MOLECULAR AND CELLULAR BIOLOGY, Dec. 2003, p. 9189–9207 Vol. 23, No. 24
0270-7306/03/$08.00+0 DOI: 10.1128/MCB.23.24.9189–9207.2003

Characterization of *Sleeping Beauty* Transposition and Its Application to Genetic Screening in Mice

Kyoji Horie,[1,2] Kosuke Yusa,[2] Kojiro Yae,[2,3] Junko Odajima,[4] Sylvia E. J. Fischer,[5] Vincent W. Keng,[2,6] Tomoko Hayakawa,[2] Sumi Mizuno,[2,6] Gen Kondoh,[2] Takashi Ijiri,[7] Yoichi Matsuda,[7,8] Ronald H. A. Plasterk,[5] and Junji Takeda[1,2,6]*

Collaborative Research Center for Advanced Science and Technology,[1] Department of Social and Environmental Medicine,[2] Department of Hematology and Oncology,[4] and Japan Science and Technology Corporation,[6] Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Division of Gene Function in Animals, Nara Institute of Science and Technology, Ikoma, Nara 630-0101,[3] and Laboratory of Cytogenetics, Division of Bioscience, Graduate School of Environmental Earth Science,[7] and Laboratory of Animal Cytogenetics, Center for Advanced Science and Technology,[8] Hokkaido University, Kita-ku, Sapporo 060-0810, Japan, and Hubrecht Laboratory, Center for Biomedical Genetics, 3584 CT Utrecht, The Netherlands[5]

Received 7 July 2003/Returned for modification 22 August 2003/Accepted 17 September 2003

The use of mutant mice plays a pivotal role in determining the function of genes, and the recently reported germ line transposition of the *Sleeping Beauty* (*SB*) transposon would provide a novel system to facilitate this approach. In this study, we characterized *SB* transposition in the mouse germ line and assessed its potential for generating mutant mice. Transposition sites not only were clustered within 3 Mb near the donor site but also were widely distributed outside this cluster, indicating that the *SB* transposon can be utilized for both region-specific and genome-wide mutagenesis. The complexity of transposition sites in the germ line was high enough for large-scale generation of mutant mice. Based on these initial results, we conducted germ line mutagenesis by using a gene trap scheme, and the use of a green fluorescent protein reporter made it possible to select for mutant mice rapidly and noninvasively. Interestingly, mice with mutations in the same gene, each with a different insertion site, were obtained by local transposition events, demonstrating the feasibility of the *SB* transposon system for region-specific mutagenesis. Our results indicate that the *SB* transposon system has unique features that complement other mutagenesis approaches.

The analysis of mutant mice plays a key role in the understanding of gene functions, and the importance of this approach is expected to increase (2) with the recent availability of the mouse genome sequence (25). However, large-scale genetic screening for mice has been lagging far behind that for other model organisms, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, because of the lack of a system allowing for both mutagenesis and subsequent rapid identification of the mutation. Large-scale generation of mutant mice has been conducted recently by using *N*-ethyl-*N*-nitrosourea (ENU), and a number of mutant mice with various phenotypes have been generated successfully (6, 26). The drawback of this approach is that identification of the causative point mutations is time-consuming. The embryonic stem (ES) cell-based gene trap is another effective approach (11, 12, 31). However, large-scale generation of mutant mice, which is a prerequisite for genetic screening, is not easy because the ES cell-based methods involve labor-intensive processes such as tissue culture or embryo manipulation.

Transposon-tagged mutagenesis has been used in a wide range of organisms, such as *D. melanogaster* (1, 32), *C. elegans* (13, 24), and plants (27). Although the mutation rate resulting from transposition is not as high as that in ENU mutagenesis, transposon-tagged mutagenesis has been used as an alternative genetic screening method for the following reasons. First, the genes responsible for the phenotypes can be identified rapidly by using the transposon sequence as a tag. Second, desired elements can be introduced into the transposon sequence to expand the application range of the mutant lines. This principle has been demonstrated in the *P* element of *D. melanogaster*, where various GAL4 enhancer trap lines have been created by *P*-element transpositions and used for the expression of a gene of interest in specific tissues (3) or for the elimination of specific cells by expression of a toxin gene (15). Misexpression of the genes downstream of the insertion sites could also be achieved by inducing transcription from a promoter introduced into the transposon (29). However, application of the transposon system to mammals has been hampered until recently by the absence of an efficient transposon.

Sleeping Beauty (*SB*) is a synthetic *Tc1*/*mariner*-like transposon system that was reconstructed from sequences found in salmonid fish (18). We (10, 16) and others (8) have reported recently that the *SB* transposon transposes efficiently in mice. In the present study, we addressed issues that are crucial for assessing the potential of the *SB* transposon system for large-scale mutagenesis in mice. The first is the distribution of transposition sites. Although the *SB* transposon has been shown to jump preferentially to the chromosome bearing the original integration site (8, 10, 21), no extensive or detailed analyses have been made of the transposition sites, such as of the distance from the original integration site and the frequency of

* Corresponding author. Mailing address: Department of Social and Environmental Medicine H3, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan. Phone: 81 6 6879 3262. Fax: 81 6 6879 3266. E-mail: takeda@mr-envi.med.osaka-u.ac.jp.

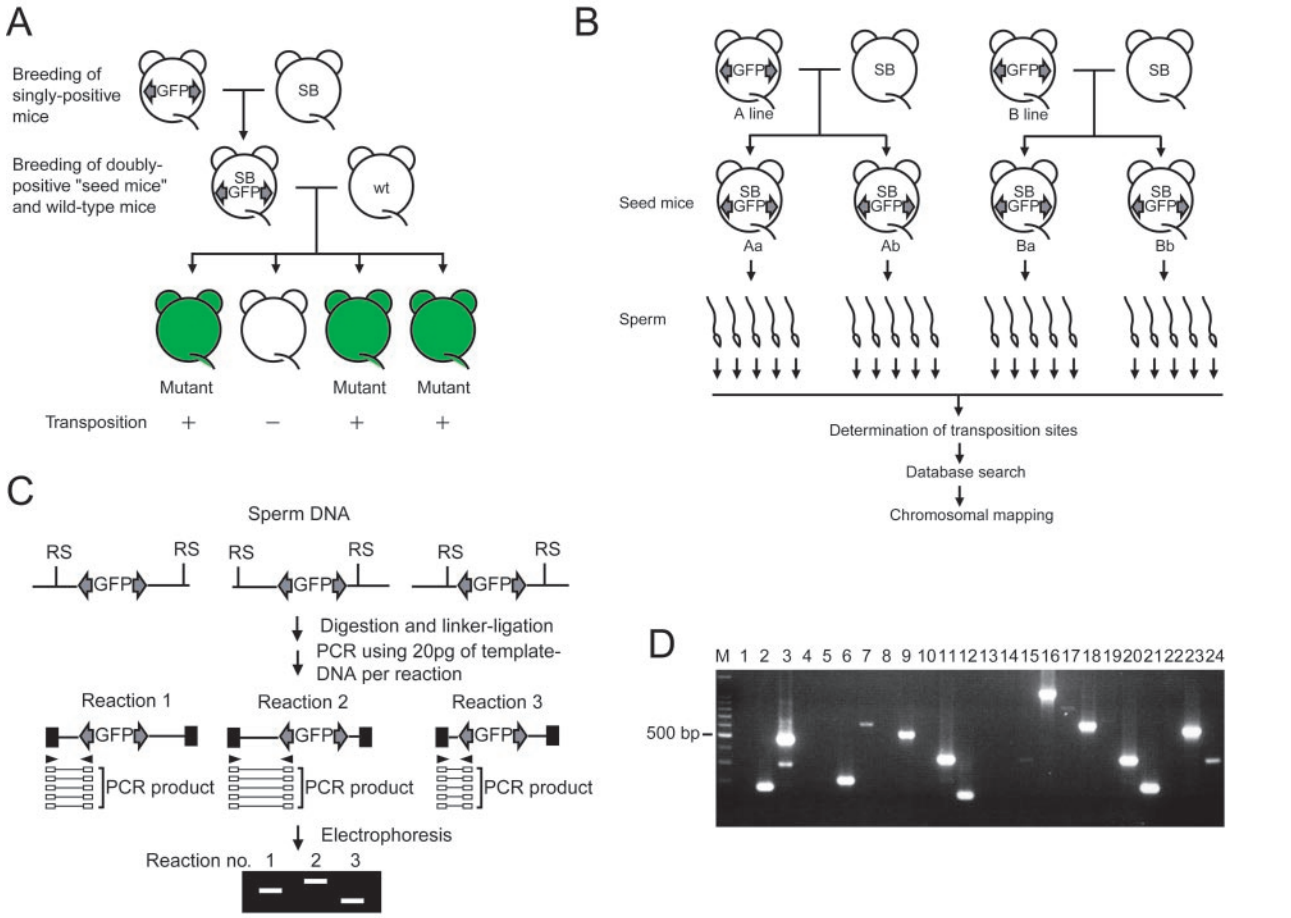


FIG. 1. Determination of *SB* transposition sites in the mouse germ line. (A) Overview of the breeding scheme used to generate mutant mice. grey arrows, IR and DR of the transposon that is the recognition sequence of the SB transposase. GFP was used as a marker to monitor transposition events; other elements located between IRs and DRs are not shown for simplicity. Two kinds of singly positive mice are generated; one carries the transposon vector, and the other expresses the SB transposase. Doubly transgenic mice (seed mice) are obtained as a result of breeding of the singly positive mice and are mated with wild-type mice. Mutant mice are generated as a result of a transposition event in the germ lines of seed mice. (B) Strategy used to study transposition sites in the sperm of seed mice. (C and D) Ligation-mediated PCR to determine independent transposition sites from sperm DNA. (C) PCR procedure; (D) amplification of one copy of a transposition site per reaction. RS, restriction site; filled boxes, linker DNAs; Rx, reaction. The marker (lane M) is a 100-bp DNA ladder.

transposition to other chromosomes. Distribution of the insertion sites with respect to the endogenous genes was addressed as well, since some transposons have a preference in this respect. For example, the *P* element has been reported to transpose with high frequency into the promoter regions or the 5′ untranslated regions of the genes (32). The second issue is the complexity of the transposition sites in the germ lines of mice bearing both the transposon and the transposase (the seed mice in Fig. 1A). This will determine how many different mutant mice can be obtained in the progeny. Based on these findings, a gene trap procedure was conducted with the mouse germ line for the rapid generation and analysis of mutant mice. The results demonstrate that both genome-wide and region-specific mutageneses are feasible. Furthermore, the analysis of homozygous mice demonstrated that our transposon vector was highly mutagenic. In combination, these results indicate that the *SB* transposon system represents a powerful genetic screening system for gene function analysis in mice.

MATERIALS AND METHODS

Construction of trap vector and generation of transgenic mice. Cloning sites of pBluescript II (Stratagene) were replaced with *Asc*I, *Xho*I, *Not*I, and *Swa*I sites via PCR amplification of pBluescript II with primers 5′ GCCGCTCGAGGGC GCGCCAGATTTAAATCAGCTTTTGTTCCCTTTAGTGAG 3′ and 5′ CGC AGCGGCCGCATTTAAATGAGGCGCGCCGCTCCAATTCGCCCTAT AGTG 3′. A 2.9-kb *Xho*I-*Not*I fragment of the PCR product was ligated to a 0.8-kb *Xho*I-*Not*I fragment of IR/DR(R,L) from pBS-IR/DR(R,L) (16), resulting in pBS-IR/DR-AS, which contains *Asc*I and *Swa*I sites flanking the inverted repeats (IRs) and direct repeats (DRs).

Linker DNAs containing *Asc*I-*Kpn*I-*Swa*I sites and *Pme*I-*Pac*I sites were created by annealing oligonucleotides 5′ GTACGGCGCGCCGGTACCATTT AAAT 3′ and 5′ GTACATTTAAATGGTACCGGCGCGCC 3′ and oligonucleotides 5′ CGTTTAAACTTAATTAAGAGCT 3′ and 5′ CTTAATTAAGTT TAAACGAGCT 3′, respectively. Each linker was inserted into the unique *Kpn*I and *Sac*I sites, respectively, of pTransCX-*GFP*:*Neo* (16) after the removal of the TransCX-*GFP* fragment, resulting in pAKS:*Neo*:PP.

