Transposon Mutagenesis of the Mouse Germline

Corey M. Carlson,*,† Adam J. Dupuy,*,† Sabine Fritz,*,† Kevin J. Roberg-Perez,† Colin F. Fletcher‡ and David A. Largaespada*,†,[1]

**The Arnold and Mabel Beckman Center for Transposon Research, Institute of Human Genetics, Department of Genetics, Cell Biology, and Development,* †*University of Minnesota Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455 and* ‡*Genomics Institute of the Novartis Research Foundation, San Diego, California 92121*

Manuscript received October 30, 2002
Accepted for publication April 30, 2003

ABSTRACT

Sleeping Beauty is a synthetic "cut-and-paste" transposon of the Tc1/mariner class. The *Sleeping Beauty* transposase (SB) was constructed on the basis of a consensus sequence obtained from an alignment of 12 remnant elements cloned from the genomes of eight different fish species. Transposition of *Sleeping Beauty* elements has been observed in cultured cells, hepatocytes of adult mice, one-cell mouse embryos, and the germline of mice. SB has potential as a random germline insertional mutagen useful for *in vivo* gene trapping in mice. Previous work in our lab has demonstrated transposition in the male germline of mice and transmission of novel inserted transposons in offspring. To determine sequence preferences and mutagenicity of SB-mediated transposition, we cloned and analyzed 44 gene-trap transposon insertion sites from a panel of 30 mice. The distribution and sequence content flanking these cloned insertion sites was compared to 44 mock insertion sites randomly selected from the genome. We find that germline SB transposon insertion sites are AT-rich and the sequence ANNTANNT is favored compared to other TA dinucleotides. Local transposition occurs with insertions closely linked to the donor site roughly one-third of the time. We find that ~27% of the transposon insertions are in transcription units. Finally, we characterize an embryonic lethal mutation caused by endogenous splicing disruption in mice carrying a particular intron-inserted gene-trap transposon.

MEMBERS of the Tc1/mariner family of transposable elements have been found in a diverse group of species (Plasterk *et al.* 1999). Several active Tc1/mariner family transposons have been cloned and studied for potential use as insertional mutagens (Eide and Anderson 1985; Medhora *et al.* 1991; van Luenen *et al.* 1993; Pelicic *et al.* 2000; Hartl 2001). There has been much interest in the Tc1/mariner transposons as insertional mutagens because they have a small target site for integration and seem to require few host cell factors for their activity. The *Sleeping Beauty* (SB) transposase is one member of the Tc1/mariner family that was recently engineered on the basis of a consensus sequence obtained from 12 remnant Tc1/mariner elements from eight different fish species (Ivics *et al.* 1997). The activity of the *Sleeping Beauty* transposase has been demonstrated in cultured mammalian cells (Ivics *et al.* 1997; Izsvak *et al.* 2000), mouse embryonic stem (ES) cells (Luo *et al.* 1998), mouse hepatocytes (Yant *et al.* 2000), the one-cell mouse embryo (Dupuy *et al.* 2002), and the mouse germline (Dupuy *et al.* 2001; Fischer *et al.* 2001; Horie *et al.* 2001).

Transposition of *Sleeping Beauty* transposons in the mouse germline offers the potential for *in vivo* insertional mutagenesis and gene trapping. The *P* element has been mobilized in the fruit fly *Drosophila melanogaster* and played an important role in Drosophila functional genetics as a random insertional mutagen. Methods using gene-trap technology with ES cells have proven useful for generating novel mutants and for assigning functions to mouse genes (Skarnes *et al.* 1992; Townley *et al.* 1997). However, these methods are technically demanding and time consuming and do not provide the potential for forward genetic screens. ES cell libraries generated thus far also show evidence of bias for certain loci. *N*-ethyl-*N*-nitrosourea (ENU) is an effective random chemical mutagen of the male germline of the mouse for generating novel phenotypes (Hrabe de Angelis *et al.* 2000; Nolan *et al.* 2000). However, detection of ENU-induced single-base-pair changes responsible for a given mutant phenotype is difficult and necessitates laborious techniques such as positional cloning. *Sleeping Beauty* provides us with a useful system to mimic random *P*-element mutagenesis of Drosophila in the mouse without these drawbacks. Multiple gene disruptions can be generated from a single mouse line *in vivo*, and determination of loci affected by a given insertion is facilitated by sequences within the transposon vector that provide a molecular tag. In addition, the loci vulnerable to different methods of functional inactivation may vary, and

[1]*Corresponding author:* Room 6-160, Jackson Hall, 321 Church St., S.E. Minneapolis, MN 55455. E-mail: larga002@umn.edu

transposon vectors may mutate genes inaccessible by ES cell gene trapping or ENU-based strategies.

Before the feasibility of a large-scale mutagenesis screen can be considered, the insertion site preferences of the *SB* transposase must be determined. The insertion site preferences of the Drosophila *P* element have been most thoroughly studied (for review see Spradling *et al.* 1995). The *P*-element transposase recognizes an 8-bp sequence in the insertion site (O'Hare and Rubin 1983). However, the target sequences within a region of DNA do not have an equal chance of being disrupted because the *P*-element transposase preferentially targets the 5′ untranslated or promoter regions of genes (Spradling *et al.* 1995). It has also been shown that *P* elements insert within ~100 kb of the donor site at a rate 46- to 67-fold higher than that of regions outside that interval (Tower *et al.* 1993). In addition, there may be additional sequence requirements flanking an insertion site that create favorable molecular interactions between the target site and the *P*-element transposase (Liao *et al.* 2000). The insertion site preferences of the Tc1 and Tc3 transposable elements have been studied, although not to the extent that the *P* element has (van Luenen and Plasterk 1994). The exact nature of insertion site preference of the Tc1/mariner family has not been elucidated, and variations are observed within the family. Aside from insertion into TA dinucleotides, recent analysis of *Sleeping Beauty*-mediated transposon insertion sites in HeLa cells suggests that the sequence ANNTANNT is favored (Vigdal *et al.* 2002). It also appears that SB transposons demonstrate a "local transposition" phenomenon as seen with *P*-element transposase (Tower *et al.* 1993; Luo *et al.* 1998; Fischer *et al.* 2001).

Work in our lab has produced a panel of mice harboring novel gene-trap transposon insertions (Dupuy *et al.* 2001). We cloned and mapped 44 transposon insertions obtained after germline transposition to better characterize the SB transposase insertion site preference and to assess the ability of gene-trap transposons to mutate genes for functional genomics studies. The insertion sites were compared to 44 randomly selected TA dinucleotides from the mouse genome for differences in sequence content as well as distance and position relative to genes. Among the cloned insertion sites, 12 were within known or predicted genes. We have demonstrated stable germline inheritance of 3 transposon insertions by Southern blotting with probes from the disrupted loci. We also show that transposon gene traps are capable of efficiently disrupting the splicing of endogenous genes by splicing to a splice acceptor in the transposon vector. Finally, we describe an embryonic lethal mutant phenotype arising from a specific disruption of a predicted gene. From these data, we conclude that the *Sleeping Beauty* transposon system is a viable *in vivo* insertional mutagen in the mouse.

MATERIALS AND METHODS

Generation of transgenic mice: Mice and transgene constructs were as previously described (Dupuy *et al.* 2001).

