pCAGGS-FLPe expression plasmid was a gift from A. Francis Stewart (18). The pCAGGS-Cre expression plasmid was derived from pCAGGS-FLPe by replacing the FLPe cDNA with the Cre cDNA of pSG5Cre (19).

ES Cell Cultures, Infections, and Electroporations. The [C57BL/6J $\times$ 129S6/SvEvTac]F_1 ES cell lines were grown on irradiated or mitomycin C-treated MEF feeder layers in the presence of 1,000 U/ml leukemia inhibitory factor (LIF) (Esgro, Chemicon, Hofheim, Germany) as described in ref. 9.

Gene trap retrovirus was produced in Phoenix-Eco helper cells by using the transient transfection strategy described previously (20). ES cells were infected with the virus-containing supernatants at a multiplicity of infection of <0.5 as described in ref. 9. Gene trap expressing ES cell lines were selected in 130 μg/ml G418 (Invitrogen), manually picked, expanded, and stored frozen in liquid nitrogen.

Electroporations were carried out by using 1×10^7 ES cells, 10 μg of plasmid DNA, and a 400-μF capacitor (Bio-Rad) as described in ref. 5. After incubating for 2 days in medium supplemented with 0.6 μg/ml puromycin (Sigma–Aldrich), the cells were trypsinized and seeded at low density (1,000 cells per dish) onto 60-mm Petri dishes. Emerging clones were manually picked after 9 days and expanded. The resulting cell lines were used for X-Gal stainings and molecular analyses.

Nucleic Acids and Protein Analyses. PCRs were performed according to standard protocols by using 300–500 ng of genomic DNA or 1 μg of reverse-transcribed total RNA in a total volume of 50 μl. The primer sequences used are available on request.

For Northern blotting, poly(A)$^+$ RNA was purified from total RNA by using the Oligotex mRNA-minikit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The mRNA (1–2 μg) was fractionated on 1% formaldehyde/agarose gels, blotted onto Hybond N+ (Amersham Biosciences) nylon membranes, and hybridized to ^{32}P-labeled cDNA probes (Hartmann Analytic, Braunschweig, Germany) in ULTRAhyb hybridization solution (Ambion, Austin, TX) according to manufacturer's instructions. The Glt28d1-cDNA probe was obtained by asymmetric RT-PCR (21) using an antisense primer complementary to exon 10 of the Glt28d1 gene.

Semiautomated 5′-RACE and sequencing were performed as described in ref. 9. The sequences of the generic and vector-specific primers used are available on request.

Western blotting was performed as described in ref. 22, by using anti-RbAp46, (Abcam, Cambridge, U.K.) and lamin A (Santa Cruz Biotechnology) primary antibodies.

GTST Analysis. GTSTs were analyzed as described in ref. 9 by using the following databases: GenBank (release 144), UniGene (build 141), RefSeq (release 8) (all at www.ncbi.nlm.nih.gov), ENSEMBL v26.33 (www.ensembl.org), MGI (www.informatics.jax.org) and GeneOntology (December 2004 release) (www.geneontology.org).

Results

Vector Design. Two gene trap vectors were designed for large-scale conditional mutagenesis in ES cells. The first vector, FlipRosaβgeo, contains a classic SA, β-galactosidase/neomycin phosphotransferase fusion gene (βgeo), pA cassette inserted into the backbone of a promoterless and enhancerless Moloney murine leukemia virus in inverse transcriptional orientation relative to the virus (Fig. 1*A*) (17). The second vector, FlipRosaCeo, is similar to FlipRosaβgeo except that SAβgeo has been exchanged with Ceo, which is an in-frame fusion between the human CD2 cell surface receptor and the neomycin-resistance genes (16). Unlike βgeo, Ceo does not