The *Sal*I-*Bam*HI fragment of pCX-*EGFP-PigA* (16), containing CAG-*EGFP*, and the 256-bp *Bam*HI-blunted *Xho*I fragment of the *Neo* cassette (17), consisting of splice donor (SD) sequences from the mouse *hprt* gene exon 8/intron 8 region and the mRNA instability signal derived from the 3′ untranslated region

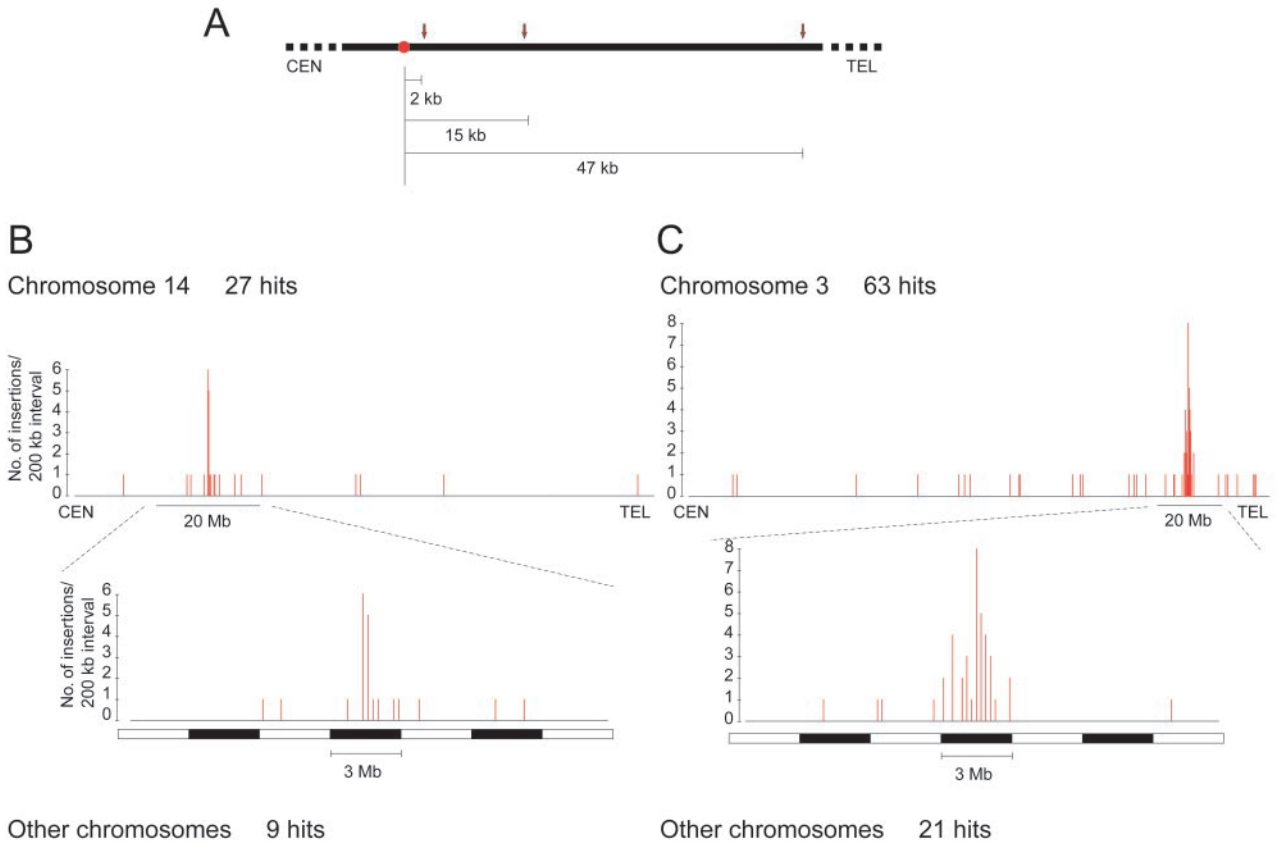


FIG. 2. Distribution of *SB* transposition sites in the mouse germ line. (A) Distribution of transposition sites from a single copy of the *SB* transposon at the DST. Twelve transposition sites described previously (10) were analyzed again by using the Celera Genomics database, and three transposition sites located within 200 kb of the DST are shown. CEN, centromeric region; TEL, telomeric region. The DST is depicted by a filled circle. Arrows indicate transposition sites of loci 1835, 1818, and 1680 (Table 1). (B and C) Distribution of transposition sites from the seed mice of line A (B) and line B (C) according to the Ensembl mouse genome database. The DST was mapped by FISH to chromosome 14 B distal-C1 proximal in line A (data not shown) and to chromosome 3 H1-H2 in line B as reported previously (16). Chromosomes bearing DSTs were divided into 200-kb intervals, and the number of transposon insertions per interval was plotted. The 20-Mb regions around the cluster of insertions are shown magnified as well, together with a 3-Mb scale of black and white boxes. Note that although some transposition sites are clustered, most of them were mapped to different locations at the nucleotide level (see Table 2).

of the human granulocyte-macrophage colony-stimulating factor cDNA (17), were inserted into *Sal*I-blunted *Not*I sites of pBluescript II, resulting in pCAG-*GFP*-SD.

An *Xba*I–blunted-*Hin*dIII fragment of the rabbit β-globin poly(A) addition signal and a *Sac*II-*Not*I fragment of the *lacZ* gene containing the nuclear localizing signal were isolated from pRTonZ (K. Horie et al., unpublished data) and were sequentially inserted at *Xba*I and *Sma*I sites and *Sac*II and *Not*I sites of pBluescriptII, respectively, resulting in p*LacZ*-BS. After the insertion of the *lacZ*-poly(A), the *Hin*dIII site was deleted by blunting and subsequent self-ligation. A *Sal*I-*Xho*I fragment of CAG-*GFP*-SD from pCAG-*GFP*-SD was inserted at the *Xho*I site of p*LacZ*-BS, resulting in p*LacZ*-CAG-*GFP*-SD. The human *bcl-2* intron 2/exon 3 splice acceptor (SA) sequence was amplified by using primers 5′ CGGCAAGCTTCTCGAGCTGTATCTCTAAGATGGCTGG 3′ and 5′ GCCACGGTCGACGCCTGCATATTATTTCTACTGC 3′, with the RET vector (17) serving as a template. The internal ribosome entry site (IRES) sequence was amplified with primers 5′ GGAGCGTCGACTACGTAAATTCC GCCCCTCTCCCTC 3′ and 5′ GGAGCGTCGACTACGTAAATTCTCCCT CCCC 3′, again with the RET vector serving as a template. A *Hin*dIII-*Sal*I SA-containing fragment and a *Sal*I-*Bam*HI IRES-containing fragment were simultaneously cloned into the *Hin*dIII and *Bam*HI sites of p*LacZ*-CAG-*GFP*-SD, resulting in pSA-IRES*LacZ*-CAG-*GFP*-SD.

The *Xho*I fragment of pSA-IRES-*LacZ*-CAG-*GFP*-SD containing SA, IRES, *lacZ*, poly(A), and CAG-*GFP*-SD was blunted and inserted at the *Eco*RI and *Bam*HI sites of pBS-IR/DR-AS after both sites were blunted, resulting in pTrans-SA-IRES*LacZ*-CAG-*GFP*-SD. The *Asc*I-*Swa*I fragment of pTrans-SA-IRES*LacZ*-CAG-*GFP*-SD was inserted at the *Asc*I and *Swa*I site of pAKS:*Neo*:PP, resulting in pTrans-SA-IRES*LacZ*-CAG-*GFP*-SD:*Neo*. pTrans-SA-IRES*LacZ*-CAG-*GFP*-SD:*Neo* was linearized with *Pac*I and injected into BDF1 × BDF1 fertilized eggs.

This vector contains the same vector backbone that was used in our previous study (16), in which the CAG promoter was epigenetically repressed. Since multiple copies of the vector DNA are likely to be integrated in a head-to-tail array at donor sites for transposition (DSTs), splicing may occur between the SD site of the green fluorescent protein (GFP) gene and the SA site within the downstream vector unit, resulting in GFP expression from DSTs. The expression of GFP from the tandem array would be suppressed by leaving the vector backbone in the injected vector DNA, because the CAG promoter will be inactivated in this configuration but will be activated at transposition sites in the next generation (16). As predicted, GFP signal was not detected in most (seven of eight) founder mice. Transgenic lines were crossed with the *SB* line (16) to generate doubly transgenic mice (referred to as seed mice). Genotyping was performed with primers specific for the GFP and *SB* genes as described before (16). Doubly transgenic mice were mated with female ICR mice, and mutant mice were obtained among the progeny. All animal studies were done in compliance with Osaka University guidelines.

Examination of GFP expression. Newborn mice were examined with a fluorescence stereomicroscope with GFP-specific filters (WILD M10; Leica). Screening was performed before the appearance of coat color to avoid reduction of detection sensitivity.

Determination of transposon insertion sites by ligation-mediated PCR. Sequences flanking the transposon insertion sites were identified by ligation-medi-

TABLE 1. Distribution of transposon insertion sites in the protamin-*SB* system[a]

Locus[b]	Chromosome	Chromosome position		Gene hit[c]			
		Ensembl	Celera	Ensembl ID	Celera ID	Gene product name	Insertion site
1657	5	80221197	74946648	None	mCG53983	PD	Intron
1797	5	80776245	75500536	None	mCG11793	PD	Intron
1814	5	82394376	77115433	None	None	None	None
1820*1	5	84487210	79158207	None	mCG1046448	PD	Intron
1842*1	5	84604878	79278427	None	mCG1046448	PD	Intron
1682*1	5	84606566	79280115	None	mCG1046448	PD	Intron
1775*1	5	84608228	79281779	None	mCG1046448	PD	Intron
1688	5	85060113	79720540	None	None	None	None
1633	5	87009939	81623030	ENSMUSESTG00000040083	None	EST	Downstream
				ENSMUSESTG00000040079	None	EST	Downstream
				None	mCG1046107	PD	Upstream
Founder*2	5	119805335	115831871	ENSMUSG00000042605	mCG12184	Ataxin 2	Upstream
1835*2	5	119807381	115833752	ENSMUSG00000042605	mCG12184	Ataxin 2	Upstream
1818*2	5	119820208	115846483	ENSMUSG00000042605	mCG12184	Ataxin 2	Upstream
1680*2	5	119852710	115879093	None	mCG12184	Ataxin 2	Intron
1836	7	55378386	60397356	ENSMUSG00000030513	mCG19967	PACE4	Intron
1576	12	55914072	59608311	None	None	None	None

[a] Transposon insertion sites that were mapped previously by means of a radiation hybrid panel (10) were analyzed again as described in the footnotes to Table 2.
[b] The locus names correspond to those previously reported (10). Founder indicates the founder mouse bearing a single copy of the transposon prior to transposition, and its transposon integration site is defined as the DST in Fig. 2A. The loci that were mapped to the same gene are indicated by *1 and *2, each representing an insertion into the same gene.
[c] Definitions of ID, PD, upstream, and downstream can be found in the footnotes to Table 2.

ated PCR as reported previously (7, 18) with some modifications. Briefly, genomic DNA was extracted from sperm or testis then digested with *Bgl*II or *Xba*I, and splinkerettes (7) were ligated to the cleavage ends. Splinkerettes compatible with *Bgl*II or *Xba*I were generated by annealing the oligonucleotide Spl-top (5′ CGAATCGTAACCGTTCGTACGAGAATTCGTACGAGAATC GCTGTCCTCTCCAACGAGCCAAGG 3′) with SplB-Bgl (5′ GATCCCTTG GCTCGTTTTTTTTTGCAAAAA 3′) or SplB-Xba (5′ CTAGCCTTGGCTCG TTTTTTTTTGCAAAAA 3′), respectively. Ligation products were diluted prior to PCR so that only a single molecule of the target DNA would be amplified (see the legend to Fig. 1). Since *Bgl*II and *Xba*I sites do not exist in vector sequences outside the transposon region, amplification from the vector concatemer could be avoided.

Estimation of the complexity of transposition sites. Conditions for nested PCR were set up to detect a single molecule of the junction fragment between transposon DNA and genomic DNA in three transposition sites each from Aa line (Aa3, Aa4, and Aa28) and Ba line (Ba49, Ba52, and Ba56). The outer primers specific to the transposition sites were 5′ TACAAGACTAACAACCA CCTTACTCATTC 3′ for Aa3, 5′ GGAGTTCCCATAAGGCAAGATAGAGC CAGG 3′ for Aa4, 5′ GTAATATCTCTGAAAGTTGGGGGCTCTT 3′ for A28, 5′ AGGAGAATGCAGAGGGACTCAGAGAATGG 3′ for Ba49, 5′ TC TCCATGATCAAGAAATTCCCCAGCTAAC 3′ for Ba52, and 5′ GACCAAC AACAGCTATAATCTCATAACAGC 3′ for Ba56. The inner primers for the sites were 5′ CCTTGTAAGTGTGATAACTGTCCTAGTTTG 3′ for Aa3, 5′ CCTGTGTGGAGAAGTAGTGATTCGTTC 3′ for Aa4, 5′ GAATCTACCCT CAGAGTTTGAAGCCAAA 3′ for A28, 5′ ATCCTGTGAGGTGCAAGTGT GAGAG 3′ for Ba49, 5′ CAGTAAGTTGAACTTCCAACGTGGAG 3′ for Ba52, and 5′ CTTCCAAAATTCACCAATAACACTCATC 3′ for Ba56. The outer and inner primers specific to the transposon region were 5′ CTTGTGTC ATGCACAAAGTAGATGTCC 3′ and 5′ CCTAACTGACTTGCCAAAACT ATTGTTTG 3′, respectively. Nested PCR was performed with the HotStarTaq system (Qiagen) under the following conditions: 1 cycle of 95°C for 15 min; followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min; followed by 1 cycle of 72°C for 10 min. One microliter of the first PCR product was used as a template in the second PCR. To examine the sensitivity of detecting the junction region, ligation-mediated PCR products from which each insertion site was determined (as described in the previous paragraph) were purified with a PCR purification kit (Qiagen), and 10 PCRs were performed with one molecule of the ligation-mediated PCR product as a template for each reaction. Serially diluted testis DNA samples were used as templates to examine the frequency of each transposition site, and the complexity was estimated as described in Results.