Fluorescent *in situ* hybridization analysis: A spleen from a transgenic mouse was isolated and sent to SeeDNA Biotech for fluorescent *in situ* hybridization analysis. Lymphocytes were isolated from the spleen of a transgenic mouse and cultured at 37° in RPMI 1640 medium supplemented with 15% fetal calf serum, 3 μg/ml concanavalin A, 10 μg/ml lipopolysaccharide and 5×10^5 M mercaptoethanol. After 44 hr, the cultured lymphocytes were treated with 0.18 mg/ml BrdU for an additional 14 hr. The synchronized cells were washed and recultured at 37° for 4 hr in α-MEM with thymidine (2.5 μg/ml). Chromosome slides were made by a conventional method of preparation (hypotonic treatment, fixation, and air dry). A plasmid containing the T/MPT-eGFPF transposon was biotinylated with dATP using the BRL BioNick labeling kit (15°, 1 hr; Heng *et al.* 1992). The procedure for fluorescence *in situ* hybridization (FISH) detection was performed according to Heng *et al.* (1992) and Heng and Tsui (1993). Briefly, slides were baked at 55° for 1 hr. After Rnase A treatment, the slides were denatured in 70% formamide in 2× SSC for 2 min at 70° followed by dehydration with ethanol. Probes were denatured at 75° for 5 min in a hybridization mix consisting of 50% formamide and 10% dextran sulfate and mouse cot I DNA and prehybridized for 15 min at 37°. Probes were loaded on the denatured slides. After overnight hybridization, slides were washed and detected as well as amplified as described in Heng *et al.* (1992); FISH signals and the 4′,6-diamidino-2-phenylindole (DAPI) banding pattern were recorded separately. Images were captured and combined by CCD camera, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI-banded chromosomes (Heng and Tsui 1993).

Splinkerette PCR using blocking primers: Genomic DNAs from tail-clips from offspring of doubly transgenic founders were each digested with *Nla*III at a concentration of 50 ng/μg. The *Sau*3AI digestions are useful for cloning from the IR/DR(L) using the primers described previously (Dupuy *et al.* 2001). The *Nla*III enzyme is used to clone from the IR/DR(R) using primers described below. Splinkerettes were made by heating equimolar amounts of the primerette-long (5′-CCTCCACTACGACTCACTGAAGGGCAAGCAGTCCTAACAACCATG-3′) with the appropriate splink to 80° for 5 min and allowing them to cool to room temperature (splink *Nla*III, 5′-GTTGTTAGGACTGCTTGGAGGGGAAATCAATCCCCT-3′, 5′-phosphate; splink *Sau*3AI, 5′-GATCCATGGTTGTTAGGACTGGAGGGGAAATCAATCCCCT-3′, 5′-phosphate). Splinkerettes were then ligated to the ends of the restriction-endonuclease-treated genomic DNA. Ligation was performed at a splinkerette concentration of 7.5 μM and a DNA concentration of 25 ng/μl using T4 DNA Ligase (New England Biolabs, Beverly, MA). Primary PCR entailed primerette-short (5′-CCTCCACTACGACTCACTGAAGGGC-3′) in conjunction with the long IR/DR(R) primer (5′-GCTTGTGGAAGGCTACTCGAAATGTTTGACCC-3′). In addition to these primers, two blocking primers were added to the reaction at a final concentration twice that of the PCR primers, AD-003 (5′-ATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTG-3′, 3′-phosphate) and AD-004 (5′-TAGGGGATCCTCTAGCTAGAGTCGACCTCGAGGGGGGGCCCGGTACC-3′, 3′-phosphate). Primary PCR involved 10 cycles of 95° for 5 sec and 70° (−0.5°/cycle) for 2 min followed by 20 cycles of 95° for 5 sec and 65° for 2 min. A secondary "nested" PCR was performed using the primary PCR products diluted 1/250 within the nested PCR reaction. The second PCR entailed primerette-nested (5′-GGGCAAGCAGTCCTAACAACCA TG-

3′) in conjunction with IR/DR(R)KJC1(5′-CCACTGGGAAT GTGATGAAAGAAATAAAAGC-3′). Blocking primers AD-003 and AD-004 were also included in the nested PCR reaction at twice the concentration of the PCR primers. Nested PCR involved 30 cycles of 95° for 5 sec, 61° for 30 sec, and 70° for 90 sec. Both primary and nested PCRs incorporated a hot start at 95° for 1 min and a final extension at 70° for 10 min. The PCR was run on a 1% agarose gel, the bands were cut out, and gel was extracted. They were cloned into the pCR 2.1-TOPO vector using the TOPO TA (Invitrogen, San Diego) cloning kit. Positive clones were sequenced and analyzed.

Insertion mapping and annotation pipeline: Mapping of insertions using public data sets was performed with an automated pipeline. Each insertion was compared to the mouse genome (ref., MGSC_2002 April11_V3) using the BLAST algorithm (Schaffer *et al.* 2001) with default settings except that the number of descriptions and alignments was limited to five each. The resulting BLAST reports were subjected to the BioPerl blast parser accessed through the Bio::SearchIO module (Stajich *et al.* 2002). Genomic position data were derived from the best BLAST hit using the contig position from the BLAST report and the assembly_contig table from the mus_musculus_core_7_3b Ensembl database (Clamp *et al.* 2003). Quality comments were assigned to each mapping on the basis of previously described criteria (Roberg-Perez *et al.* 2003). Of the 44 insertion-flanking sequences examined, all had a best BLAST hit with 95% or greater identity. A total of 5 insertion-flanking sequences (01-0005, 01-0007, 01-0010, 01-0023, and 01-0043) were found to have a second-best BLAST hit with a match length ≥90% of the first, and a fraction of identical residues ≥95% of the first. These were flagged as "best blast hits are very similar" and should be considered with caution. The remaining sequences passed these criteria and were considered distinct. BLAST reports and quality comments are available through http://mouse.ccgb.umn.edu/transposon.

Nearest gene information based on the mouse Ensembl and Ensembl_espressed sequence tag (EST) gene annotation assignments were determined using the mus_musculus_core_gene and the mus_musculus_est_gene_gene tables of the ensembl_mart_7_3 database. The position of the insertion relative to the nearest was determined using the mus_musculus_core_gene_structure and the mus_musculus_est_gene_gene_structure tables of the same database. To facilitate the use of the National Center for Biotechnology Information (NCBI) gene annotation assignments, gene identifier, position, and structure information were extracted from the chr_GenomeScan.gtf file (downloaded using the NCBI ftp serve from /genomes/M_musculus/MGSCv3_Release1/maps) and formatted to match the Ensembl tables mentioned above. Nearest genes were identified by querying for gene termini present within a given range from the insertion site. The search was initiated with a range of 500 kb. If no genes were identified, the search range was progressively increased in 500-kb increments until genes were found. If a set of genes was identified, the gene closest to the insertion site was selected. Finally, positions within genes were defined as being in either exons or introns by querying the appropriate gene_structure table.

Generation of probes and Southern blotting: Primers were designed against sequence flanking each transposon. Standard PCR conditions were used to amplify probes from wild-type FVB/n strain mouse DNA. For insertion 01-0001, primers were 5′-TCGACGGAGTTGGCAGAAA-3′ and 5′-AAGTGTGG GCCCTGAGTGTC-3′. For insertion 01-0004, primers were 5′-CAAGCAACGCATCTACAAAT-3′ and 5′-ACTTGCCACAC AACCTCTAA-3′. For insertion 01-0024, primers used were 5′-TGGGAATTTGGGAAACTTGT-3′ and 5′-GGAACCGGCCAA TCATTATT-3′. PCR products were gel purified and cloned into the pCR2.1-TOPO vector (Invitrogen). Southern blotting was performed essentially as previously described (Jenkins *et al.* 1982). Genomic DNA was digested with *Eco*RV, run out on a 1% agarose gel, and transferred to a membrane.