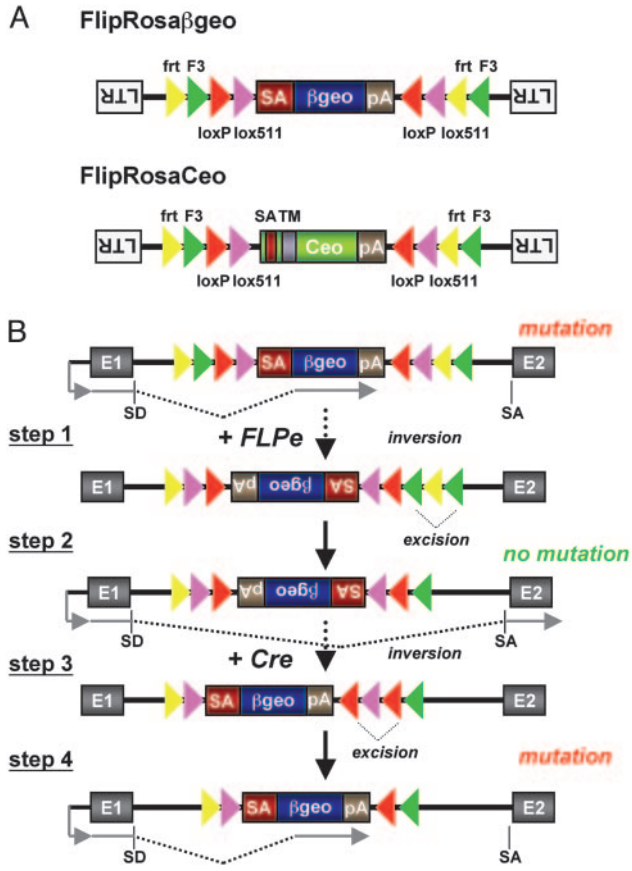

Fig. 1. Conditional gene trap vectors and mechanism of gene inactivation. (*A*) Schematic representation of the retroviral gene trap vectors. LTR, long terminal repeat; frt (yellow triangles) and F3 (green triangles), heterotypic target sequences for the FLPe recombinase; loxP (red triangles) and lox511 (purple triangles), heterotypic target sequences for the Cre-recombinase; SA, splice acceptor; βgeo, β-galactosidase/neomycin phosphotransferase fusion gene; pA, bovine growth hormone polyadenylation sequence; TM, human CD2 receptor transmembrane domain. (*B*) Conditional gene inactivation by a SAβgeopA cassette. The SAβgeopA cassette flanked by recombinase target sites (RTs) in a FlEx configuration is illustrated after integration into an intron of an expressed gene. Transcripts (shown as gray arrows) initiated at the endogenous promoter are spliced from the splice donor (SD) of an endogenous exon (here, exon 1) to the SA of the SAβgeopA cassette. Thereby the βgeo reporter gene is expressed and the endogenous transcript is captured and prematurely terminated at the cassette's pA causing a mutation. In step 1, FLPe inverts the SAβgeopA cassette onto the antisense, noncoding strand at either frt (shown) or F3 (not shown) RTs and positions frt and F3 sites between direct repeats of F3 and frt RTs, respectively. By simultaneously excising the heterotypic RTs (step 2), the cassette is locked against reinversion because the remaining frt and F3 RTs cannot recombine. This inversion reactivates normal splicing between the endogenous splice sites, thereby repairing the mutation. Cre-mediated inversion in steps 3 and 4 repositions the SAβgeopA cassette back onto the sense, coding strand and reinduces the mutation. Note that the recombination products of steps 1 and 3 are transient and transformed into the stable products of step 2 and 4, respectively (15).