Examination of the mutagenicity of transposon insertion by reverse transcription-PCR. cDNA was synthesized from 2 μg of total RNA extracted from the tails of line TM67 and TM88 mice by using Superscript II (Invitrogen) with gene-specific primers. The primer for line TM67 was 5′GTTTGGGGTGAGTG TTTGCTTTCTTGTCTG 3′, and that for line TM88 was 5′ GGTTTCCTTGG GTTTTGATGTTCTGATGAG 3′. To examine the expression of the gene bearing the transposon insertion, 0.1 μg of cDNA was PCR amplified with the HotStarTaq system (Qiagen) by using primer pairs annealed to the exons flanking the transposon insertion site. Amplification conditions were as follows: 1 cycle of 95°C for 15 min; followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min; followed by 1 cycle of 72°C for 7 min. Primer sequences were as follows: upstream primer for line TM67, 5′ GCCAAGGAGGAAACAGCA GGCACCCAAGCG 3′; downstream primer for line TM67, 5′ CTCGTTTTCT GCATCCTGATTGGACAGGTG 3′; upstream primer for line TM88, 5′ CAG TCAAGAGAAGCATCCCTCCAGAAACAG 3′; downstream primer for line TM88, 5′ TTCATCCATGCATTTAGAGAGTGGTTGTAG 3′. The EGFP5U primer (5′ GCGATCACATGGTCCTGCTGGAGTTCGTG 3′) was used with each downstream primer to amplify the transcript generated with the poly(A) trap scheme. This reaction also served as a control for the integrity of RNAs from heterozygous and homozygous mice.

3′ Rapid amplification of cDNA ends (3′ RACE). Total RNA was isolated from the tails of GFP-positive mice by using TRIzol (Invitrogen). cDNA was synthesized by using the 92-nucleotide oligo(dT) primer 5′ GGAGCAAGCAG TGGTAACAACGCAGAGTACCGATCAGTTGCTCTGGTGTCCGTGTCC TACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN 3′ with Superscript II (Invitrogen) according to the manufacturer's recommendations. Trapped sequences were subsequently amplified by using nested PCR and the Expand high-fidelity PCR system (Roche Diagnostics). Conditions used were as follows: 94°C for 5 min; followed by 20 cycles of 94°C for 1 min, 60°C for 1 min, and 68°C for 1 min; followed by 68°C for 5 min. Primers for the first reaction were EGFP4U (5′ CCCTGAGCAAAGACCCCAACGAGAAGC 3′) and RC1 (5′ GGAGCAAGCAGTGGTAACAACGCAGAGTAC 3′), and primers for the second reaction were EGFP5U (5′ GCGATCACATGGTCCTGCTGGAGTT CGTG 3′) and RC2 (5′ CGATCAGTTGCTCTGGTGTCCGTGTCCTAC 3′). The PCR products were directly sequenced.

β-Galactosidase staining. Tissues or embryos were fixed with 1% paraformaldehyde–0.2% glutaraldehyde–0.02% NP-40 in phosphate-buffered saline (PBS) (pH 7.3) for 15 to 60 min, washed in PBS with 0.02% NP-40 three times, and stained for 3 to 6 h at 37°C in a solution containing 1 mg of 4-chloro-5-bromo-3-indolyl-β-D-galactopyranoside (X-Gal) (Sigma) per ml, 2 mM $MgCl_2$, 4 mM $K_3Fe(CN)_6$, and 4 mM $K_4Fe(CN)_6$ in PBS (pH 7.3). To examine *lacZ* gene induction in the thymuses of TM88 mice, animals at the postnatal age of 10 days were injected intraperitoneally with dexamethasone (30 μg/g of body weight),

TABLE 2. Distribution and transposition sites in lines A and B[a]

Locus[b]	Chromosome	Chromosome position[c]		Gene hit[d]			
		Ensembl	Celera	Ensembl ID	Celera ID	Product name	Insertion site
Aa1	5	133339641	128744770	ENSMUSG00000029675	mCG16716	Elastin precursor	Intron
Aa2	6	33263044	30087408	None	mCG1029516	PD	Intron
Aa3	7	84409604	88953421	None	None	None	None
Aa4	9	124539645	120559087	ENSMUSG00000025786	mCG117266	RIKEN cDNA 1810006O10	Intron
				ENSMUSG00000025785	None	RIKEN cDNA 2610002K22	Upstream
Aa5	14	18064759	16317947	ENSMUSG00000021778	None	RIKEN cDNA 1700112E06	Intron
Aa6	14	30096601	28201695	ENSMUSG00000041093	mCG50091	Glutamate receptor delta chain	Intron
Aa7	14	30960037	29087128	None	None	None	None
Aa8	14	33434637	31519205	ENSMUSESTG00000017053	mCG1034632	RIKEN cDNA 4930581F23	Intron
Aa9*[1]	14	34035613	32117336	None	mCG52624	PD	Intron
Aa10*[1]	14	34036471	32118194	None	mCG52624	PD	Intron
Aa11*[1]	14	34046049	32129793	None	mCG52624	PD	Intron
Aa12*[1]	14	34061403	32145076	None	mCG52624	PD	Intron
Aa13*[1]	14	34062455	32146128	None	mCG52624	PD	Intron
Aa14*[1]	14	34146022	32230175	None	mCG52624	PD	Intron
Aa15*[1]	14	34239318	32323804	None	mCG52624	PD	Intron
Aa16*[1]	14	34350876	32435487	None	mCG52624	PD	Intron
Aa17	14	34383025	32467640	None	None	None	None
Aa18	14	34571020	32654834	None	mCG1034432	PD	Intron
Aa19	14	34720654	32806705	None	None	None	None
Aa20	14	35357392	33431825	None	None	None	None
Aa21	14	36288530	34335172	ENSMUSG00000021795	mCG10221	Surfactant associated protein D	Intron
Aa22	14	39282319	40628445	None	None	None	None
Aa23	14	40524943	41877684	ENSMUSESTG00000018516	None	EST	Intron
				None	mCG12117	Otx2	Upstream
Aa24	14	44414710	46537768	ENSMUSESTG00000022818	None	EST	Upstream
				None	mCG18700	PD	Upstream
Aa25	14	62261532	65877739	None	mCG60014	Proteoglycan 3	Intron
Aa26	14	63049150	66650853	None	mCG1037549	PD	Upstream
Aa27	14	78965759	82384631	None	None	None	None
Aa28	15	42404980	39726489	None	mCG1044549	PD	Intron
Ab1	4	50509424	48625949	None	None	None	None
Ab2	8	110975125	112292150	ENSMUSESTG00000041903	mCG61536	EST	Downstream
				ENSMUSESTG00000041911	None	EST	Upstream
Ab3	10	92362613	91142273	ENSMUSESTG00000002022	None	EST	Intron
				None	mCG62941	PD	Upstream
Ab4	12	14320535	11576377	None	None	None	None
Ab5*[1]	14	34219590	32304087	None	mCG52624	PD	Intron
Ab6	14	34385669	32470284	None	None	None	None
Ab7	14	35540838	33615751	ENSMUSG00000037842	mCG49009	Protein tyrosine phosphatase IVA	Upstream
Ab8	14	115846358	121067394	ENSMUSG00000025551	None	FGF-14	Intron
				None	mCG1025834	PD	Intron
Ba1	1	73055946	69942808	ENSMUSG00000026187	mCG121781	Ku autoantigen	Intron
Ba2	3	12507203	9455812	None	mCG1028528	PD	Upstream
Ba3	3	13772180	10716340	None	None	None	None
Ba4	3	47252396	43663327	None	mCG1049778	PD	Intron
Ba5	3	64738743	60756385	ENSMUSG00000027824	mCG6370	Putative pheromone receptor V2	Intron
Ba6	3	76346410	72743600	None	None	None	None
Ba7	3	79419969	No hit	None	None	None	None
Ba8	3	90606788	86803854	ENSMUSG00000027950	mCG22243	Chrnb2	Upstream
				ENSMUSG00000042579	None	RIKEN cDNA 4632404H12	Downstream
Ba9	3	92937841	No hit	None	None	None	None
Ba10	3	93338427	91452377	None	mCG1042765	PD	Upstream
Ba11	3	108208659	107046014	None	None	None	None
Ba12	3	110468799	109301041	ENSMUSG00000027819	None	Netrin G1	Intron
Ba13	3	111191473	110015146	None	None	None	None
Ba14	3	124274422	No hit	None	None	None	None
Ba15	3	125666710	128699579	None	None	None	None
Ba16	3	126228269	129185593	None	None	None	None
Ba17	3	136872386	139812954	None	None	None	None
Ba18	3	139247325	No hit	ENSMUSG00000028152	None	Tspan-5	Intron
Ba19	3	139655523	142578508	None	None	None	None
Ba20	3	139734836	142630806	ENSMUSESTG00000013903	mCG1045576	EST	Upstream
Ba21	3	140033976	142957610	None	None	None	None
Ba22	3	140115986	143021269	None	None	None	None
Ba23	3	140177697	143087243	None	None	None	None

Continued on following page

TABLE 2—*Continued*

Locus[b]	Chromosome	Chromosome position[c]		Gene hit[d]			
		Ensembl	Celera	Ensembl ID	Celera ID	Product name	Insertion site
Ba24	3	140737961	143644800	None	None	None	None
Ba25	3	141003421	143912948	None	None	None	None
Ba26*[2]	3	141129016	144038837	None	mCG1045798	PD	Intron
Ba27*[3]	3	141226208	144135636	None	mCG1045797	PD	Intron
Ba28*[3]	3	141226409	144135834	None	mCG1045797	PD	Intron
Ba29*[3]	3	141251735	144169231	None	mCG1045797	PD	Intron
Ba30*[4]	3	141442119	144355226	None	mCG1045552	PD	Intron
Ba31*[4]	3	141453041	144366202	None	mCG1045552	PD	Upstream
Ba32	3	141662278	144583162	None	None	None	None
Ba33	3	141687661	144608744	None	None	None	None
Ba34	3	141957604	144873099	None	mCG1045645	PD	Intron
Ba35	3	142406420	145316721	ENSMUSESTG00000008780	None	EST	Intron
Ba36	3	142479121	145393050	None	None	None	None
Ba37	3	149178150	152125671	None	mCG1045751	PD	Intron
Ba38	3	151453999	No hit	None	None	None	None
Ba39	3	151936034	154829309	None	None	None	None
Ba40	3	154761954	157648380	ENSMUSG00000028202	None	EST	Intron
				None	mCG1045569	PD	Upstream
Ba41	3	159073309	161916377	None	mCG11661	PD	Upstream
Ba42	3	159422563	162257755	None	mCG1045689	PD	Intron
Ba43	4	109016854	106620571	ENSMUSESTG00000012197	None	EST	Downstream
				None	mCG7419	PD	Downstream
Ba44	5	30712555	26517380	ENSMUSG00000037511	None	EST	Intron
Ba45	5	74806690	69770958	None	None	None	None
Ba46	5	97900137	94681456	ENSMUSG00000029324	None	RIKEN cDNA 1810024J13	Downstream
				None	mCG7543	PD	Upstream
Ba47	5	149049499	144802025	ENSMUSESTG00000018949	None	EST	Downstream
				None	mCG1029339	PD	Intron
Ba48	7	41962607	46721472	None	mCG1033115	PD	Upstream
Ba49	8	25899555	25649727	ENSMUSG00000031488	mCG2247	RIKEN cDNA 4833414G05	Intron
				ENSMUSESTG00000035764	None	EST	Upstream
Ba50	8	66760649	66917274	ENSMUSESTG00000029727	None	EST	Upstream
				None	mCG1048987	PD	Upstream
Ba51	9	53636482	47566899	ENSMUSG00000034584	mCG129220	EST	Intron
Ba52	10	101584951	100363591	None	None	None	None
Ba53	11	108198964	116126072	None	mCG54370	PD	Downstream
Ba54	13	45692264	No hit	None	None	None	None
Ba55	14	78816293	82243983	None	None	None	None
Ba56	17	37466479	38760848	ENSMUSG00000013094	None	Olfactory receptor	Upstream
				None	mCG1034068	PD	Intron
Ba57	18	40781043	38938508	None	None	None	None
Ba58	3	No hit	72496721	None	None	None	None
Ba59	3	No hit	132473648	None	None	None	None
Ba60	3	No hit	139912164	None	None	None	None
Ba61	3	No hit	144726649	None	None	None	None
Ba62	3	No hit	146815889	None	mCG1045783	PD	Upstream
Ba63	3	No hit	150622778	None	mCG64937	PD	Intron
Bb1	1	65738552	63658310	ENSMUSG00000026888	mCG22301	Grb14	Intron
Bb2	3	78123913	74530121	None	None	None	None
Bb3	3	128921395	131946395	None	None	None	None
Bb4	3	134603031	137542408	None	mCG57759	None	Intron
Bb5	3	136643002	139591321	None	None	None	None
Bb6	3	140069006	142985245	None	None	None	None
Bb7	3	140442067	143357645	None	None	None	None
Bb8	3	140463385	143379060	None	None	None	None
Bb9	3	140721318	143627508	None	mCG53502	PD	Intron
Bb10	3	140734580	143641510	None	mCG53502	PD	Downstream
Bb11	3	140959824	143863788	None	None	None	None
Bb12*[2]	3	141123243	144033139	None	mCG1045798	PD	Intron
Bb13*[2]	3	141140455	144050261	None	mCG1045798	PD	Upstream
Bb14*[2]	3	141140455	144050261	None	mCG1045798	PD	Upstream
Bb15	3	141153125	144063917	None	None	None	None
Bb16	3	141153125	144063917	None	None	None	None
Bb17	3	141161352	144072145	None	None	None	None
Bb18*[3]	3	141216021	144124766	None	mCG1045797	PD	Intron
Bb19*[4]	3	141370931	144282066	None	mCG1045552	PD	Intron
Bb20	3	141533210	144447454	None	None	None	None
Bb21	3	141550152	144468179	None	None	None	None

Continued on following page

TABLE 2—*Continued*

Locus[b]	Chromosome	Chromosome position[c]		Gene hit[d]			
		Ensembl	Celera	Ensembl ID	Celera ID	Product name	Insertion site
Bb22	3	141791266	144708994	None	None	None	None
Bb23	3	159306381	162143341	None	None	None	None
Bb24	9	52064672	46001369	None	None	None	None
Bb25	10	52847168	50527975	None	mCG1028869	None	Upstream
Bb26	11	111938647	119846977	None	None	None	None
Bb27	13	112145008	115874575	None	None	None	None
Bb28	3	No hit	56662465	None	None	None	None
Bb29	3	No hit	142724083	None	mCG1045576	PD	Intron

[a] Sequences flanking transposon insertion sites were searched with the mouse genome databases at both Ensembl and Celera Genomics to determine the chromosome positions of the insertion sites and to identify any insertion events nearby or in the genes. Transposition sites that were aligned with the array of transposon sequences located at the DST or that could not be mapped owing to the repetitive nature of the sequences are not presented. In cases where insertion sites are mapped to two genes, both genes are presented. Eleven insertion sites were determined from testis DNA, and remaining 128 sites were from sperm DNA.