RT-PCR: Tissues (liver, lung, spleen, thymus) were extracted from wild type and mice heterozygous for insertion 01-0032, and total RNA was extracted using Trizol (Invitrogen). Primers were designed for RT-PCR using predicted exon sequences from the Celera whole mouse genome assembly. To assess upstream splicing of the poly(A) trap, primers were designed specific for sequences just upstream of the poly(A) signal (5′-TTAGGAAAGGACAGTGGGAGTG-3′) and within an upstream exon of the endogenous gene (5′-TCAAACCCG TGAAGCACA-3′). Splicing of the green fluorescent protein (GFP) reporter into a downstream exon was also assessed using primers within GFP (5′-CTGCCCGACAACCACTA CCT-3′) and the predicted exon (5′-AGACACCTGTGCCC TCTGCT-3′). Gapdh primers are as follows: 5′-TGTCTCC TGCGACTTCAACAGC-3′ and 5′-TGTAGGCCATGAGGTC CACCAC-3′. RT-PCR was performed with 0.5 μg of total RNA using the QIAGEN OneStep RT-PCR kit. Amplification consisted of reverse transcription (50°, 30 min), initial denaturation (95°, 15 min), polymerase chain reaction (94°, 1 min; 61°, 1 min; 72°, 1 min; 35 cycles), and a final extension (72°, 10 min). RT-PCR products were gel purified using the Q-BIOgene GENECLEAN II kit and cloned into the pCR4 TOPO Vector (Invitrogen). Sequencing was performed using M13 forward (−20; 5′-GTAAAACGACGGCCAG-3′) and M13 reverse primers (5′-CAGGAAACAGCTATGAC-3′).

Northern blotting: Fifty micrograms of total RNA was electrophoresed on a 1.3% agarose, 1× MOPS, 18% formaldehyde gel with 1× MOPS running buffer at 4°. RNA was transferred to Amersham Pharmacia Biotech Hybond-N+ nitrocellulose in 10× SSC and hybridized with the appropriate probes. Probes were generated by RT-PCR of predicted exon sequences from 0.5 μg wild-type total RNA. The upper (5′-CCG GAAGTAGTTGCTCCA-3′) and lower (5′-CATGTGCTTCAC GGGTTT-3′) primers generated a probe of ~387 bp using the QIAGEN one-step RT-PCR kit as described above with an annealing temperature of 59.4°. This product was cloned, subsequently labeled with [^{32}P]dCTP isotope, and hybridized to the nitrocellulose containing the total RNA. The blot was subsequently stripped and probed with a GAPDH probe.

Quantitative real-time RT-PCR: cDNA was produced from 500 ng total RNA using the SuperScript II first strand synthesis for RT-PCR kit and treated with RNase H. A total of 10 ng cDNA was subsequently amplified using 100 nM of primers listed above for RT-PCR specific for immediately flanking exons and primers for GAPDH (also listed above) as a reference cDNA. Reactions (25 μl) were performed with the SYBR Green Master mix and run/analyzed on the ABI Prism 7700 Q-PCR machine.

Genotyping PCR: Mouse genomic DNA was used in a three-primer PCR with two primers flanking a given transposon insertion and one within the transposon. Primers for insertion 01-0032 used in Figure 9 include 0032 upper (5′-CCAGGC ATGAGAAATCTTCTTTTG-3′), 0032 lower (5′-ATGGAGAT AGGAATCACACTGGTT G-3′), and 0032 transposon lower (5′-CCTAACTGACCTTAAGACAGGGAATCT-3′). PCR entailed 5 min at 94°; 30–35 cycles of 30 sec at 94°, 57°, and 68°; and a final extension of 10 min at 68°. The wild-type product is 476 bp and the transposon insertion yields a 374-bp product.

RESULTS

Generation of transgenic mouse lines and mapping transgene insertions: We previously created two trans-

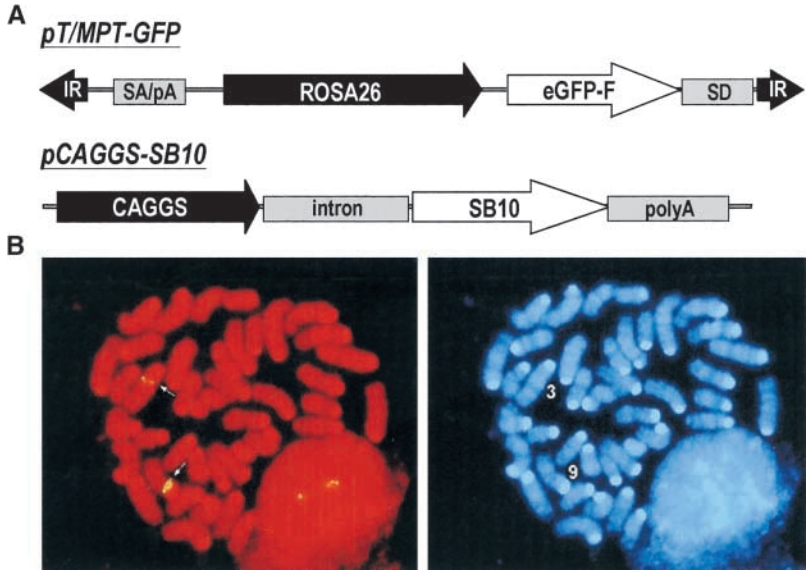


FIGURE 1.—Transgene constructs and FISH mapping of transgene insertions. (A) We created a mutagenic poly(A) trap transposon (pT/MPT-GFP) and a cassette that expresses the SB transposase from the ubiquitous CAGGS promoter (DUPUY *et al.* 2001). (B) Splenocytes from a transgenic mouse were used for fluorescent *in situ* hybridization to map the transposon transgene insertion to chromosome 9A2-3.

genic lines of FVB/n strain mice, one that ubiquitously expresses the *Sleeping Beauty* transposase from the CAGGS promoter (NIWA *et al.* 1991) and another that harbors a mutagenic gene-trap transposon (Figure 1A; DUPUY *et al.* 2001). Previous work from our lab demonstrated germline mobilization of transposons to novel genomic sites (DUPUY *et al.* 2001). Before cloning and mapping these novel sites, we first used FISH to map the transposon concatomer to chromosome 9A2-3 and the transposase concatomer to chromosome 3 (Figure 1B).

Cloning and sequencing transposon insertion sites: We used splinkerette PCR to amplify transposon junctions from genomic DNA of mice harboring novel transposon insertions (DUPUY *et al.* 2001, 2002). Briefly, genomic DNA was digested with either the *Nla*III or the *Sau*3AI restriction enzyme and ligated to a linker containing a region of nonhomology called a splinkerette. Ligated products were then used as a template for a PCR reaction using a transposon-specific primer along with a splinkerette-specific primer. However, unmobilized transposons from within the concatomer would yield a repetitive PCR product that competes with the novel transposon insertions for amplification (Figure 2). To reduce this background amplification, we included two 47-bp primers in the reaction that are complementary to the plasmid vector sequence flanking the transposons in the concatomer (Figure 2). These primers were phosphorylated at the 3′ end and designed to

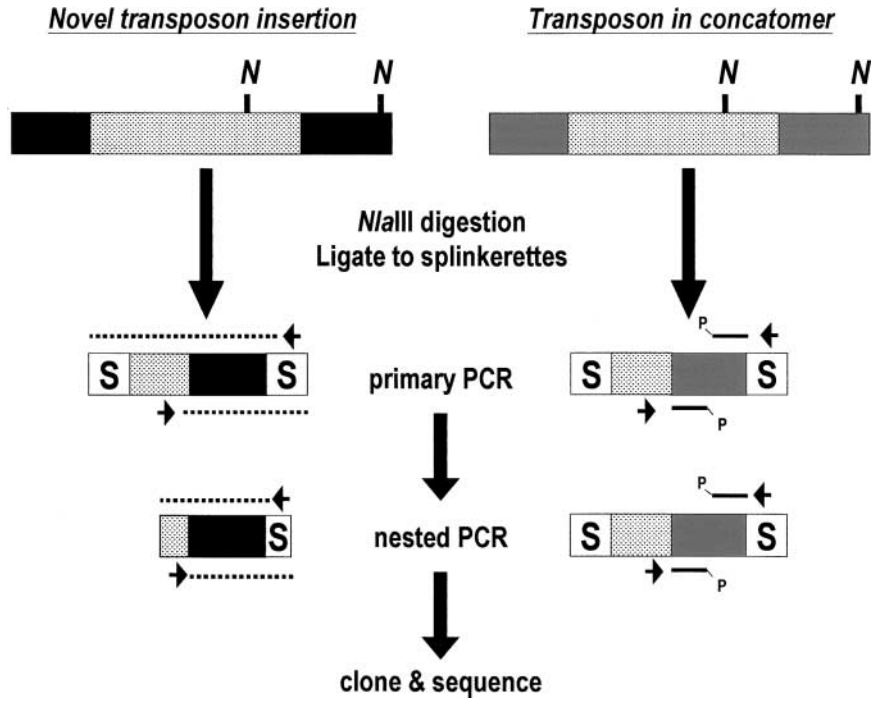


FIGURE 2.—Splinkerette PCR using blocking primers. Genomic DNA from mice harboring new transposon insertions was digested with *Nla*III restriction enzyme and ligated to a double-stranded splinkerette (S). In some cases, the new transposon insertion (solid bar) cosegregates with the transposon concatomer (stippled bar). Blocking primers were used in the PCR reaction to anneal to both strands of target molecules containing plasmid vector sequence. The 3′ phosphate prevents extension from occurring, and the primer prevents the polymerase from amplifying through the plasmid vector.