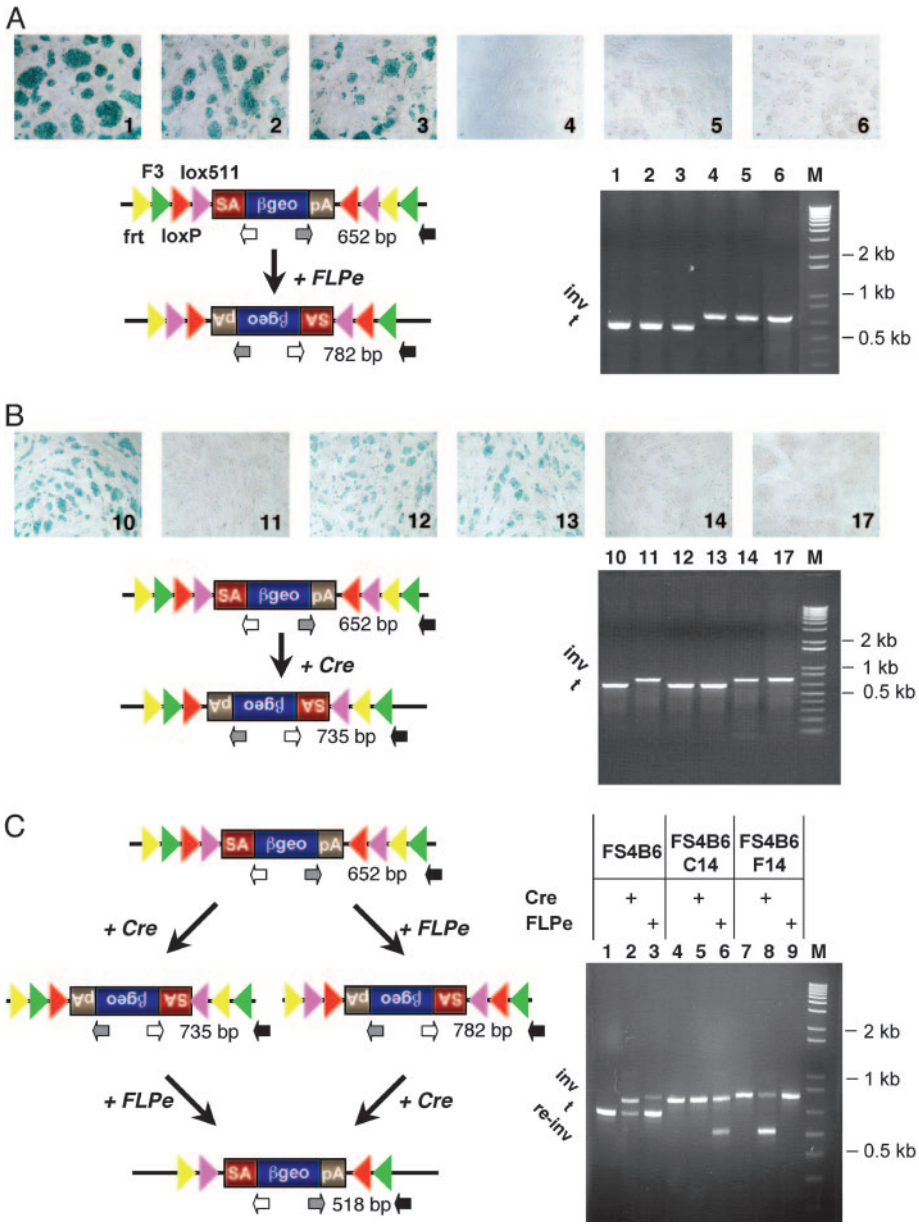

Fig. 2. Site-specific recombinase-induced inversions in FlipRosaβgeo-trapped ES cell lines. (*A* and *B*) ES cells were infected with FlipRosaβgeo virus and selected in G418. X-Gal-positive sublines (blue) were electroporated with FLPe (*A*) or Cre (*B*) expression plasmids and stained with X-Gal after incubating for 10 days. DNA extracted from blue and white sublines was subjected to a multiplex PCR to identify inversions. Primer positions within FlipRosaβgeo are indicated by large arrows; allele-specific amplification products are visualized on ethidium bromide-stained gels to the right. (*C*) Sublines of the FS4B6 ES cell line harboring Cre- or FLPe-inverted gene trap insertions were electroporated with both FLPe- and Cre-expression plasmids. The amplification products obtained from the progeny lines by allele-specific PCR are visualized on the ethidium bromide-stained gel to the right. t, trapped allele; inv, inverted allele; re-inv, reinverted allele; M, molecular weight marker (1 kb + ladder, Invitrogen). FS4B6 (1anes 1–3), parental FlipRosaβgeo-trapped ES cell line; FS4B6 C14 (lanes 4–6), Cre-inverted subline; and FS4B6 F14 (lanes 7–9), FLPe-inverted subline.

require an extra splice acceptor site for trapping because it contains a powerful cryptic 5′ splice site close to its 5′ end. Moreover, Ceo encodes a type II transmembrane domain, which favors the capture of signal sequence and/or transmembrane domain-encoding genes, i.e., secretory pathway genes (Fig. 1*A*) (16). Previous studies involving the isolation of 3,620 ES cell lines with the retroviral gene trap vector (U3Ceo) indicated that Ceo captures secretory pathway genes with >80% efficiency. This is in contrast to the classic βgeo vectors, of which only 19% insert into such genes. Thus, the classic and the secretory pathway gene trap vectors are complementary and, therefore, we equipped both with a conditional mechanism. The mechanism relies on two site-specific recombination systems (FLPe/frt and Cre/loxP) that enable gene trap cassette inversions from the sense, coding strand of a trapped gene to the antisense, noncoding strand and back. As a result, the gene trap vectors allow (*i*) high-throughput selection of gene trap lines by using G418, (*ii*) inactivation of gene trap mutations before ES cell line conversion into mice by blastocyst injection, and (*iii*) reactivation of the mutations at prespecified times and in selected tissues of the resulting mice.

We used an adaptation of a recently published site-specific recombination strategy termed FlEx (15). FlEx uses pairs of inversely oriented heterotypic RTs such as loxP and lox511 or frt and F3. When inserted upstream and downstream of a gene trap cassette, Cre or FLPe recombinases invert the cassette and place a homotypic RT pair near to each other in a direct orientation. Recombination between this pair of directly repeated RTs excises one of the other heterotypic RTs, thereby locking the recombination product against reinversion to the original orientation. Thus, by flanking the gene trap cassettes of FlipRosaβgeo and FlipRosaCeo with pairs of heterotypic lox and frt sites (Fig. 1*A*), a successive delivery of FLPe and Cre to a trapped ES cell line will induce two directional inversions, thereby first repairing and then reinducing the gene trap mutation as exemplified for the SAβgeopA gene trap cassette in Fig. 1*B*.

Gene Trap Cassette Inversions in ES Cells. To test for recombinase-mediated inversions, several FlipRosaβgeo-trapped ES cell lines were selected for high levels of βgeo expression by using X-Gal staining. X-Gal-positive (blue) cell lines were then transiently

transfected with FLPe or Cre expression plasmids, and emerging subclones were stained with X-Gal. As shown in Fig. 2, exposure of the gene trap lines to either FLPe (Fig. 2*A*) or Cre (Fig. 2*B*) yielded a mixture of X-Gal-positive (blue) and X-Gal-negative (white) subclones, indicating that several cell lines have ceased to express βgeo. To test whether this was caused by recombination, we isolated DNA from both the blue and the white sublines, and subjected it to an allele-specific PCR. Fig. 2 *A* and *B* shows that, in each case, the amplification products obtained from the blue and white clones corresponded to a normal and to an inverted gene trap allele, respectively. Taken together, the results indicate that both FLPe and Cre can disrupt the gene trap expression by simply flipping it to the antisense, noncoding strand.