[b] The first two letters of the locus name correspond to the name of the parental seed mouse (shown in Fig. 1B and 4D) from which each insertion site was determined. The names of the loci that were mapped to the same gene are indicated by *1 to *4, each representing insertion into the same gene.

[c] No hit indicates that a transposition site could not be determined.

[d] ID, identification number; PD, gene predicted according to the Celera Genomics database. Based on Celera database parameters, 10 kb upstream and downstream of the transcription units is associated with the same gene identification number. Therefore, insertion sites occuring in these regions are defined as upstream and downstream insertion sites, respectively.

and their thymuses were stained with X-Gal as described above at 16 h after administration.

RESULTS

Breeding scheme to generate mutant mice. A previously described breeding scheme (16) used to demonstrate the transposition of the *SB* transposon is outlined in Fig. 1A. The *SB* transposon vector contains the GFP expression unit, consisting of the ubiquitously active CAG promoter, the GFP gene, and a polyadenylation signal. Multiple copies of the *SB* transposon vector were introduced at the original integration site and served as a DST. From the founder mice, we selected the ones in which GFP expression was repressed at the DST. In the doubly transgenic mice bearing both the *SB* transposon and the *SB* transposase gene, one or a few copies of the transposons were excised and reintegrated into the genome by SB transposase. During this process, the repressed status of the GFP gene was expected to be removed, followed by activation of the gene at the new locus. Green mice were obtained in the progeny of the doubly transgenic mice, indicating that the *SB* transposon had transposed into the germ line of the doubly transgenic mice. We refer to the doubly transgenic mice as seed mice (Fig. 1A), since they will produce many mutant progeny.

Determination of transposition sites in the germ lines of seed mice. As an initial step aimed at investigating both the distribution and complexity of the *SB* transposon in the mammalian germ line, we determined the integration sites of a large number of transpositions which had occurred in the sperm of the seed mice (Fig. 1B). We analyzed two mouse lines, referred to as line A and line B (Fig. 1B). These same mice were previously characterized (16), and lines A and B contain DSTs on chromosomes 14 and 3, respectively. After digestion of sperm DNA with a 6-base cutter (*Bgl*II or *Xba*I), an oligonucleotide linker was ligated to the cleavage ends and transposition sites were amplified by PCR with a transposon-specific primer and a linker-specific primer (Fig. 1C). Since numerous transposition sites exist in the template DNA, a smear consisting of multiple PCR products would be amplified. We therefore diluted the template DNA prior to PCR in order to determine the conditions under which only a single integration site is amplified per reaction. When the amount of the template was reduced to 20 pg per reaction, a discrete band was amplified (data not shown). Once this condition had been determined, the PCR procedure was scaled up for a large number of identical PCRs (Fig. 1D), and transposition sites were subsequently determined by direct sequencing of the PCR products.

Distribution of transposition sites in the germ lines of seed mice. Out of a total of 215 sequenced transposition sites, 57 (27%) were aligned with the sequence of the transposon vector (data not shown). Since the transposon vector is 10 kb in length and approximately 20 such copies exist at the DST (see reference 16 for line B; data not shown for line A), there must be vector sequences spanning as much as 200 kb at the DST. We therefore assumed that 27% of the transposition events had occurred within 200 kb of the DST. In order to confirm this assumption, we mapped the transposition sites of another transgenic line (10) by database searching. This transgenic line contains a single copy of the *SB* transposon at the DST. Although the chromosomal positions of the transposition sites were determined previously with the aid of a radiation hybrid panel (10), the mouse genome database at Ensembl and Celera Genomics makes it possible to determine the distance of the sites from the DST at the nucleotide level, and it was found that 3 out of 12 transposition sites (25%) were mapped within 200 kb from the DST (Fig. 2A; Table 1). This result confirms the observation made for lines A and B. The remaining 158 transposition sites were analyzed by using the mouse genome database. Although some sites could not be mapped due to the presence of repetitive sequences and sequence gaps in the database, exact locations could be determined for 128 sites: 120 sites were determined with Ensembl, and 122 sites were determined with the Celera Genomics database (Table 2; Fig. 2B and C). Three-quarters of the transposition sites were mapped to the chromosome bearing the DST in both mouse lines (Fig. 2B and C), and preferential transposition near the DST was apparent. Most of the transpositions near the DST

A

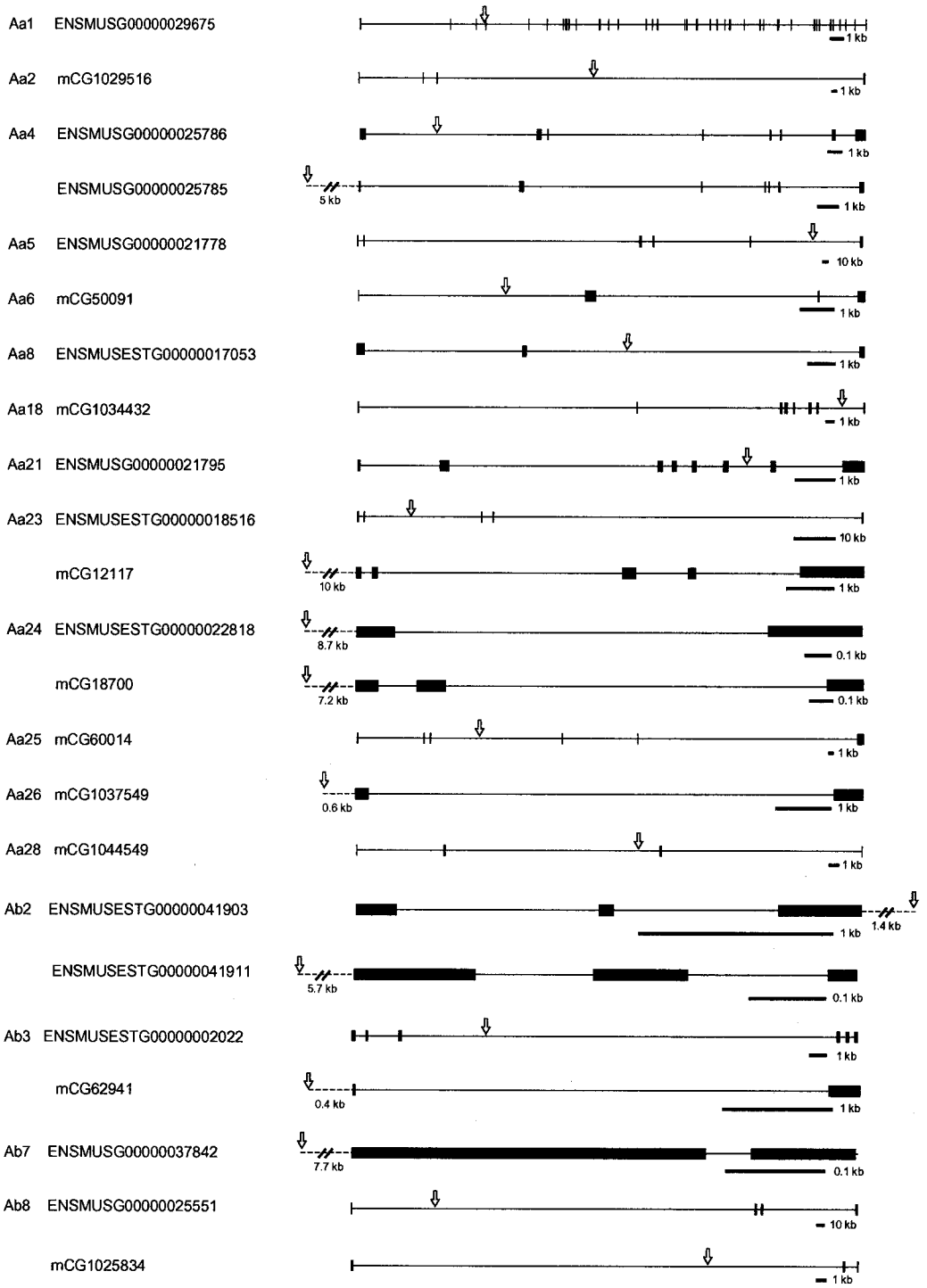


FIG. 3. Distribution of transposon insertion sites at known or predicted genes. Transposon insertion sites that were mapped between 10 kb upstream and 10 kb downstream of the transcription units in Tables 1 and 2 are shown. They are classified into two patterns: single insertion per gene (A) and multiple insertions in a single gene (B). When a gene is registered in both the Ensembl and Celera databases (Tables 1 and 2), the structure of the gene and the corresponding accession number are shown according to the Ensembl database. Boxes, exons; arrows, transposon insertion sites. The orientation of genes is from left to right. Scale bars for each of the genes are shown.

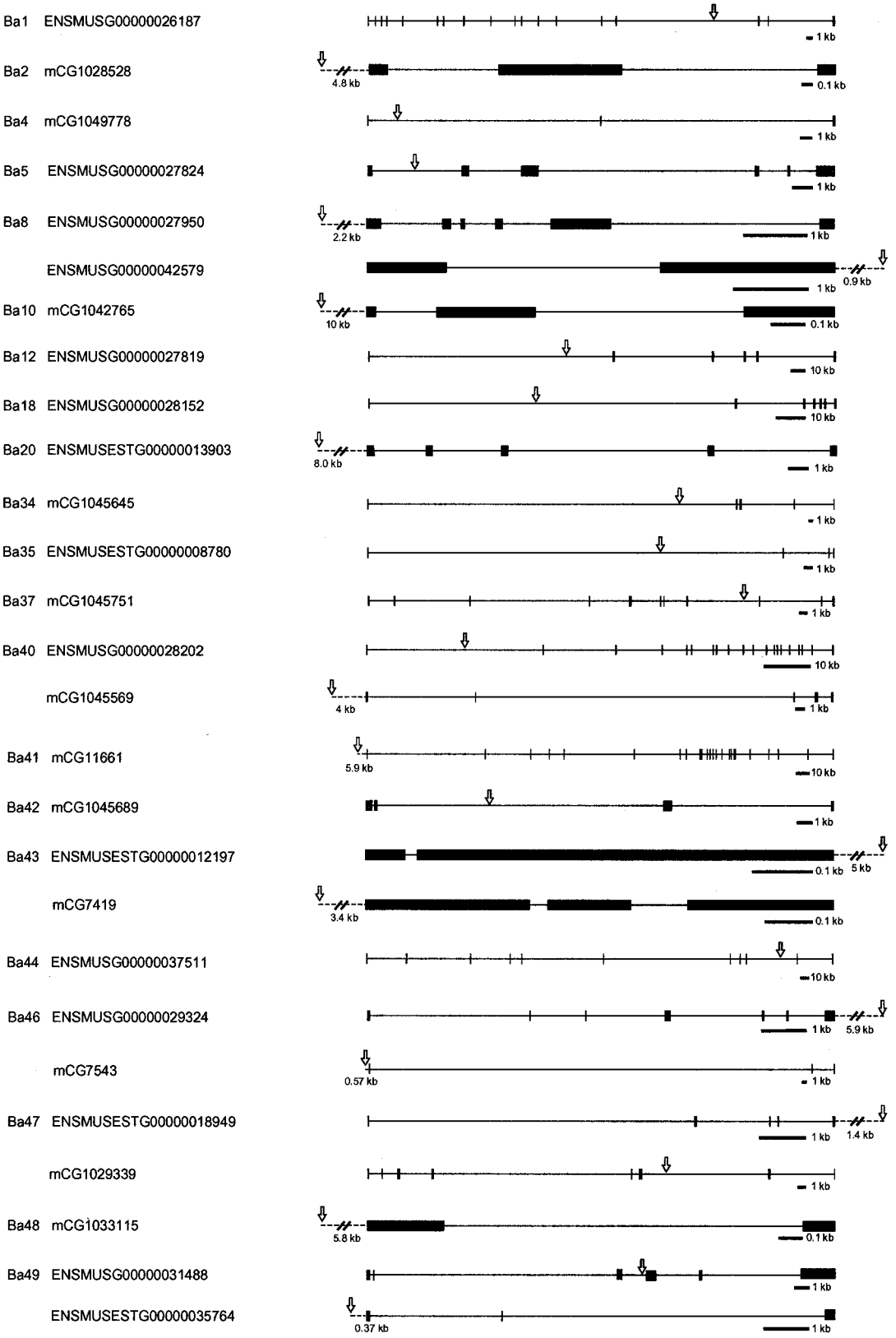


FIG. 3—*Continued*.