TABLE 1

Mapping of transposon insertions using the Ensembl and Celera databases

Transposon ID	Ensembl chr.	Ensembl (Mb)	Celera chr.	Celera (Mb)	Disrupted gene	Distance to nearest transcribed region (kb)
01-0001	8	8,516,260	8	5,745,814	—	10–15
01-0002	7	85,141,875	7	89,683,257	—	45–50
01-0003	9	8,811,762	9	3,218,627	—	25–30
01-0004	9	30,985,562	—	—	—	40–45*
01-0005	9	20,299,642	9	14,508,962	—	10–15
01-0006	17	20,549,623	7	NA	mCG67976	—
01-0007	9	26,322,905	—	—	—	10–15*
01-0008	1	117,875,898	1	116,946,118	—	<5
01-0009	9	19,945,903	9	14,154,737	mCG1034501	—
01-0010	9	17,923,293	9	12,215,649	—	25–30
01-0011	11	116,909,893	11	124,775,913	—	<5
01-0012	7	18,672,019	7	18,003,615	—	5–10
01-0013	9	12,638,679	9	7,101,683	—	>50
01-0014	3	11,002,646	3	7,886,889	—	10–15
01-0015	14	34,268,024	14	32,352,513	mCG52624	—
01-0016	9	11,357,475	9	5,737,713	—	>50
01-0017	9	13,555,504	9	7,867,636	—	30–35
01-0018	9	25,767,533	9	19,935,962	—	5–10
01-0019	12	76,314,937	12	80,074,264	—	10–15
01-0020	9	9,902,121	9	4,309,928	—	35–40
01-0021	9	40,499,567	9	34,541,781	mCG127192	—
01-0022	3	58,788,547	3	54,903,145	—	30–35
01-0023	10	97,383,559	3	NA	mCG9496	—
01-0024	7	83,593,175	7	88,114,871	mCG1028279	—
01-0025	3	5,043,118	3	2,110,951	—	>50
01-0026	9	16,139,314	9	10,469,239	—	20–25
01-0027	9	5,706,848	3	NA	mCG1044682/mCG5982	—
01-0028	18	77,350,897	—	—	—	<5*
01-0029	5	13,320,833	12	NA	mCG1039718	—
01-0030	17	78,259,419	17	80,326,798	mCG12054	—
01-0031	7	58,501,429	7	63,498,130	—	15–20
01-0032	6	95,318,094	6	93,423,488	mCG127714	—
01-0033	9	20,249,486	9	14,433,000	—	10–15
01-0034	14	42,090,998	14	43,480,760	—	30–35
01-0035	4	47,789,747	4	45,952,681	—	40–45
01-0036	2	97,337,023	2	94,758,373	—	>50
01-0037	14	42,509,495	14	43,905,509	—	<5
01-0038	9	88,633,709	9	81,740,569	—	20–25
01-0039	9	16,187,918	9	10,420,637	—	15–20
01-0040	9	12,609,088	9	6,967,282	—	10–15
01-0041	8	4,660,958	8	1,931,386	mCG1814	—
01-0042	9	11,815,533	9	6,206,464	—	5–10
01-0043	12	29,696,962	10	75,801,846	—	<5
01-0044	9	99,726,511	9	95,742,434	mCG1032876	—

Chromosome and megabase position are shown for both assemblies if available. In addition, distance from the nearest transcribed region and genes disrupted are shown for the Celera assembly. An asterisk indicates a distance based on Ensembl. Updated insertion position information using the NCBI30 version of the mouse genome and recent genomic annotation data sets are available at the Mouse Transposon Insertion Database (http://mouse.ccgb.umn.edu/transposon).

have melting temperatures at least 10° higher than those used for splinkerette PCR. During the PCR reaction, these "blocking primers" should anneal to the DNA fragments from the concatomer and block extension of the splinkerette primers through these regions. Following PCR, products were gel purified and cloned. Sequence for positive clones was obtained by high-throughput preparation of DNA followed by sequencing. Sequences were then processed to remove remaining transposon and splinkerette sequences prior to mapping.

Mapping novel transposon insertion sites: Cloned sequences were compared against Celera's whole mouse