To test whether the FLPe- or Cre-inverted cell lines would reinvert after a second recombinase exposure, we reexpressed FLPe and Cre in each of the cell lines and checked their progeny for reinversions by the allele-specific PCR. Fig. 2*C* shows that FLPe readily reinverted the Cre-inverted subline FS4B6 C14 (lane 6) but not the FLPe-inverted subline FS4B6 F14 (lane 9). Conversely, Cre readily reinverted the FLPe-inverted subline FS4B6 F14 (lane 8) but not the Cre-inverted subline FS4B6 C14 (lane 5). Taken together, the results indicate that gene trap reinversions are inducible only by the recombinase that was not involved in the original inversion, suggesting that the recombination products obtained with either recombinase are stable. Inversions induced by Cre and FLPe in FlipRosaCeo-trapped ES cell lines were similarly stable and efficient (see below). In this context, it is noteworthy that under certain circumstances relating to excessive exposure to Cre enzyme, by long periods of exposure in culture, during development, or by very high levels of Cre expression, some background recombination between heterotypic loxP/lox511 sites can occur (23, 24). However, in gene trap lines stably transduced with a Cre-expression vector, we were unable to detect recombination between loxP and lox511 sites even after several weeks in culture (data not shown), suggesting that background recombination does not significantly affect conditional gene trapping.

Reversibility of Gene Trap Mutations. To test whether the mutations induced by the conditional gene trap vectors are reversible, we selected the Q017B06 and M117B08 gene trap lines for further analysis. In Q017B06, the FlipRosaβgeo gene trap vector disrupted the retinoblastoma binding protein 7 (RBBP7) gene at the level of the first intron. In M117B08, the FlipRosaCeo gene trap vector disrupted the glycosyltransferase 28 domain containing 1 gene (Glt28d1) in the 10th intron. Both genes are located on the X chromosome of a male-derived ES cell line, which provided a haploid background for the mutational analysis. As shown in Figs. 3 *B* and *C* and 4 *B* and *C* , the RBBP7 (Fig. 3) and Glt28d1 (Fig. 4) genes were both expressed in the wild-type cells as expected. However, expression was either blocked (RBBP7, Fig. 3) or severely repressed (Glt28d1, Fig. 4) by the gene trap insertions. Both trapped cell lines instead expressed fusion transcripts as a result of splicing the upstream exons to the gene trap cassettes (Figs. 3*B* and 4 *B* and *C*).

A critical issue that could be addressed with these trapped ES cell lines was whether endogenous gene expression would resume after Cre- or FLPe-induced inversions. Toward this end, we expressed Cre or FLPe in the Q017B06 and M117B08 cell lines, isolated several sublines, and genotyped them by allele-specific PCR (Figs. 3*A* and 4*A*). Inverted sublines were then analyzed for RBBP7, Glt28d1, and gene trap cassette expression by using RT-PCR in combination with Northern and Western blotting. Figs. 3 *B* and *C* and 4 *B* and *C* show that in both cell lines the endogenous gene expression was restored to wild-type levels and the fusion transcripts disappeared, indicating that the antisense gene trap insertions do not interfere with gene expression. Finally, to test whether relocating the gene traps back to their original position on the sense, coding strand would reinduce the mutation, we exposed inverted