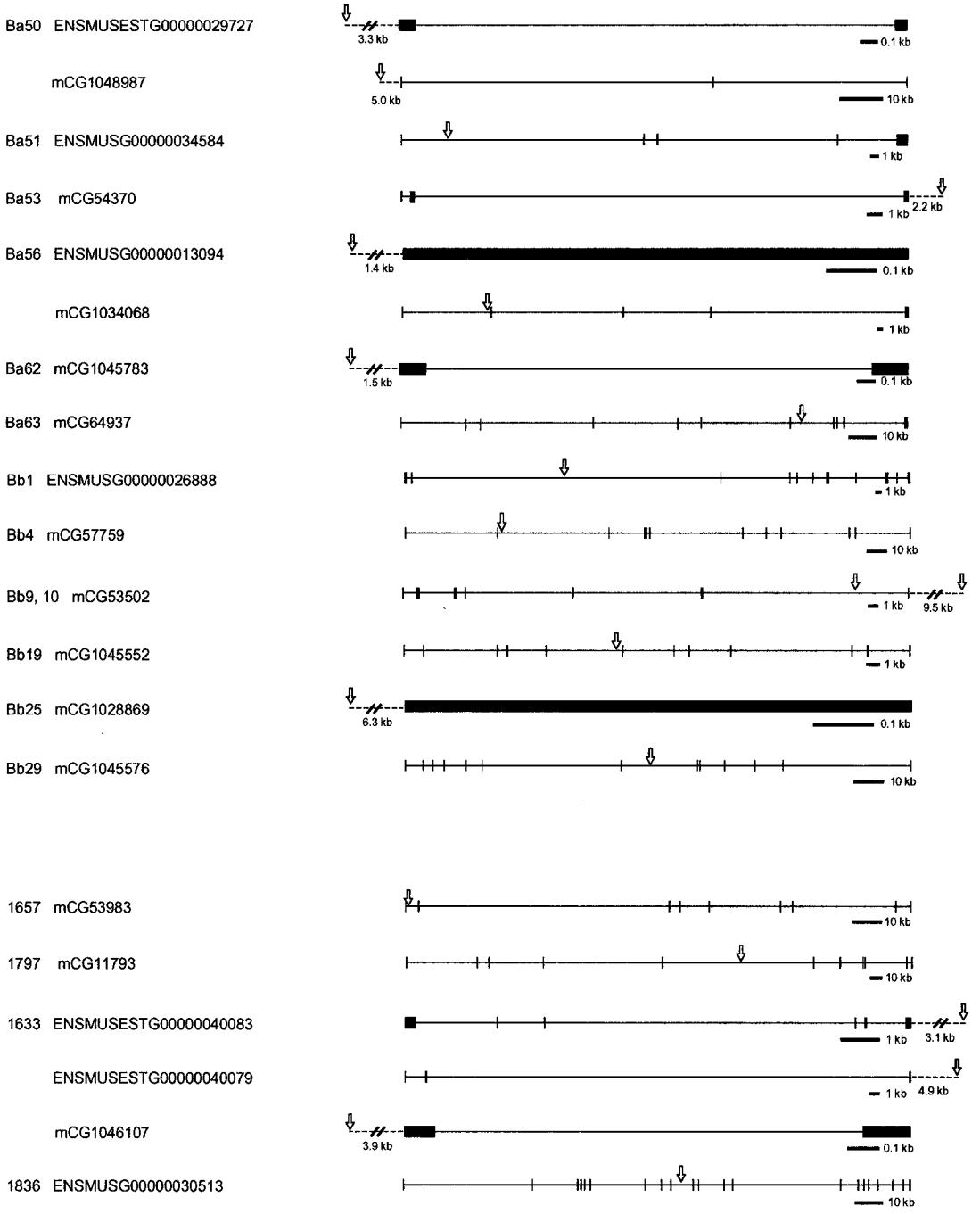


FIG. 3—*Continued*.

were clustered within the 3-Mb region (Fig. 2B and C). Interestingly, transpositions outside this cluster were widely distributed, demonstrating the potential for genome-wide mutation (Fig. 2B and C). Indeed, 16% of all transposition sites analyzed by Ensembl were mapped within transcription units (19 out of 120 sites) (Table 2) predicted on the basis of information about known genes and/or experimental data such as expressed sequence tags (ESTs). The ratio reached as much as 39% if the predictions by Celera Genomics were included (50 out of 128 sites) (Table 2). Furthermore, 24% of them were located on the chromosomes without the DSTs (12 out of 50 sites) (Table 2). This is consistent with the result that 25% of the transpositions that exclude insertions into the transposon sequence were mapped on the chromosomes without the DSTs in both mouse lines (Fig. 2B and C). In total, transposition sites were mapped to 16 different chromosomes when the results of both mouse lines were combined (Table 2). These findings demonstrate that a large number of genes at various chromosomal locations can be mutated by using the *SB* transposon despite the preference for local transposition. Transposition sites were

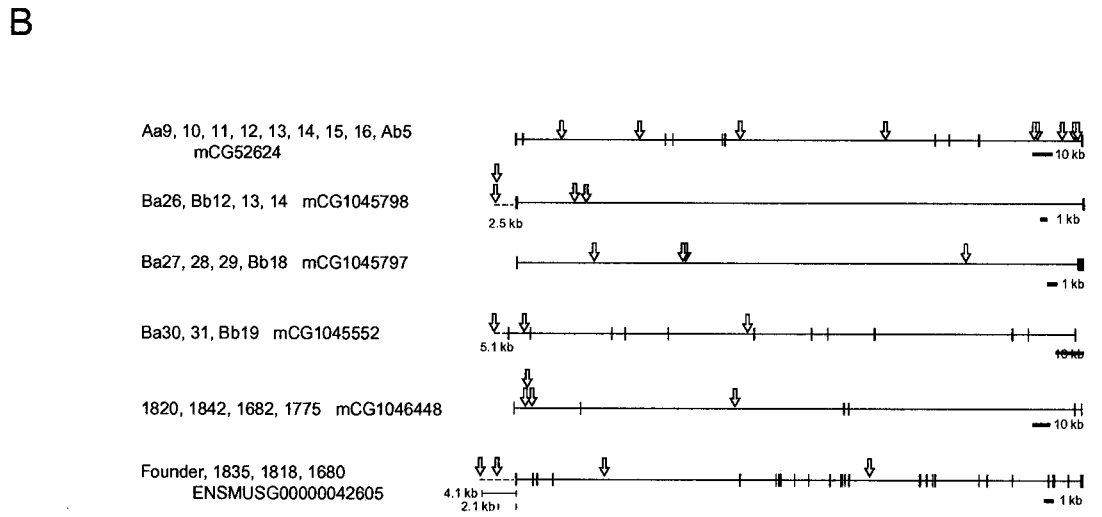


FIG. 3—*Continued*.

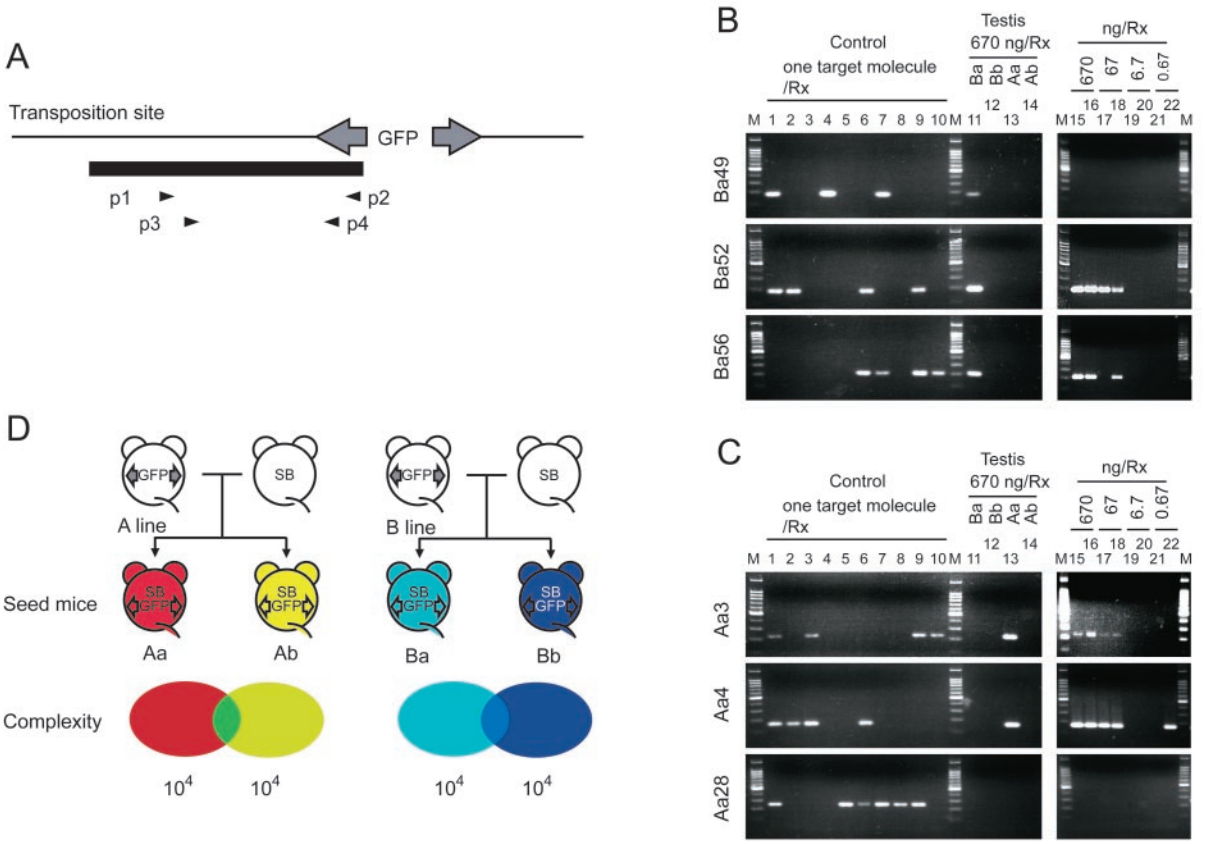


FIG. 4. Complexity of transposition sites within the germ lines of seed mice. (A) Nested PCR primers for the detection of an individual transposition. Primers p1 and p2 are for the first reaction, and primers p3 and p4 are for the second reaction. The region represented by a black bar was isolated by PCR as shown in Fig. 1C, and this fragment was used as a template in lanes 1 to 10 of panels B and C. (B and C) Estimation of the complexity of transposition sites in line B (B) and line A (C). The transposition sites being studied are indicated on the left of each panel. Lanes 1 to 10 show the PCR sensitivity. Lanes 11 to 14 show that each transposition site was detected only in the parental mouse. Lanes 15 to 22 show the amount of testicular DNA in which the target molecule exists. The marker (lanes M) is a 100-bp DNA ladder. Rx, reaction. (D) Schematic diagram of complexity of transposition sites in seed mice. Overlapping circles depicting complexity indicate that some transposition sites are clustered close to the donor site in the same mouse strain.