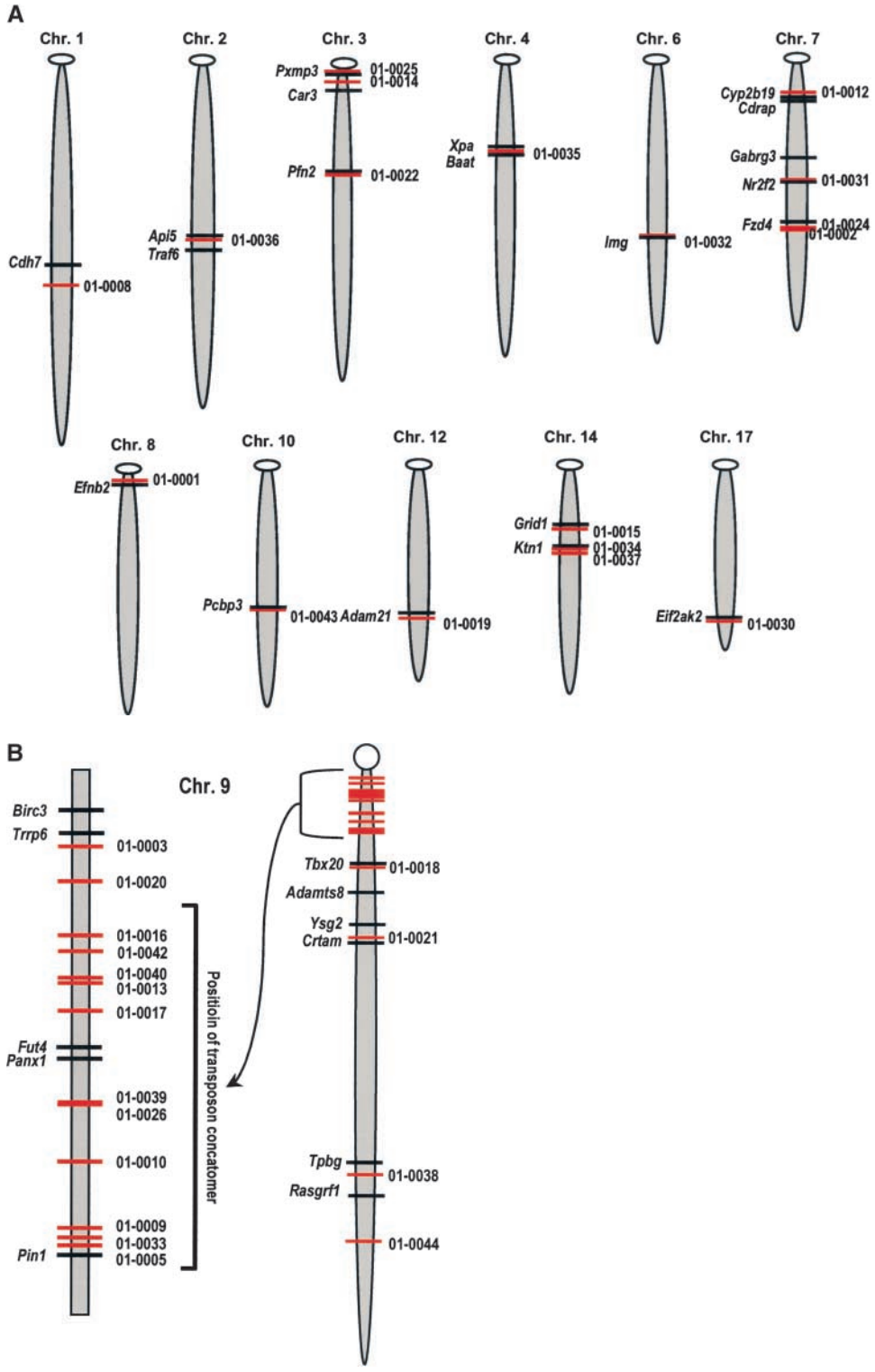


FIGURE 3.—Map position of transposon insertion sites. (A) We obtained 44 novel transposon insertions sites and mapped them using Celera's whole mouse genome assembly. The position of each insertion is shown by a red line with the transposon ID to the right. The map position of known genes is indicated by a black line with the gene name shown to the left. (B) An enlarged view of chromosome 9. The position of the concatomer is indicated as determined by FISH analysis.

genome assembly using the BLAST search tool (ALTSCHUL *et al.* 1990). The average length of cloned flanking sequence analyzed was ~133 bp, and the average percentage of identity with specific sites in the Celera assembly was ~99%. Insertions were also later mapped using the Ensembl database (http://www.ensembl.org; GREGORY *et al.* 2002; HUBBARD *et al.* 2002). We determined the chromosome and nucleotide position for each insertion as well as the distance to the nearest transcribed region according to Celera for those insertions not in genes (Table 1). Clones were eliminated from further analysis if the map position was ambiguous due to the presence of repetitive sequence or if the clone represented an insertion into an adjacent transposon (four clones). We obtained a total of 44 mapped transposon insertion sites with 19 of those mapping

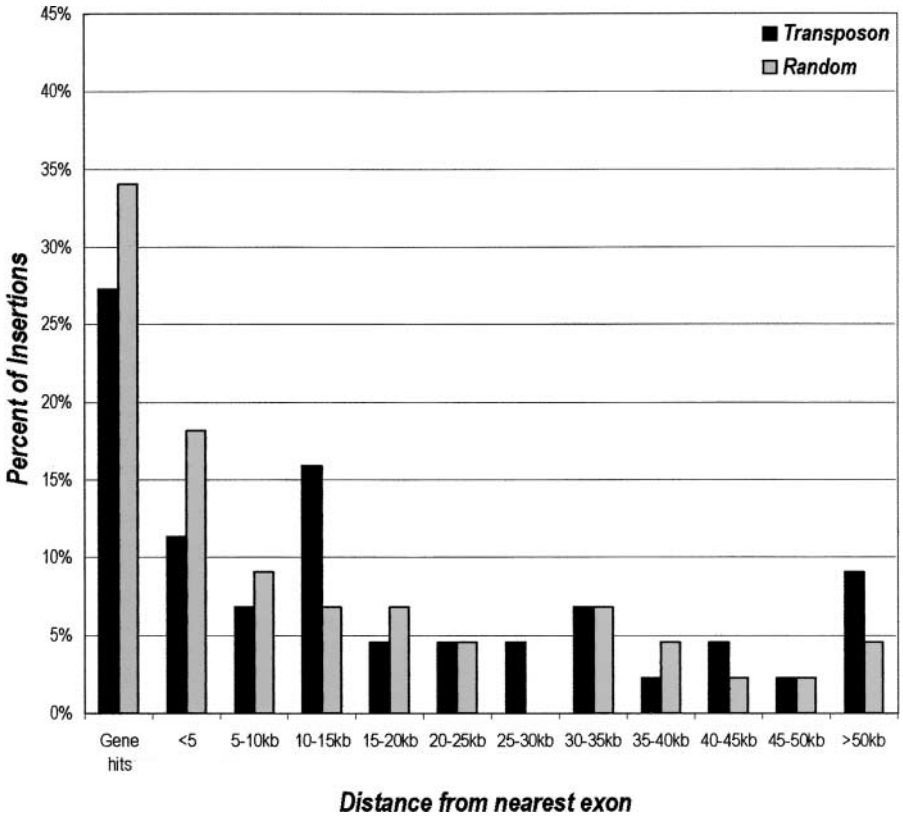


FIGURE 4.—Distribution of transposon insertions relative to genes. The distance to the nearest transcription unit was calculated for each insertion as well as for each randomly selected TA dinucleotide. These distances were put into groups of 5 Mb. The percentage of insertions for each category as well as the frequency of gene hits is shown.

to chromosome 9 where the transposon concatomer is located (Figure 3B). The remaining insertions occurred on other chromosomes without any obvious preference for chromosome or region (Figure 3A).

Of the 19 insertions that mapped to chromosome 9, 13 are within the interval containing the transposon concatomer (Figure 1B). These transposition events can be attributed to local transposition in which an excised transposon tends to integrate near the donor site. However, unlike *P* elements in which local transposition occurs over a 100-kb interval, *Sleeping Beauty* transposons have a much larger local transposition interval (TOWER *et al.* 1993). The *Sleeping Beauty* local transposition interval appears to be between 5 and 15 Mb, depending on the exact location of the concatomer within band 9A2-3. The cluster of local transposition events detected occurred between the *Trrp6* gene (3.2 Mb, 1 cM) and the *Pin1* gene (14.5 Mb, 4 cM).

Determination of insertion site preferences: To determine any bias in transposon insertion sites, we used a random number generator to select 44 TA dinucleotides from the genome. Random TA dinucleotides were noted if they occurred within a transcribed region. Otherwise, the distance to the nearest transcribed region was determined. The number of hits within known and predicted transcription units in the control group was 34% (14 of 44) compared to 27% (12 of 44) for the transposon insertion group (Figure 4). Outside transcribed regions, there does not appear to be any obvious preference for SB transposons to insert near or distant from genes. As shown in Figure 4, the distribution of transposon insertions relative to transcription units is nearly the same as randomly selected TAs, indicating the randomness of transposition. Previous work has demonstrated that *P* elements preferentially integrate into the 5′ region of genes (SPRADLING *et al.* 1995). We examined more closely those insertions that occurred within 20 kb of a known transcribed region. Of the 16 insertions occurring near genes, we determined that 8 of them were 5′ while the remaining 8 were 3′ of the nearest gene (data not shown).

We also compared the sequence flanking the TA dinucleotide between the transposon insertion and control groups to detect any differences in nucleotide content. Transposable elements in the Tc1/mariner family require only a TA dinucleotide for insertion (PLASTERK *et al.* 1999). If this were the sole requirement, we would expect to find no differences in the sequence flanking the TA dinucleotides in either group. We compared 25 bp flanking both sides of each transposon integration site and randomly selected TA dinucleotides. An unpaired *t*-test was performed to compare the nucleotide content of flanking sequence from each group. Analysis of each individual nucleotide revealed a decrease in the percentage of cytosine in the sequence flanking the transposon insertions ($P = 0.0091$). We also found that the transposon insertion sites occurred in regions with higher AT content ($P = 0.015$) and lower GC content ($P = 0.015$) when compared to the control group.