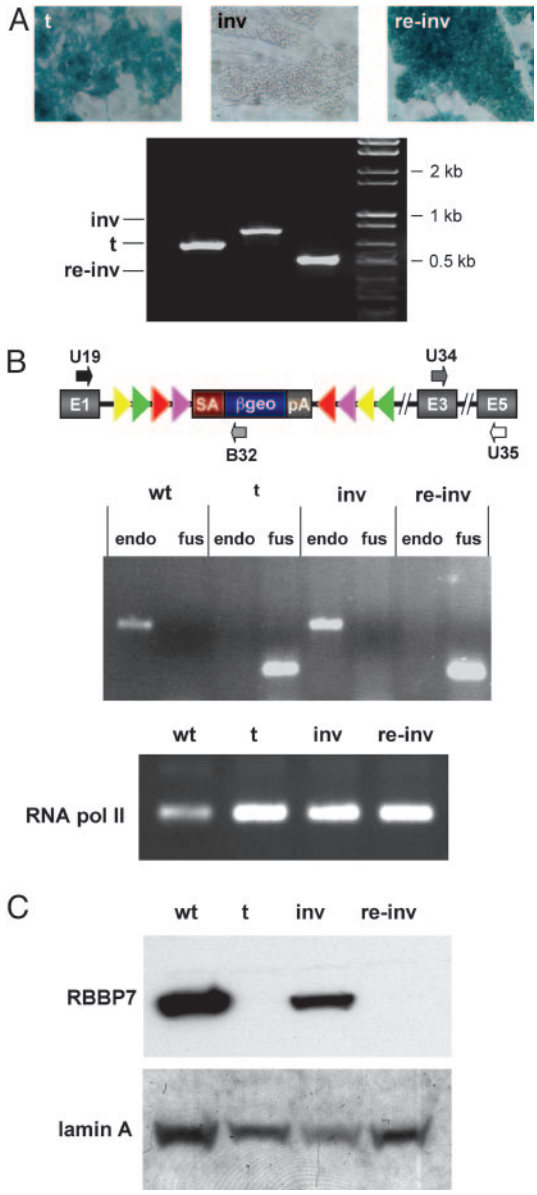

Fig. 3. Conditional mutation induced by a FlipRosaβgeo gene trap insertion in the RBBP7 gene (ENSEMBL ID: ENSMUSG00000031353). The Q017B06 gene trap cell line (t) was transiently transfected with a FLPe expression plasmid, and several sublines with inverted gene trap cassettes were identified by X-Gal staining and allele-specific PCR (inv). Inverted sublines were then electroporated with a Cre-expression plasmid and enriched for reinversions by selecting in G418 (re-inv). (*A*) X-Gal staining (*Upper*) and allele-specific PCR amplification products (*Lower*) from the trapped RBBP7 locus in trapped (t), inverted (inv), and reinverted (re-inv) Q017B06 cell lines. Primers used for the multiplex PCRs were identical to those shown in the diagrams of Fig. 2. (*B*) RT-PCR for the amplification of RBBP7 wild-type and trapped fusion transcripts expressed in Q017B06 cells before and after exposure to FLPe and Cre recombinases. The positions of the primers used are shown on top, wherein U19 = 5′-GCT CTT GAC TAG CGA GAG AGA AG-3′, B32 = 5′-CAA GGC GAT TAA GTT GGG TAA CG-3′, U34 = 5′-CCA GAA GGA AAG GAT TAT GC-3′, and U35 = 5′-ACA GAG CAA ATG ACC CAA GG-3′. Amplification products are visualized below on ethidium bromide-stained gels. Amplification of the RNA polymerase II (RNA pol II) transcript serves as a positive control. wt, parental ES cells; t, trapped Q017B06 cells; inv, inverted Q017B06 subline; re-inv, reinverted Q017B06 subline; endo, endogenous transcript; fus, fusion transcript. (*C*) Western blot analysis of the RBBP7 protein expressed in Q017B06 cells. Crude cell lysates from the F_1 (wt), Q017B06 (t), inverted Q017B06 (inv), and reinverted Q017B06 (re-inv) ES cells were resolved by SDS/PAGE and analyzed by Western blotting using the anti-RbAp46 antibody. The anti-lamin A antibody served as a loading control.