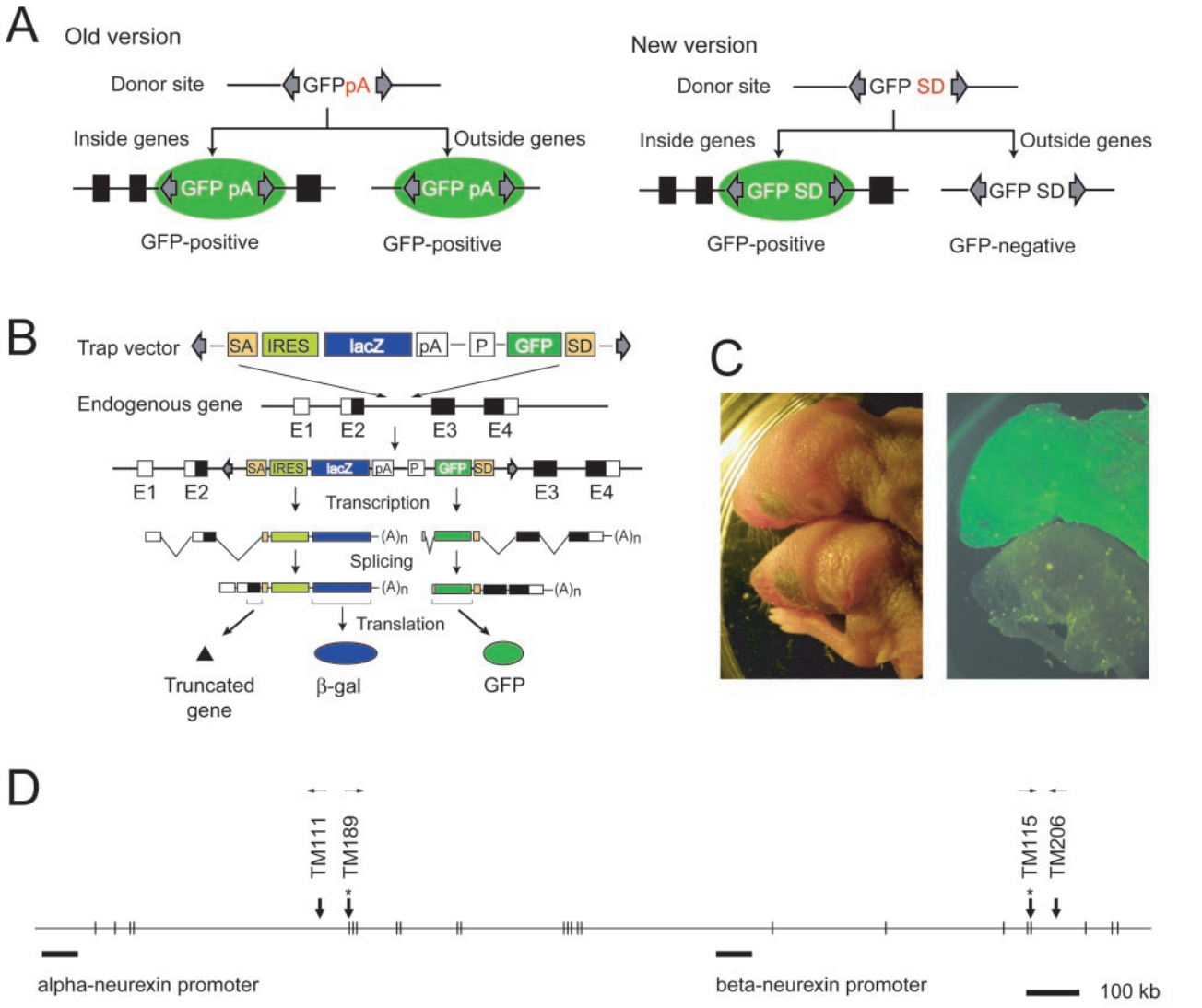


FIG. 5. Screening for mutant mice generated by the gene trap strategy. (A) Strategy to identify transposition events by using the previously reported old version of the *SB* vector (16) (left) (also see Fig. 1A) and the new *SB* trap vector used in the present study (right). pA, poly(A) addition signal; SD, splice donor; filled boxes, exons. (B) Outline of the gene trap scheme. In this example, the transposon vector is inserted into intron 2 of an endogenous gene. Transcription of the endogenous gene results in a chimeric transcript of the endogenous transcript and vector-derived sequences. As a result, translation from the endogenous gene is disrupted and β-galactosidase is expressed, reflecting the expression pattern of the endogenous gene. Transcription by the ubiquitously active CAG promoter generates a chimeric transcript of the GFP sequence and the endogenous transcript, resulting in ubiquitous expression of GFP. White boxes, untranslated regions of exons; black boxes, translated regions of exons. (C) Screening for mutant mice performed by GFP expression. Newborn mice were examined by fluorescence stereomicroscopy. The left panel is a bright-field image, and the right panel is a dark-field fluorescent image taken with GFP filters. The mouse at the top is GFP positive (and therefore presumably a mutant), and the one at the bottom is GFP negative. (D) Multiple insertions into the neurexin 3 gene by local transposition. The neurexin 3 gene contains two promoters, one transcribing alpha-neurexin 3 and the other transcribing beta-neurexin 3, and is located in the vicinity of the DST. Thick arrows indicate the locations of trapped sites, and those with asterisks indicate that the vector-derived SD site was spliced to exons. The orientation of the transposon insertion is shown by thin arrows.

distributed throughout various genes without an apparent preference with respect to the gene structure (Table 2; Fig. 3). It should be noted that some integration sites were mapped to the same gene located near the DST (Table 2; Fig. 3B). This result suggests that the preferential transposition near the DST can be used to introduce multiple mutations in a specific region of the genome.

Complexity of transposition sites within the germ lines of seed mice. The overall complexity of transposition sites within the germ lines of seed mice (i.e., total number of different transposition sites in the germ line) was estimated by determining the relative frequencies of the individual transposition sites described above. Since the overall complexity of transposition sites inversely correlates with the relative frequency of individual transposition sites, the complexity of transposition sites would be represented by the reciprocal of the frequency of individual transposition sites per germ cell. Each germ cell contains approximately one

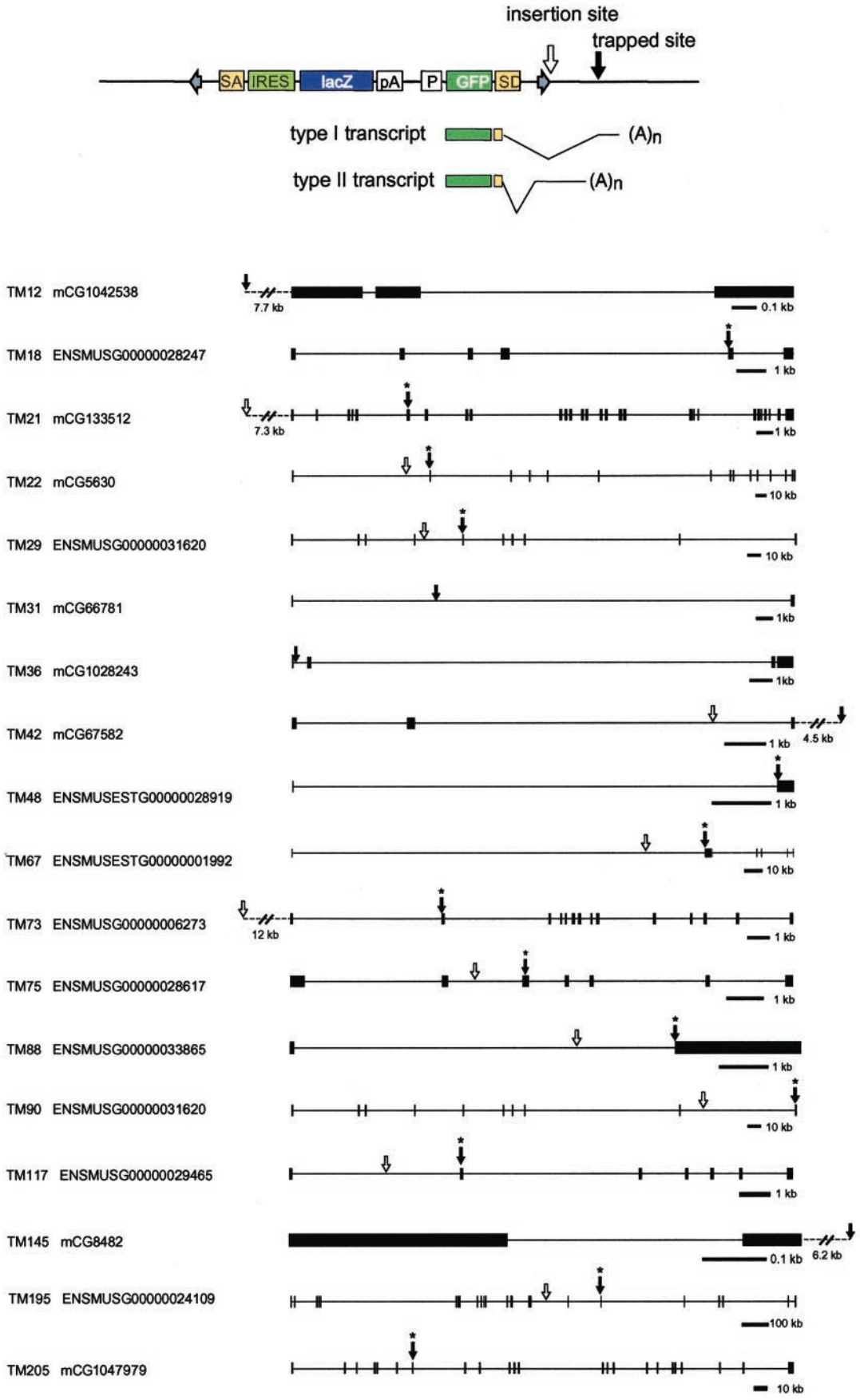
TABLE 3. Homology search of trapped sequences[a]

Mouse ID	Founder	Ensembl ID	Celera ID	Gene product name	Location of trapped sequences	Orientation of trapped sequence relative to trapped gene	Chromosome
TM1	T1	ENSMUSG00000027568	mCG6723	Neurotensin receptor	Downstream	Reverse	2
TM3	T1	None	mCG21578	PD	Intron	Reverse	3
TM4	T2	ENSMUSG00000027835	mCG113365	Programmed cell death 10	Intron	Reverse	3
TM6	T2	ENSMUSG00000031632	mCG1048909	Testican3	Intron	Reverse	8
TM11	T1	None	mCG1048007	PD	Downstream	Reverse	12
TM12	T1	None	mCG1042538	PD	Upstream	Forward	14
TM17	T1	None	mCG1042476	PD	Intron	Reverse	14
TM18	T1	ENSMUSG00000028247	mCG2931	Hexaprenyldihydroxybenzoate methyltransferase	Exon	Forward	4
TM21	T3	None	mCG133512	Integrin alpha D	Exon	Forward	7
TM22	T2	None	mCG5630	Acetylglucosaminyltransferase like	Exon	Forward	8
TM29	T2	ENSMUSG00000031620	mCG13289	RIKEN cDNA 1700007B14	Exon	Forward	8
TM31	T1	None	mCG66781	PD	Intron	Forward	15
TM36	T2	ENSMUSESTG00000023444	None	RIKEN cDNA A730054M09	Intron	Reverse	7
		None	mCG1028243	PD	Intron	Forward	7
TM42	T3	None	mCG67582	PD	Downstream	Forward	9
TM44	T2	None	mCG1048972	PD	Intron	Reverse	8
TM48	T2	ENSMUSESTG00000028919	mCG11950	Aldosterone receptor	Exon	Forward	8
TM54	T2	ENSMUSESTG00000025653	None	RIKEN cDNA C730040P17	Intron	Reverse	16
TM67	T1	ENSMUSESTG00000001992	mCG56153	Teashirt2	Exon	Forward	2
TM73	T2	ENSMUSG00000006273	mCG16916	Vascuolar ATP synthase subunit B, brain isoform	Exon	Forward	8
TM75	T3	ENSMUSG00000028617	mCG16086	RIKEN cDNA A930011F22	Exon	Forward	4
TM83	T1	None	mCG58496	PD	Intron	Reverse	12
TM88	T1	ENSMUSG00000033865	mCG122280	T-cell death-associated gene 8	Exon	Forward	12
TM90	T2	ENSMUSG00000031620	mCG13292	RIKEN cDNA 1700007B14	Exon	Forward	8
TM111	T4	None	mCG8477	Neurexin 3	Intron	Reverse	12
TM115	T4	ENSMUSG00000033935	mCG6347	Neurexin 3	Exon	Forward	12
TM117	T5	ENSMUSG00000029465	mCG12172	Actin-related protein 2/3 complex subunit3	Exon	Forward	5
TM129	T4	ENSMUSG00000044326	mCG128599	Vomeronasal 1 receptor A1	Downstream	Reverse	6
TM145	T4	None	mCG8482	Glyceraldehyde 3-phosphate dedydrogenase	Downstream	Forward	12
TM180	T4	ENSMUSG00000051350	mCG50210	60S ribosomal protein L31	Downstream	Reverse	12
TM189	T4	ENSMUSESTG00000010144	mCG8477	Neurexin 3	Exon	Forward	12
TM195	T4	ENSMUSG00000024109	mCG15583	Neurexin 1	Exon	Forward	12
TM199	T4	None	mCG124030	Seven transmembrane helix receptor	Intron	Reverse	15
TM205	T4	None	mCG1047979	40S ribosomal protein S6	Exon	Forward	12
TM206	T4	None	mCG6347	Neurexin 3	Intron	Reverse	12

[a] From the database analysis of 81 sequences trapped by the poly(A) trap scheme shown in Fig. 5B, the results for 34 mice with sequences mapped near or within genes are presented. Five different founder mice (T1 to T5) bearing the trap vector were used. The effect of the DST on phenotypes was tested in lines T1 and T4. Mice bearing the DST were identified by FISH analysis, and no overt phenotypes were observed, indicating that these lines are suitable for phenotypic analysis of mutations subsequently introduced by transposition. Definitions can be found in the footnotes to Table 2.

copy of a transposition site (16); therefore, the complexity of transposition sites equals 1/frequency of individual transposition sites per germ cell.

Three transposition sites were selected from each line of seed mice, Aa and Ba (Fig. 1B). Transposition sites Ba49, Ba52, and Ba56 originated from the Ba mouse, and Aa3, Aa4, and Aa28 originated from the Aa mouse (Fig. 4A to C). We isolated individual transposition sites (Fig. 4A) according to the protocol shown in Fig. 1C, and primers for the site-specific nested PCR were designed based on the sequences of the individual transposition sites (Fig. 4A). The fragment from an individual transposition site was used as a template for the PCR under the condition that one fragment molecule exists per reaction (Fig. 4B and C, lanes 1 to 10). Six reactions could be expected to be PCR positive according to the Poisson distribution if one target molecule can be amplified, and the overall result was compatible with this prediction. This indicates that the presence or absence of an individual transposition site in the template DNA can be determined by the presence or absence of the PCR-amplified band. The frequency of these sites in the germ lines of the seed mice Aa, Ab, Ba, and Bb (Fig. 1B) was examined by using the DNA from the germ line of each mouse as a template for the PCR (Fig. 4B and C). Only the Ba mouse gave a positive signal when 670 ng of template DNA was used (Fig. 4B, lanes 11 to 14). No positive signal was detected in the progeny from line A or in the Bb mouse, a littermate of the Ba mouse. Therefore, the repertoire of transposition sites did not overlap between line A and line B or even between littermates. When the template DNA was diluted up to 67 ng per reaction and analyzed in duplicate, no


insertion site
trapped site
SA
IRES
lacZ
pA
P
GFP
SD
type I transcript
(A)n
type II transcript
(A)n
TM12 mCG1042538
7.7 kb
0.1 kb
TM18 ENSMUSG00000028247
1 kb
TM21 mCG133512
7.3 kb
1 kb
TM22 mCG5630
10 kb
TM29 ENSMUSG00000031620
10 kb
TM31 mCG66781
1kb
TM36 mCG1028243
1kb
TM42 mCG67582
1 kb
4.5 kb
TM48 ENSMUSESTG00000028919
1 kb
TM67 ENSMUSESTG00000001992
10 kb
TM73 ENSMUSG00000006273
12 kb
1 kb
TM75 ENSMUSG00000028617
1 kb
TM88 ENSMUSG00000033865
1 kb
TM90 ENSMUSG00000031620
10 kb
TM117 ENSMUSG00000029465
1 kb
TM145 mCG8482
0.1 kb
6.2 kb
TM195 ENSMUSG00000024109
100 kb
TM205 mCG1047979
10 kb

signal was obtained in Ba49, both reactions were positive in Ba52, and one reaction was positive in Ba56 (Fig. 4B, lanes 15 to 22). Since 67 ng of genomic DNA corresponds to approximately 10,000 cells, the frequency of individual transposition is less than 1/10,000 in Ba49, more than 1/10,000 in Ba52, and around 1/10,000 in Ba56. We therefore estimate that the complexity of the transposition in the germ line of the seed mice is approximately 10,000 according to the equation presented above. This means that a seed mouse is capable of generating 10,000 different mutant mice. Similar data were obtained from the analysis of the transposition sites obtained from line A (Fig. 4C). Taking into account the distribution of transposition sites (Fig. 2B and C), we have summarized the complexity of transposition sites in Fig. 4D. Each seed mouse would be predicted to generate 10,000 different mutant mice in which transposition had occurred independently. Different mutant mice would be obtained from different seed mice bearing the same DST, although a portion of the transposition sites would be mapped in close proximity to one another, since the *SB* transposon demonstrates preferential transposition in close proximity to the DST (Fig. 2B and C). It would be easy to increase the complexity by generating different lines of seed mice bearing DSTs at different locations (Fig. 4D), because the distribution of the transposition sites is completely different as shown in Fig. 2B and C.