Although these differences were statistically significant, they were slight. We aligned the junction sequences to determine if the differences in nucleotide

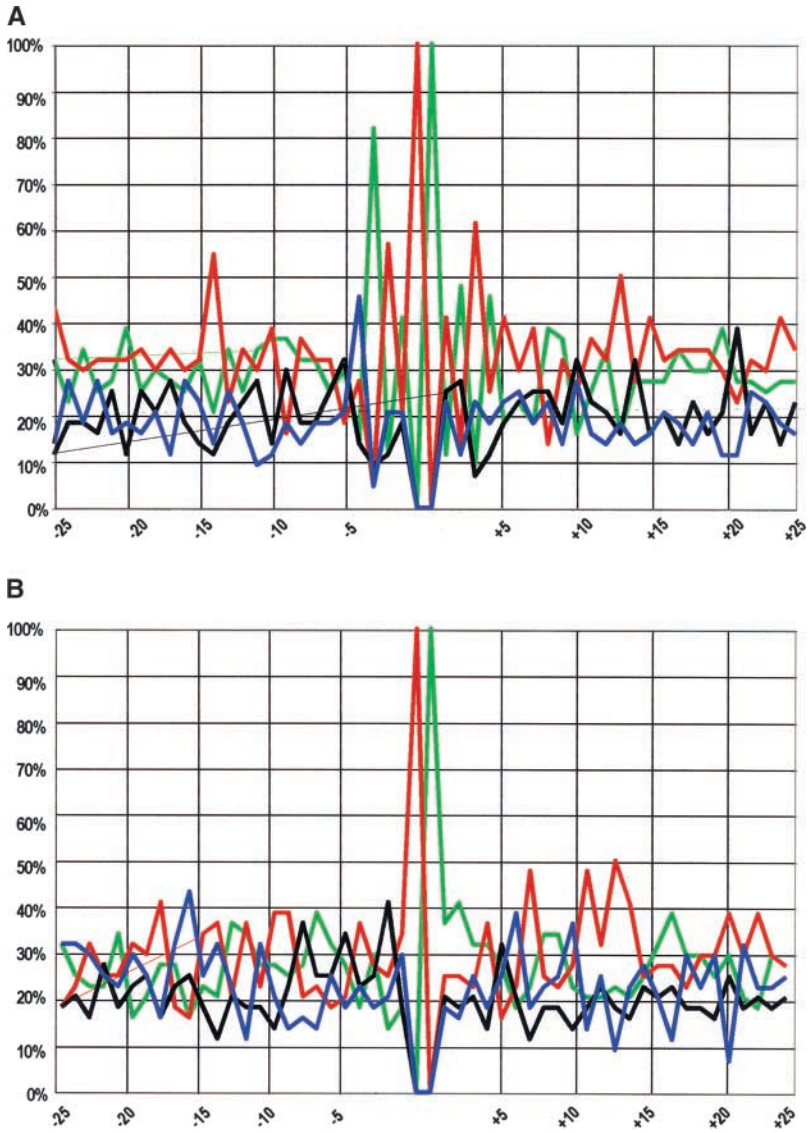


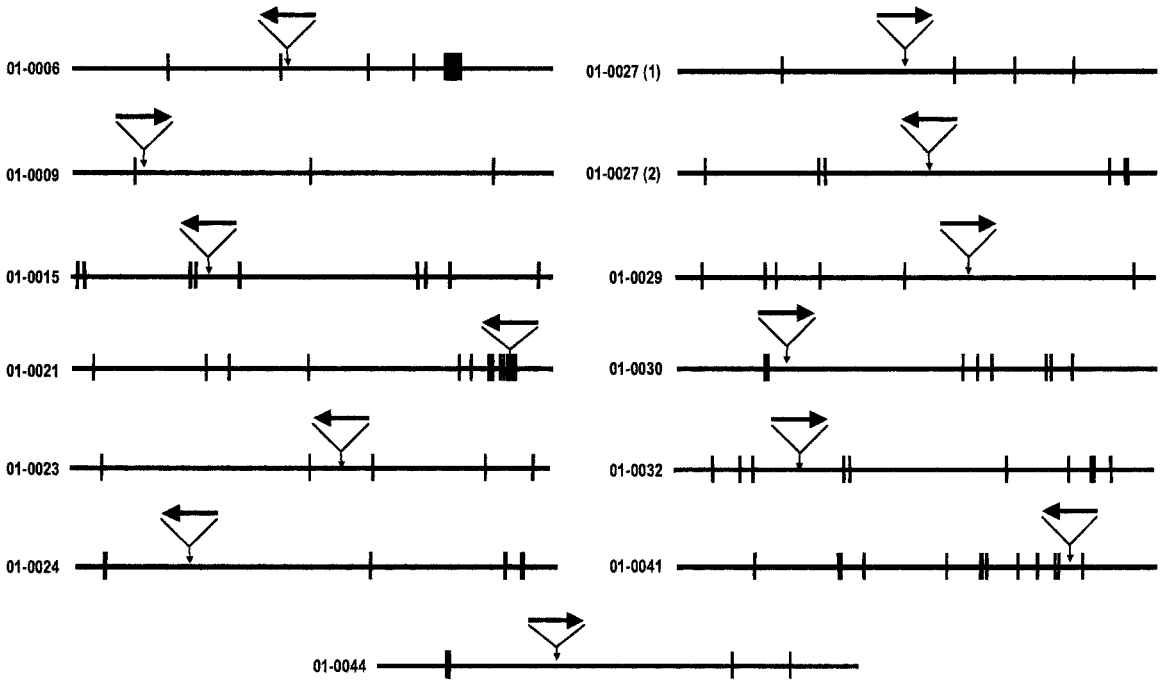
FIGURE 5.—Nucleotide content in regions flanking the site of insertion. A total of 25 bp of sequence flanking each side of the insertion site was analyzed. The percentage of each nucleotide at each position was calculated and graphed for both the (A) transposon insertion group and the (B) randomly selected TA dinucleotide group. (Green) A, (red) T, (black) G, (blue) C.

content could be attributed to a consensus, other than the TA dinucleotide, used by the transposase (Figure 5). Although we did not find any consensus nucleotides strictly required, other than the TA, we did detect strong preferences. Most of the insertions had an adenine at position −3 and a thymine at position +3 (82 and 61%, respectively). Therefore, the SB transposase appears to prefer an expanded consensus of ANNTANNT, consistent with a recent report for SB transposon insertions in zeocin-selected HeLa cells (VIGDAL *et al.* 2002).

Analysis of transposon insertions in genes: We cloned 12 transposon insertions that are within 13 genes according to the Celera mouse genome assembly (Figure 6). Of these insertions, 6 are in the same orientation as transcription and would be predicted to disrupt the gene. All 12 insertions are within introns and are spread throughout the length of genes. Twelve of the insertions are within introns flanked by coding exons and 1 is within the 3′ untranslated region (mCG1814). Of the 13 transcripts, 6 are predicted by Celera and are not supported by homology to mouse or human cDNA clones (Figure 6). Eight of the predicted transcripts do not contain a complete open reading frame. Thus, other transposon insertions that mapped close to transcription units may in fact be within them, because unidentified upstream or downstream exons exist.

Germline transmission of transposon insertions: Several of the mice harboring novel transposon insertions were bred to demonstrate germline transmission of the transposons. We performed PCR on wild-type mouse genomic DNA to amplify sequences flanking transposon insertion sites for use as probes. Southern blotting was performed on both offspring and parental tail-biopsy DNA was digested with *Eco*RV restriction enzyme. We were able to demonstrate germline transmission in roughly Mendelian ratios (Figure 7). The rearranged and wild-type bands corresponded to the predicted sizes on the basis of the sequence obtained from the Celera database (data not shown).