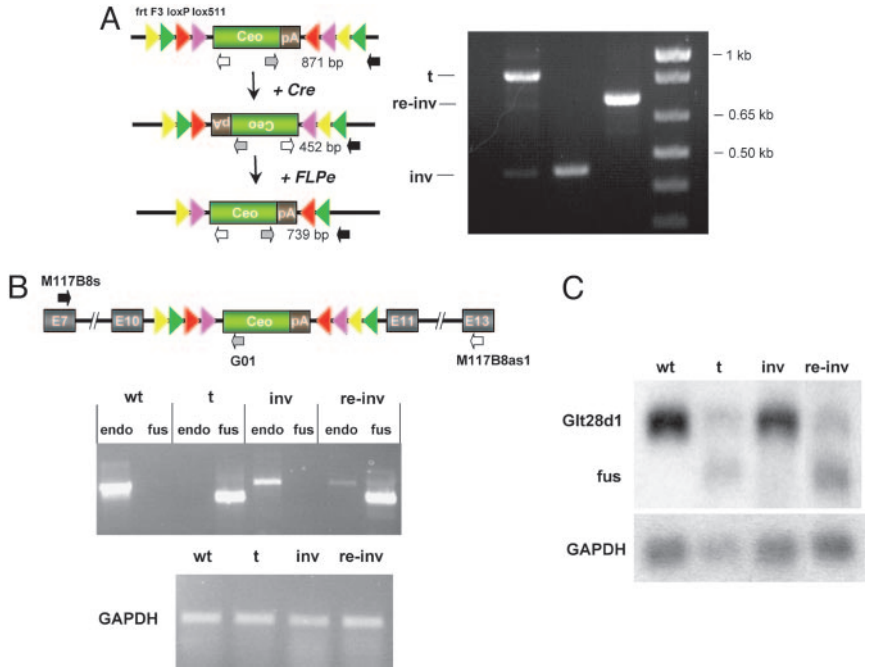

Fig. 4. Conditional mutation induced by a FlipRosa-Ceo gene trap insertion in the Glt28d1 gene (ENSEMBL ID ENSMUST00000040338). The M117B08 gene trap line was treated with recombinases and processed as described for Q017B06 in the legend to Fig. 3, except that Cre was used for the first inversion and FLPe for the second. (*A*) Allele-specific PCR of the trapped Glt28d1 locus in trapped (t), inverted (inv), and reinverted (re-inv) cell lines. (*B*) RT-PCR of Glt28d1wild type and Glt28d1/gene trap fusion transcripts expressed in M117B08 cells before and after exposure to Cre and FLPe recombinases. The positions of the respective primers within the trapped gene are shown on top, wherein M117B8s = 5′-GAG AGT GCT GGC CAG CTG GAA C-3′, G01 = 5′-CAA GTT GAT GTC CTG ACC CAA G-3′, and M117B8as1 = 5′-CCA CCA TAC TCC ACA CAC TCT G-3′. Amplification products are visualized on ethidium bromide-stained gels below. Amplification of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript serves as a positive control. wt, parental ES cells; t, trapped M117B08; inv, inverted M117B08 subline; re-inv, reinverted M117B08 subline; endo, endogenous transcript; fus, fusion transcript. (*C*) Northern blot analysis of Glt28d1 transcripts expressed in M117B08 cells. Samples (2 μg) of polyadenylated RNAs from wild-type (wt), Q017B06 (t), inverted M117B08 (inv), and reinverted M117B08 (re-inv) ES cells were fractionated on 1% formaldehyde/agarose gels and hybridized to a ^{32}P-labeled Glt28d1 cDNA probe. The Glt28d1 probe was obtained by asymmetric RT-PCR using a reverse primer in exon 10 to amplify sequences upstream of the insertions site. The loading of each lane was then assessed by using a GAPDH probe. endo, endogenous transcript; fus, fusion transcript; Glt28d1, Glt28d1 transcript.

subclones to FLPe or Cre. Figs. 3 and 4 show that the reinverted sublines lost the endogenous gene expression and re-expressed the fusion transcripts, like the original trapped lines. The results suggest that the FlipRosaβgeo- and FlipRosaCeo-induced mutations can be repaired and reinduced by the successive activation of the two recombination systems.

Large-Scale Conditional Mutagenesis in ES Cells. We isolated 4,525 ES cell lines with conditional gene trap insertions and recovered 4,138 GTSTs by 5′-RACE. Of these, 3,257 were derived from FlipRosaβgeo and 881 from FlipRosaCeo integrations. Ninety percent of the FlipRosaβgeo and 99% of the FlipRosaCeo GTSTs belonged to RefSeq annotated genes (Table 1). The number of annotated genes was nearly double that found in our previous analysis (9), reflecting the swift progress in genome annotation. The overall efficiency of trapping was similar to that observed in previous studies, as was the number of preferred insertions sites (i.e., hot spots) (Table 1). Insertions occurred in all chromosomes, including one on the Y chromosome, and their number correlated with the number of genes per chromosome (data not shown). Collectively, these observations indicate that the heterotypic lox and frt sites built into the gene trap vectors do not affect the efficiency of trapping. Regardless of the vector, the vast majority of gene trap insertions occurred into first and second introns, confirming the reported preference of retroviral integrations near the 5′ ends of genes (Fig. 5) (25). As expected, the major difference between the vectors was their ability to capture signal sequence genes. Whereas >80% of the FlipRosaCeo insertions were in genes encoding secreted or transmembrane proteins, only 21% of FlipRosaβgeo insertions captured secretory pathway genes according to Gene-Ontology. Thus, like the nonconditional vectors, the two types of conditional gene trap vectors complement each other in gene trapping.

Discussion

In the present study, we have combined gene trap mutagenesis with site-specific recombination to develop an approach suitable for the large-scale induction of conditional mutations in ES cells. The strategy is based on a recently described site-specific recombination strategy (FlEx) (15) that enables directional inversions of gene trap cassettes at the insertion sites. By using gene trap integrations into X-chromosomal genes, we have shown that gene trap vectors equipped with the FlEx system cause mutations that can be repaired and reinduced. Thus, ES cell lines expressing these gene trap vectors can be used for generating mice with either null or conditional mutations. For example, to obtain straight knock-outs the cell lines can be converted directly into mice by blastocyst injection. However, to obtain conditional mutations, one would first repair the mutation in ES cells, preferentially with FLPe to reserve the more efficient Cre for *in vivo* recombination, and then proceed to mouse production. Resulting mice would lack germ-line mutations but would be vulnerable to somatic mutations inducible by Cre. De-

Table 1. Trapping efficiency with conditional gene trap vectors

Gene trap vector	No. of ES cell lines	No. of GTSTs	No. of insertions into annotated genes (%)	No. of unique genes trapped	No. of "hot spots"* (%)
FlipRosaβgeo	3,604	3,257	2,944 (90)	924	505 (16)
FlipRosaCeo	921	881	873 (99)	275	98 (11)
Total	4,525	4,138	3,817 (92)	1,000	603

Analysis is based on National Center for Biotechnology Information mouse genome build 33 and RefSeq release 8.
*All genes with ≥2 insertions were classified as hot spots.