Screening and analysis of mutant mice generated with the gene trap strategy. The high degree of complexity of transposition in the seed mice led us to believe that the *SB* transposon system could be employed as a tool for the large-scale generation of mutant mice. For this purpose, we constructed a new version of the transposon vector which was designed to utilize the poly(A) trap method (17, 35) (Fig. 5A and B). Our original version of the vector was designed to detect transposition events irrespective of their location or incorporation into endogenous genes (16) (Fig. 5A, left; also see Fig. 1A). In contrast, the new vector was designed to utilize the poly(A) trap method (17, 35) in order to select transposition events occurring in endogenous genes (Fig. 5A [right] and B). This vector contains the same vector backbone used in our previous study (16) in order to minimize GFP expression prior to transposition (see Materials and Methods). Approximately 7% of the newborns from seed mice were GFP positive (Fig. 5C), suggesting that the *SB* transposon had inserted into endogenous genes. Although a majority of newborns were GFP negative, use of the noninvasive GFP reporter enabled us to focus on potential mutant mice soon after birth (Fig. 5C) and to avoid working with the vast majority of mice that are not likely to carry gene mutations. Genomic sequences trapped by the poly(A) trap scheme were determined by 3′ RACE. Eighty-one sequences were analyzed with the mouse genome database, resulting in the mapping of 34 sequences at sites of known or predicted genes (Table 3), and splicing to endogenous exons was observed in 15 of these sequences, (Table 3, Fig. 6), thus validating the principle of the poly(A) trap scheme. Although many of the trapped sequences were not mapped to exons, it cannot be ruled out that many of them might be located at the sites of unknown genes. In fact, it has been demonstrated that the poly(A) trap scheme is useful for the identification of novel genes (14, 35).

It is notable that four individual insertions were mapped within the neurexin 3 gene (Fig. 5D). The neurexin 3 gene encodes a longer alpha-neurexin 3 and a shorter beta-neurexin 3 because of distinct promoters, as shown in Fig. 5D (33). Interestingly, two insertions were in the region specific to the alpha-neurexin 3, and the remaining insertions were in the region that is common to both alpha- and beta-neurexin 3 (Fig. 5D). The neurexin 3 gene is reported to be at chromosome 12D3 (33), and the DST was mapped at chromosome12D3-E by fluorescent in situ hybridization (FISH) analysis (data not shown), indicating that multiple insertions in the neurexin 3 gene are the result of preferential local transposition of the *SB* transposon (Fig. 2). This result demonstrates that the *SB* transposon is a valuable tool for region-specific mutagenesis.

Another advantage of our new *SB* vector is that it contains elements for a promoter trap (11, 12, 31), thereby facilitating the visualization of endogenous locus-specific expression patterns by *lacZ* gene activity (Fig. 5B). Various patterns of *lacZ* gene expression were observed. Ubiquitous expression was observed in 11.5-day-postcoitum (dpc) embryos of the TM75 line (Fig. 7A), and testis-specific expression was detected in TM90 adult mice (Fig. 7B). TM67 mice (Fig. 7C) contain a transposon insertion in the teashirt2 gene (4) (*mtsh2*), a mouse ortholog of the *Drosophila* teashirt gene that is required for specification of trunk segments during embryogenesis (9). In this line, the *lacZ* gene was expressed in specific regions of the 11.5-dpc embryo, such as somites and limb buds (Fig. 7C). This is similar to the reported expression pattern of the *mtsh2* gene detected by in situ hybridization (4). TM88 mice (Fig. 7D) contain a transposon insertion in the T-cell death-associated gene 8 (TDAG8) (5). TDAG8 is a putative G protein-coupled

FIG. 6. Distribution of trapped sites at known or predicted genes. From the trapped sites that were mapped at the genes shown in Table 3, 18 sites with transposon insertions in the same orientation relative to the trapped genes are shown. Insertions into neurexin 3 genes (TM115 and TM189) are shown in Fig. 5D and therefore are not shown here. Splicing patterns revealed by 3′ RACE are shown at the top. In addition to the predicted splicing between the vector-derived SD site and genomic sequences (type I transcript), we occasionally observed unexpected splicing between the SD site and a cryptic SA site within the trap vector (type II transcript). Since the type II transcript contains the junction of the transposon sequence and the genomic sequence, we sequenced it to determine the transposon insertion site (for TM29, TM42, TM67, and TM73). When a type II transcript was not observed, transposon insertion sites were determined either by ligation-mediated PCR (7, 10) (for TM21, TM22, TM75, TM90, TM117, and TM195) or by PCR between transposon-specific primers (T/BAL and T/DR) (18) and reverse primers that were designed at or upstream of the trapped site (TM88). TM numbers correspond to the mouse identification numbers in Table 3. When a gene is registered in both the Ensembl and Celera databases (Table 3), the structure of the gene and the corresponding accession number are presented according to the Ensembl database. Black arrows indicate the locations of trapped sites, and asterisks indicate that the vector-derived SD site was spliced to known or predicted exons. White arrows indicate transposon insertion sites. The orientation of genes is from left to right. Scale bars for each of the genes are shown. For TM67, see the legend for Fig. 7E.

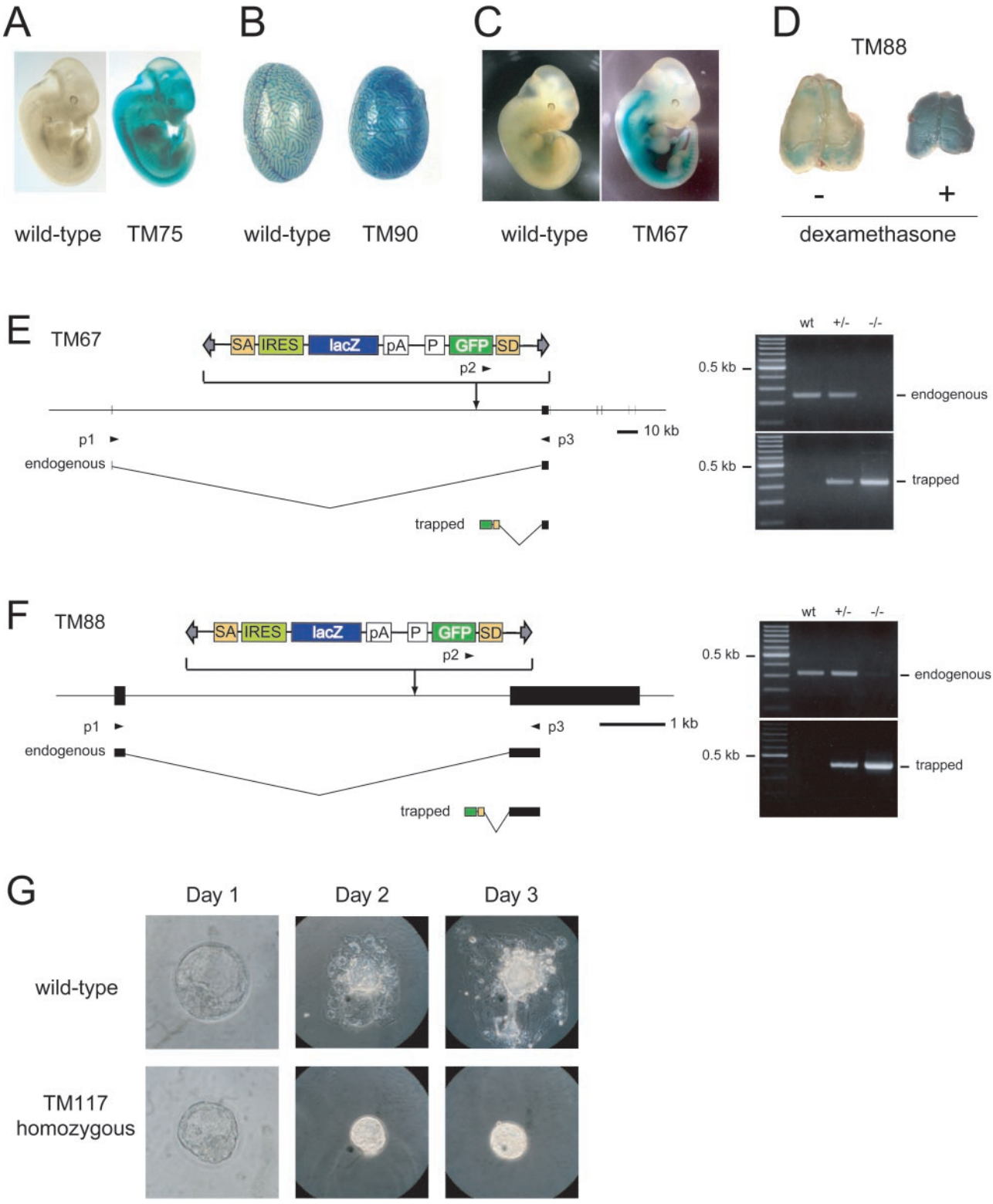


FIG. 7. Generation and analysis of mutant mice. (A to C) X-Gal staining of an 11.5-dpc embryo from TM75 (A), an adult testis from TM90 (B), and an 11.5-dpc embryo from TM67 (C). (D) *lacZ* gene induction by dexamethasone in the thymuses of TM88 mice. Heterozygous mice (10 days of age, 6.5 g) were injected intraperitoneally with 200 μg of dexamethasone, and the thymus was stained with X-Gal 16 h after injection. (E and F) Disruption of gene expression in homozygous TM67 mutant mice (E) and TM88 mutant mice (F). p1 and p3 were used to detect the wild-type (wt) transcripts of the mutated genes, and p2 and p3 were used to detect splicing events occurring between the transposon vector and a downstream exon. We found an uncharacterized EST (National Center for Biotechnology Information accession no. 11506469) derived from the upstream region of the *mtsh2* gene and incorporated it in the gene structure presented in panel E. (G) TM117 heterozygous mice were established after segregating the DST during breeding and were intercrossed. Blastocysts were isolated and observed for 3 days, followed by genotyping with PCR. Hatching was impaired in homozygous embryos.

TABLE 4. Genotypes of progeny from intercrosses of heterozygous mice with transposon insertions[a]

Mouse	No. of progeny		
	Wild type	Heterozygous	Homozygous
TM21	9	13	8
TM67	12	18	9
TM75	20	24	10
TM88	15	21	5
TM115	3	12	9
TM117	8	15	0
TM150	7	12	5
TM173	2	10	2
TM195	2	5	4
TM222	2	4	2

[a] All mouse lines are presented in Table 3 except for TM150, TM173, and TM222. Their insertion sites are as follows: TM150, intron 1 of ENSMUS ESTT00000002853 (Ensembl identification number); TM173, intron 3 of two-pore-domain potassium channel gene; TM222, intron 2 of armadillo repeat gene deleted in velo-cardio-facial syndrome.

receptor that is highly expressed in T cells undergoing apoptosis. Indeed, *lacZ* gene induction was observed in thymocytes of TM88 mice upon intraperitoneal injection of dexamethasone, a strong inducer of apoptosis in thymocytes (Fig. 7D). These results demonstrate that the expression patterns of the mutated genes can be examined by *lacZ* reporter gene activity.

Homozygous mice were generated from 10 different transposon-gene insertions in order to test the mutagenicity of the transposon vector (Table 4). Homozygous TM67 mice were viable and did not show any overt abnormality, but they were smaller than their heterozygous littermates (at 24 days old, the mean weight ± standard deviation was 8.8 ± 1.4 g for homozygous mice versus 12.0 ± 0.5 g for heterozygous mice). Analysis of the expression of the *mtsh2* gene by reverse transcription-PCR showed no normal transcripts in homozygous mice (Fig. 7E). We analyzed the TM75 and TM88 homozygous mice as well and found that normal transcripts were almost entirely eliminated (Fig. 7F for TM88 mice; data not shown for TM75 mice). In TM117 mice, no homozygotes were obtained at birth or from 7.5 dpc. We therefore analyzed the growth and differentiation of blastocysts by using in vitro culture (Fig. 7G). In the wild type and heterozygotes, blastocysts hatched and attached to the dish, and growth of the inner cell mass on the trophectodemal layer was observed. In contrast, homozygous blastocysts did not hatch, although the inner cell mass continued to proliferate. TM117 mice have a transposon insertion in the actin-related protein 2/3 complex subunit 3 (arpc3) gene, which is known to regulate actin polymerization (28). The results indicate a role of actin polymerization during the hatching process. These results demonstrate the mutagenicity of the transposon vector and indicate that transposon-tagged mutagenesis is an efficient system for generating mutant mice.