RNA analysis of mutant transcripts: For transposon-



Transposon ID	Gene	Predicted gene function	Homology to ESTs	Other Homology
01-0006	mCG67976	Krueppel-related C2H2-type zinc-finger	AA178652	Zpf118
01-0009	mCG1034501	No known function	none	none
01-0015	mCG52624	No known function	none	none
01-0021	mCG127192	No known function	BF782386, BB624245	RIKEN cDNA 8030448I23
01-0023	mCG9496	Guanylate cyclase alpha 2 subunit	BB664048, BB667027	RIKEN cDNA E530014A09, E030024I21
01-0024	mCG1028279	Reverse Transcriptase-related	AA755787	none
01-0027(1)	mCG1044682	No known function	none	none
01-0027(2)	mCG59825	No known function	none	none
01-0029	mCG1039718	No known function	none	none
01-0030	mCG12054	No known function	BI106850	none
01-0032	mCG127714	Mitochondrial carrier protein-related	BB852792, BM454154	none
01-0041	mCG1814	Shc SH2-binding protein 1	BM461238, BM234685, UniGene #Mm.37801	mPAL (AF017152)
01-0044	mCG1032876	No known function	none	none

FIGURE 6.—Distribution and orientation of transposon insertions in genes. (Top) Twelve transposons are inserted within transcription units of known or predicted genes. The transposon ID is shown next to a diagram of each insertion. The direction of transcription is from left to right for each gene, and exons are represented with black vertical lines. The approximate location of each transposon is also indicated. The arrow represents the orientation of the mutagenic cassette within the transposon. (Bottom) Additional information for each gene is listed.

tagged mutagenesis to be successful in the mouse, gene-trapping elements within the transposon must be capable of producing a mutant transcript upon gene insertion. The gene trap used here is designed to truncate endogenous transcripts via splicing and polyadenylation. If the vector functions as designed, then splicing from an upstream exon of the disrupted gene will join it to the splice acceptor within the transposon vector. The vector includes stop codons in all three frames and a polyadenylation signal (Figure 8A). The downstream portion of the transposon is a poly(A) trap that is predicted to express GFP when it is provided with a poly(A) signal from an endogenous gene via splicing to a downstream exon. The GFP gene is driven by the ubiquitous ROSA26 promoter (KISSENBERTH *et al.* 1999), followed by a splice donor from the HPRT gene. To test the splicing efficiency of intron-inserted gene-trap transposons, total RNA was analyzed by RT-PCR, Northern blot, and real-time quantitative RT-PCR. Insertion 01-0032 lies within the mCG127714 transcription unit as annotated by the Celera whole mouse genome assembly (Figure 6). Predicted exons of this locus were analyzed with the NCBI BLAST search tool for identity to characterized ESTs with known expression patterns to determine appropriate tissues for transcript analysis. The predicted mCG127714 exons showed significant identity to ESTs isolated from brain, intestine, liver, lung, spleen, testes, and thymus.

RT-PCR was employed to assess the efficiency of both the upstream splice acceptor and the downstream splice donor within the transposon vector. Primers were designed for the predicted upstream and downstream exons as well as for sequences predicted to be transcribed within the gene-trap vector (Figure 8A). RT-PCR was performed on total RNA of heterozygous mice (Figure 8B). This analysis revealed specific upstream and down-

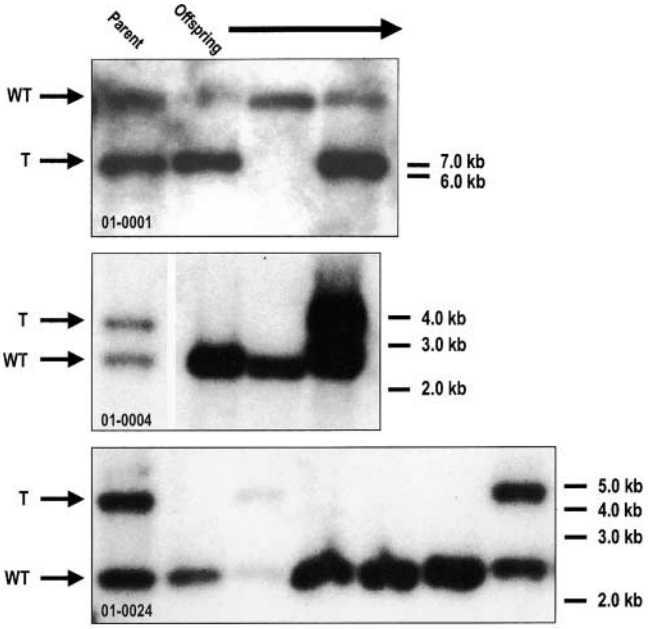


FIGURE 7.—Southern blotting of parental and offspring DNA. Several mice harboring mapped insertion sites were bred to wild-type mice, and offspring were genotyped by Southern blotting using probes specific for the sequence flanking one side of each insertion. Southern blotting results are shown for three mice. In each case, parental DNA is shown at the left with the offspring DNA in the adjacent lanes. The wild-type allele is indicated by "WT" and the transposon allele is indicated by "T." The transposon ID corresponding to each probe is also indicated.

stream splicing of the gene-trap vector with the endogenous gene with no detectable product in the wild-type controls. Cloning and sequencing of RT-PCR products revealed the expected sequence resulting from the given splicing reactions (Figure 8C), including the presence of stop codons in all three reading frames from within the upstream chimeric transcript. Thus, the gene-trap transposon is capable of splicing with endogenous genes to create chimeric transcripts.

Since RT-PCR has extraordinary sensitivity, the frequency of transcript mutation needed to be demonstrated by other means. Thus, a Northern blot was performed on total liver and spleen RNA from wild-type mice and heterozygous mice carrying insertion 01-0032 to determine whether the mutant transcript occurs at a significant frequency relative to the wild-type transcript. A probe composed of predicted exon sequences upstream of the gene-trap insertion was produced by RT-PCR and thus is predicted to hybridize with both the wild-type and the predicted truncated transcript. Figure 8D shows the Northern blot, which reveals the presence of a novel, smaller transcript within the liver and spleen RNA of the heterozygous carrier mouse, but which is absent in the wild type. This indicates that efficient transcript truncation has occurred in the carrier mice. The intensity of the band indicates that splicing of the endogenous gene into the mutagenic gene trap occurs at a high frequency at levels similar to normal exon-to-exon splicing from the wild-type allele.

To further validate that the gene-trap insertion within this particular transcription unit actually results in reduction of the amount of wild-type transcript produced, real-time quantitative RT-PCR was performed on liver and spleen RNA from heterozygous carrier and wild-type mice. Primers specific for exons immediately flanking the intron into which the transposon inserted were used to amplify a wild-type cDNA. These primers are incapable of amplifying a product from the mutant transcript cDNA due to its premature truncation. Transcript levels of the gene in question were compared to *Gapdh* as a reference and revealed that wild-type transcript levels were reduced by half within the liver of carrier mice relative to wild type (Figure 8E). Levels were also decreased in the spleen to a lesser extent. These results prove that transposon gene traps can decrease the amount of wild-type transcript produced when inserted into an intron.

Mouse phenotype analysis: Finally, we attempted to generate mice homozygous for several insertions to assess any resultant phenotypes. Heterozygous carrier mice were intercrossed and offspring were genotyped by three-primer PCR (DUPUY *et al.* 2002). Figure 9A displays the genotypic results of intercrosses for several different transposon insertions within predicted genes. The lack of mice homozygous for insertion 01-0032 is statistically significant ($P = 0.011$), indicating an embryonic lethal phenotype presumably caused by the loss of the disrupted gene's product. We performed timed pregnancies for the intercross of mice heterozygous for insertion 01-0032 to further define the phenotype. We were able to isolate embryos at E8.5 that were homozygous for the insertion and Figure 9B shows the appearance of these embryos compared to wild-type and heterozygous littermates. Although gastrulation and somitogenesis appears to have initiated, mutant embryos are growth retarded, underdeveloped, and starting to recess. The gene mutated by insertion 01-0032 is a mitochondrial carrier protein-related gene with the predicted amino acid sequence indicated in Figure 9C. This predicted gene has not been fully characterized, but the predicted protein contains sequence motifs and a tripartite structure characteristic of the mitochondrial carrier protein family (PALMIERI 1994). Similar to insertion 01-0032, we have been unable to obtain mice homozygous for insertion 01-0009, suggesting that it also causes a recessive embryonic lethal mutant phenotype. In contrast, homozygous mice for several insertions within and outside of genes have been generated. Notably, these insertions are in an antisense orientation with respect to the disrupted gene. The insertions noted in Figure 9A that produce viable homozygotes show no significant observable phenotype, but are currently being screened more aggressively, since preliminary data suggest that some