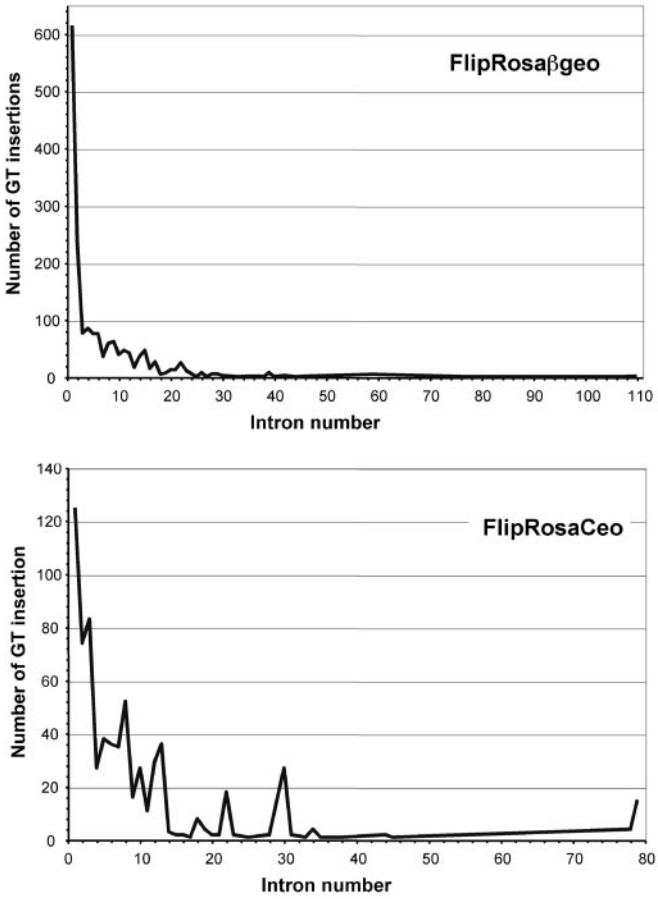

Fig. 5. Distribution of gene trap insertions according to the position of the trapped intron within genes. The data are based on National Center for Biotechnology Information (NCBI) mouse genome build 33 and RefSeq release 8.

pending on the type of Cre and the form of its delivery, the mutations can be reactivated in prespecified tissues at prespecified times.

Because of the inherent recombinase target sites, the vector insertions create multipurpose alleles enabling a large variety of postinsertional modifications by recombinase-mediated cassette exchange (26). Examples include replacing the gene trap cassettes with Cre recombinase genes to expand the Cre-zoo, or with point mutated minigenes to study point mutations. A further option is the insertion of toxin genes for cell lineage-specific ablations.

The quality of the conditional mutations induced by the gene trap insertions will largely depend on the gene trap's ability to be neutral from its position on the antisense, noncoding strand. In the two examples described, the antisense insertions were innocuous, however, this will not always be the case. Factors likely to influence the antisense neutrality include cryptic splice sites and transcriptional termination signals. In line with this, we have shown that aberrant splicing induced by an antisense gene trap insertion resulted in a partial gene inactivation and an interesting phenotype (22). Thus, the most likely outcome of antisense insertions that interfere with gene expression are hypomorphic mutations, which have a merit of their own. However, *in silico* analysis failed to identify sequences that might interfere with gene expression from the antisense strands of the present vectors (data not shown), suggesting that the majority of their insertions create *bona fide* conditional alleles.

By using the vectors in high-throughput screens, we have assembled the largest library of ES cell lines with conditional mutations of single protein-coding genes, including secretory pathway genes. Presently it contains 1,000 potentially conditional alleles (Table 2, which is published as supporting information on the PNAS web site), which is about 10 times the number produced within the last 10 years by gene targeting. Considering that these gene trap lines were isolated in less than a year, conditional gene trapping seems significantly more efficient than conditional gene targeting. However, analysis of the existing gene trap resources indicates that gene trapping is more efficient than gene targeting only up to ≈50% of all mouse genes, after which the mutation rate falls to a level comparable to gene targeting (12). Moreover, effective gene trapping is restricted to the ≈70% of the genes expressed in ES cells (27, 28). Although gene trapping strategies have been described for genes that are not expressed in ES cells, their performance is quite inconsistent, making them unsuitable for high-throughput approaches (29–31). We believe that for a comprehensive mutagenesis of the mouse genome, a balance between gene trapping and gene targeting, performed with generic gene trap cassettes inserted into the targeting vectors, is likely to be the most efficient and cost-effective.