DISCUSSION

In the postgenomic era, novel methods that allow for efficient analysis of a large number of uncharacterized genes need to be developed. We believe that our mutagenesis scheme meets this need in the following respects. First, our method is not labor intensive, nor does it require extensive tissue culture or manipulation of embryos, in contrast to the ES cell-based technology. Second, the mutated genes can be rapidly identified by using transposon sequences as tags. Third, the high complexity of transposition sites in the germ cells of seed mice indicates that a large number of mutant mice can be generated. Fourth, use of the GFP reporter allows for rapid and noninvasive screening for the identification of mutant mice. Since a male seed mouse can generate a large number of progeny by successive breeding with wild-type female mice, screening for mice carrying transposon insertions in the intragenic regions becomes an impediment for large-scale mutagenesis. Use of the GFP reporter gene in the poly(A) trap scheme has overcome this problem.

Distribution analysis of transposition sites is helpful in determining the maximum potential of the transposon system. The transposon shows a preference to jump locally, and most of the local transpositions were clustered within the 3-Mb region near the DST (Fig. 2B and C). This distribution range would be suitable for the extensive introduction of mutations into a particular locus of interest, such as a gene cluster or a chromosomal location in which the presence of a tumor suppressor gene is implicated on the basis of cytogenetic analysis of human cancer cells. This approach will also complement the region-specific ENU mutagenesis programs in which one of the homologous chromosomes is segmentally deleted and the other is mutagenized by ENU (19). Use of the *SB* transposon in place of ENU will help to introduce a variety of mutations into the specific chromosomal region under study. The principle of region-specific mutagenesis was demonstrated by the generation of four neurexin 3 mutant mice (Fig. 5D), each with a different insertion site in the neurexin 3 gene. There are three neurexin genes in vertebrates (the neurexin 1, 2, and 3 genes). Each encodes alpha- and beta-neurexin from distinct promoters, and more than 1,000 forms of neurexins are generated by alternative splicing (33). Alpha-neurexin knockout mice were described recently, and the role of alpha-neurexins in calcium-triggered neurotransmitter release was demonstrated (23). Since we have two independent insertions, at the region specific the alpha-neurexin 3 and at the region common to both types of neurexins, analysis of these mice may help to distinguish the functions of alpha- and beta-neurexin 3 as well as to elucidate the functions of different domains of neurexin 3. Multiple mutations in a single gene were also shown from analyses of different mouse lines (Fig. 3B), further demonstrating the feasibility of region-specific mutagenesis by the *SB* transposon. Since local transposition sites will be often linked to the DST, the mutant mice that are homozygous for a transposition site will contain the DST at both alleles. Therefore, we need to use a DST that does not affect phenotypes for homozygosity. We initially used FISH to identify such DSTs (see Table 3, footnote *a*), but we later devised an easy screening protocol for homozygosity of a DST by using real-time PCR. At present, this protocol is routinely used to identify appropriate DSTs before we introduce the SB transposase by breeding. Transposition sites outside the 3-Mb cluster showed a wide distribution (Fig. 2B and C). In fact, 24% of the mapped transposition sites within the transcription units were distributed on various chromosomes without DSTs (Table 2). This result indicates that a large number of genes can be mutagenized from a particular DST. It also indicates that the mutations introduced by trans-

positions can often be segregated from the DSTs. Genome-wide mutagenesis would be facilitated further by using different DSTs, each positioned on a different chromosome. For this purpose, we are currently establishing many mouse lines that contain appropriate DSTs for phenotypic analysis of mutant mice by using the real-time PCR protocol described above.

Of the trapped sequences that were located at genes by database searching, nearly half were mapped in the reverse orientation (Table 3). Insertions upstream and downstream of genes were also observed (Table 3). There is a possibility that unknown exons were trapped, and some of them may be located in the antisense orientation relative to the known genes. There is increasing recognition that antisense transcripts play important roles in the regulation of gene expression (30), and recent analysis of the mouse transcriptome suggests that the number of antisense transcripts is much higher than that was previously believed (20). In some cases, a cryptic SA site or cryptic polyadenylation signal may have been utilized. In fact, a cryptic SA site within the trap vector was occasionally observed (Fig. 6). In order to improve the efficiency of the gene trap, we are currently testing new version of the trap vector in which the cryptic SA site is eliminated.

We estimate that approximately 10,000 transposition sites exist in the germ cells of seed mice. Interestingly, this number is close to the number of stem cells per testis, which was reported to be about 20,000 to 35,000 (22, 34). This implies that transposition may have occurred in the stem cell stage. Also worth noting is that transposition efficiencies in mouse germ cells were several orders of magnitude higher than that in ES cells, which has previously been reported as 3.5×10^{-5} events/cell per generation (21). Germ line stem cells may therefore possess a mechanism or cellular factor that enhances transposition efficiency.

In contrast to ES cell-based mutagenesis, the transposon system can be used in any mouse genetic background. Most of the ES cell-derived gene knockout mice have the 129 genetic background because of the ease in isolating pluripotent euploid ES cell lines from this strain. This genetic background is inappropriate for some biological analyses, such as immunological and behavioral studies. Since the technology that we described is not restricted to a particular mouse strain, mutant mice can be created in any desired genetic background. The transposon system could be even more useful in model animals in which ES cells are not available. Since the activity of *SB* was demonstrated in cell lines from fish and human (18), it would be functional in many species, including the rat. Zayed et al. reported recently that *SB* transposition is enhanced by the DNA-bending protein HMGB1 (36). Further study of transposon structure may allow improvement of the efficiency of transposition in vivo.

Our findings reported here indicate that the *SB* transposon system can be expected to facilitate the study of gene function in mammalian model organisms and to become an essential tool in functional genomics.

ACKNOWLEDGMENTS

We acknowledge Y. Ishida for providing the RET vector; H. Koike, N. Komazawa, Kuroiwa, K. Yokota, K. Kuratani, and K. Yoshino for technical assistance; and K. Hadjantonakis, M. Kouno, R. Ikeda, and M. Nagai for comments on the manuscript.

This work was supported in part by a grant from New Energy and Industrial Technology Development Organization (NEDO) of Japan and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

REFERENCES

1. **Bellen, H. J., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. K. Pearson, and W. J. Gehring.** 1989. *P*-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev. **3:**1288–1300.
2. **Bradley, A.** 2002. Mining the mouse genome. Nature **420:**512–514.
3. **Brand, A. H., and N. Perrimon.** 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118:**401–415.
4. **Caubit, X., N. Core, A. Boned, S. Kerridge, M. Djabali, and L. Fasano.** 2000. Vertebrate orthologues of the *Drosophila* region-specific patterning gene teashirt. Mech. Dev. **91:**445–448.
5. **Choi, J. W., S. Y. Lee, and Y. Choi.** 1996. Identification of a putative G protein-coupled receptor induced during activation-induced apoptosis of T cells. Cell. Immunol. **168:**78–84.
6. **de Angelis, M. H., H. Flaswinkel, H. Fuchs, B. Rathkolb, D. Soewarto, S. Marschall, S. Heffner, W. Pargent, K. Wuensch, M. Jung, A. Reis, T. Richter, F. Alessandrini, T. Jakob, E. Fuchs, H. Kolb, E. Kremmer, K. Schaeble, B. Rollinski, A. Roscher, C. Peters, T. Meitinger, T. Strom, T. Steckler, F. Holsboer, and T. Klopstock.** 2000. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat. Genet. **25:**444–447.
7. **Devon, R. S., D. J. Porteous, and A. J. Brookes.** 1995. Splinkerettes-improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res. **23:**1644–1645.
8. **Dupuy, A. J., S. Fritz, and D. A. Largaespada.** 2001. Transposition and gene disruption in the male germline of the mouse. Genesis **30:**82–88.
9. **Fasano, L., L. Roder, N. Core, E. Alexandre, C. Vola, B. Jacq, and S. Kerridge.** 1991. The gene teashirt is required for the development of *Drosophila* embryonic trunk segments and encodes a protein with widely spaced zinc finger motifs. Cell **64:**63–79.
10. **Fischer, S. E., E. Wienholds, and R. H. Plasterk.** 2001. Regulated transposition of a fish transposon in the mouse germ line. Proc. Natl. Acad. Sci. USA **98:**6759–6764.
11. **Friedrich, G., and P. Soriano.** 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. **5:**1513–1523.
12. **Gossler, A., A. L. Joyner, J. Rossant, and W. C. Skarnes.** 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science **244:**463–465.
13. **Greenwald, I.** 1985. lin-12, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell **43:**583–590.
14. **Harrington, J. J., B. Sherf, S. Rundlett, P. D. Jackson, R. Perry, S. Cain, C. Leventhal, M. Thornton, R. Ramachandran, J. Whittington, L. Lerner, D. Costanzo, K. McElligott, S. Boozer, R. Mays, E. Smith, N. Veloso, A. Klika, J. Hess, K. Cothren, K. Lo, J. Offenbacher, J. Danzig, and M. Ducar.** 2001. Creation of genome-wide protein expression libraries using random activation of gene expression. Nat. Biotechnol. **19:**440–445.
15. **Hidalgo, A., J. Urban, and A. H. Brand.** 1995. Targeted ablation of glia disrupts axon tract formation in the Drosophila CNS. Development **121:**3703–3712.
16. **Horie, K., A. Kuroiwa, M. Ikawa, M. Okabe, G. Kondoh, Y. Matsuda, and J. Takeda.** 2001. Efficient chromosomal transposition of a *Tc1/mariner*-like transposon *Sleeping Beauty* in mice. Proc. Natl. Acad. Sci. USA **98:**9191–9196.
17. **Ishida, Y., and P. Leder.** 1999. RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. Nucleic Acids Res. **27:**e35.
18. **Ivics, Z., P. B. Hackett, R. H. Plasterk, and Z. Izsvak.** 1997. Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. Cell **91:**501–510.
19. **Justice, M. J., J. K. Noveroske, J. S. Weber, B. Zheng, and A. Bradley.** 1999. Mouse ENU mutagenesis. Hum. Mol. Genet. **8:**1955–1963.
20. **Kiyosawa, H., I. Yamanaka, N. Osato, S. Kondo, Y. Hayashizaki, et al.** 2003. Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res. **13:**1324–1334.
21. **Luo, G., Z. Ivics, Z. Izsvak, and A. Bradley.** 1998. Chromosomal transposition of a *Tc1/mariner*-like element in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA **95:**10769–10773.
22. **Meistrich, M. L., and M. E. A. B. van Beek.** 1993. Spermatogonial stem cells, p. 266–295. *In* C. Desjardins and L. L. Ewing (ed.), Cell and molecular biology of testis. Oxford University Press, New York, N.Y.
23. **Missler, M., W. Zhang, A. Rohlmann, G. Kattenstroth, R. E. Hammer, K. Gottmann, and T. C. Sudhof.** 2003. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature **423:**939–947.
24. **Moerman, D. G., G. M. Benian, and R. H. Waterston.** 1986. Molecular

cloning of the muscle gene unc-22 in Caenorhabditis elegans by Tc1 transposon tagging. Proc. Natl. Acad. Sci. USA **83:**2579–2583.
25. **Mouse Genome Sequencing Consortium.** 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420:**520–562.
26. **Nolan, P. M., J. Peters, M. Strivens, D. Rogers, J. Hagan, N. Spurr, I. C. Gray, L. Vizor, D. Brooker, E. Whitehill, R. Washbourne, T. Hough, S. Greenaway, M. Hewitt, X. Liu, S. McCormack, K. Pickford, R. Selley, C. Wells, Z. Tymowska-Lalanne, P. Roby, P. Glenister, C. Thornton, C. Thaung, J. A. Stevenson, and R. Arkell.** 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat. Genet. **25:**440–443.
27. **Osborne, B. I., and B. Baker.** 1995. Movers and shakers: maize transposons as tools for analyzing other plant genomes. Curr. Opin. Cell Biol. **7:**406–413.
28. **Robinson, R. C., K. Turbedsky, D. A. Kaiser, J. B. Marchand, H. N. Higgs, S. Choe, and T. D. Pollard.** 2001. Crystal structure of Arp2/3 complex. Science **294:**1679–1684.
29. **Rorth, P.** 1996. A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. Proc. Natl. Acad. Sci. USA **93:**12418–12422.
30. **Rougeulle, C., and E. Heard.** 2002. Antisense RNA in imprinting: spreading silence through Air. Trends Genet. **18:**434–437.
31. **Skarnes, W. C., B. A. Auerbach, and A. L. Joyner.** 1992. A gene trap approach in mouse embryonic stem cells: the *lacZ* reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev. **6:**903–918.
32. **Spradling, A. C., D. M. Stern, I. Kiss, J. Roote, T. Laverty, and G. M. Rubin.** 1995. Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. Proc. Natl. Acad. Sci. USA **92:**10824–10830.
33. **Tabuchi, K., and T. C. Sudhof.** 2002. Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. Genomics **79:**849–859.
34. **Tegelenbosch, R. A., and D. G. de Rooij.** 1993. A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. Mutat. Res. **290:**193–200.
35. **Zambrowicz, B. P., G. A. Friedrich, E. C. Buxton, S. L. Lilleberg, C. Person, and A. T. Sands.** 1998. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature **392:**608–611.
36. **Zayed, H., Z. Izsvak, D. Khare, U. Heinemann, and Z. Ivics.** 2003. The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucleic Acids Res. **31:**2313–2322.