Finally, the ES cell library presented here is freely available to the scientific community. Cell lines can be ordered directly from the German Gene Trap Consortium.

We thank Julia Schmidt, Corinna Strolz, Silke Garkisch, Carsta Werner, Beata Thalke, and Dorotha German for excellent technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft (to H.v.M.) and from the Bundesministerium für Bildung und Forschung (to the German Gene Trap Consortium).

1. 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) *Nat. Genet.* **36,** 921–924.
2. Auwerx, J., Avner, P., Baldock, R., Ballabio, A., Balling, R., Barbacid, M., Berns, A., Bradley, A., Brown, S., Carmeliet, P., *et al.* (2004) *Nat. Genet.* **36,** 925–927.
3. Cox, R. D. & Brown, S. D. (2003) *Curr. Opin. Genet. Dev.* **13,** 278–283.
4. Brown, S. D. & Balling, R. (2001) *Curr. Opin. Genet. Dev.* **11,** 268–273.
5. Floss, T. & Wurst, W. (2002) *Methods Mol. Biol.* **185,** 347–379.
6. Mansouri, A. (2001) *Methods Mol. Biol.* **175,** 397–413.
7. Stanford, W. L., Cohn, J. B. & Cordes, S. P. (2001) *Nat. Rev. Genet.* **2,** 756–768.
8. Wiles, M. V., Vauti, F., Otte, J., Füchtbauer, E. M., Ruiz, P., Füchtbauer, A., Arnold, H. H., Lehrach, H., Metz, T., von Melchner, H., *et al.* (2000) *Nat. Genet.* **24,** 13–14.
9. Hansen, J., Floss, T., Van Sloun, P., Füchtbauer, E. M., Vauti, F., Arnold, H. H., Schnütgen, F., Wurst, W., von Melchner, H. & Ruiz, P. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9918–9922.
10. 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) *Nucleic Acids Res.* **31,** 278–281.
11. 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) *Proc. Natl. Acad. Sci. USA* **100,** 14109–14114.
12. 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) *Nat. Genet.* **36,** 543–544.
13. Mitchell, K. J., Pinson, K. I., Kelly, O. G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P. A., Goodrich, L. V., Lu, X., Avery, B. J., *et al.* (2001) *Nat. Genet.* **28,** 241–249.
14. von Melchner, H. & Stewart, A. F. (2004) in *Handbook of Stem Cells*, eds. Lanza, R., Blau H., Gearhart J., Hogan B., Melton D., Moore M., Pedersen, R., Thomas, E. D., Thomson J., Verfaillie, C., *et al.* (Academic, Oxford), Vol. 1, pp. 609–622.
15. Schnütgen, F., Doerflinger, N., Calleja, C., Wendling, O., Chambon, P. & Ghyselinck, N. B. (2003) *Nat. Biotechnol.* **21,** 562–565.
16. Gebauer, M., von Melchner, H. & Beckers, T. (2001) *Genome Res.* **11,** 1871–1877.
17. Friedrich, G. & Soriano, P. (1991) *Genes Dev.* **5,** 1513–1523.
18. Rodriguez, C. I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., Stewart, A. F. & Dymecki, S. M. (2000) *Nat. Genet.* **25,** 139–140.
19. Feil, R., Wagner, J., Metzger, D. & Chambon, P. (1997) *Biochem. Biophys. Res. Commun.* **237,** 752–757.
20. Nolan, G. P. & Shatzman, A. R. (1998) *Curr. Opin. Biotechnol.* **9,** 447–450.
21. Buess, M., Moroni, C. & Hirsch, H. H. (1997) *Nucleic Acids Res.* **25,** 2233–2235.
22. Sterner-Kock, A., Thorey, I. S., Koli, K., Wempe, F., Otte, J., Bangsow, T., Kuhlmeier, K., Kirchner, T., Jin, S., Keski-Oja, J. & von Melchner, H. (2002) *Genes Dev.* **16,** 2264–2273.
23. Kolb, A. F. (2001) *Anal. Biochem.* **290,** 260–271.
24. Lauth, M., Moerl, K., Barski, J. J. & Meyer, M. (2000) *Genesis* **27,** 153–158.
25. Bushman, F. D. (2003) *Cell* **115,** 135–138.
26. Baer, A. & Bode, J. (2001) *Curr. Opin. Biotechnol.* **12,** 473–480.
27. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. (2002) *Science* **298,** 597–600.
28. Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A. & Lemischka, I. R. (2002) *Science* **298,** 601–604.
29. Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C. & Sands, A. T. (1998) *Nature* **392,** 608–611.
30. Horie, K., Yusa, K., Yae, K., Odajima, J., Fischer, S. E., Keng, V. W., Hayakawa, T., Mizuno, S., Kondoh, G., Ijiri, T., *et al.* (2003) *Mol. Cell. Biol.* **23,** 9189–9207.
31. Osipovich, A. B., Singh, A. & Ruley, H. E. (2005) *Genome. Res.* **15,** 428–435.