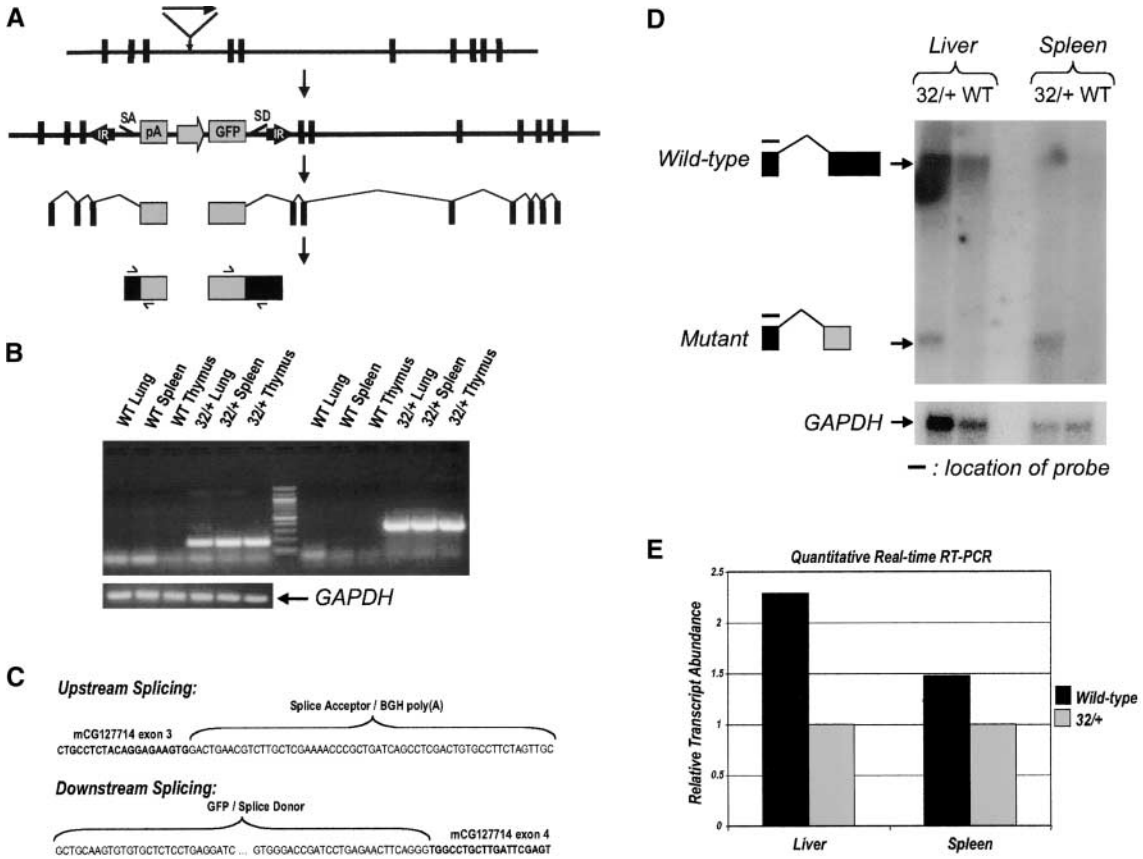


FIGURE 8.—RNA analysis of gene disruption. Insertion 01-0032 was assessed for the ability of the transposon gene trap to function as designed for putative transcriptional disruption of the gene in question. (A) Primers (indicated on figure) were designed to assess splicing of the upstream exon with the splice acceptor within the transposon. Likewise, splicing of the splice donor within the transposon with the downstream endogenous exon was examined using the same assay with appropriate primers. (B) Total RNA was extracted from appropriate tissues of wild-type mice and mice heterozygous for insertion 01-0032 and subjected to RT-PCR. (C) RT-PCR products were cloned and sequenced to demonstrate the expected spliced products as shown. (D) Northern blot analysis of total RNA from wild type and mice heterozygous for insertion 01-0032 was also performed with a probe specific for exons upstream of the gene-trap insertion, which should recognize the wild-type full-length transcript and the mutant truncated transcript. (E) Real-time quantitative RT-PCR was performed with primers that amplify across the intron disrupted by the gene-trap transposon insertion on RNA from wild-type and heterozygous tissues; an average of two experiments is shown.

of these partially suppress gene expression. We can conclude from insertion 01-0032 that SB transposon gene traps can be very efficient insertional mutagens yielding novel phenotypes.

DISCUSSION

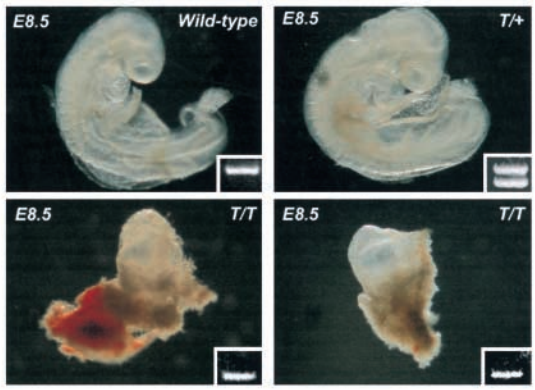
The insertion site preferences for several commonly used transposable elements have been examined (VAN LUENEN and PLASTERK 1994; LIAO *et al.* 2000). These studies have indicated that there are sequences outside the consensus target site that are preferred by the transposase. We have cloned and mapped 44 *Sleeping Beauty* transposon insertions from a panel of 30 mice to determine if the SB transposase displays any insertion site preferences. Generally, we are unable to identify any insertion site preference that would severely restrict the number of potential genomic integration sites for *Sleeping Beauty* transposons. Furthermore, we demonstrate that gene-trap transposons are capable of mutating disrupted endogenous genes upon intronic insertion, reducing wild-type transcript levels, and producing mutant phenotypes. These data suggest that *Sleeping Beauty* will be useful as a random germline insertional mutagen in mice.

It is apparent that the SB transposase displays a local transposition tendency similar to *P*-element transposase (TOWER *et al.* 1993). We estimate the local transposition interval for SB transposase to be between 5 and 15 Mb compared to 100 kb for the *P*-element transposase. Although ~43% of our insertions mapped to the donor chromosome, one-third of our total mapped insertions are within our estimated "local hopping" interval. The reported frequency of local transposition using the *Sleeping Beauty* transposon system has varied between 50% (LUO *et al.* 1998; DUPUY *et al.* 2001) and 83% (FISCHER *et al.* 2001). It is not clear whether these differences are significant and what factors (*e.g.*, location of the donor

A

Insertion #	Wild-type	Heterozygous	Homozygous
01-0001	9	10	6
01-0006	10	15	7
01-0009	5	12	0
01-0024	7	16	7
01-0032	10	16	0
01-0041	5	10	4

B



C

MDAPGFTASLVAGGVAGVSVDLILFPLDTIKTRLQSPQGFNKAGGFRGIYAGVPSAAVGSFPNAAAFFLTYEY
VKSLLHTDSTSHFKPVKHMLAASTGEVVACLIRVPSEVVKQRAQVSASSKTLQIFLTILSEEGIQGLYRGYKST
VLREIPFSLVQFPLWESLKALWAWRRGHVVDSWQSAVCGAFAGGFAAAVTTPLDVAKTRIMLAKAGSSTAV
GNVLSAMHGVWRSQGLAGLFAGVLPRMAAISMGGFIFLGAYDQARSLLLEVGRKSP*

Figure 9.—Genotype and phenotype analysis of insertions. (A) Male and female mice heterozygous for given insertions were bred in attempts to generate homozygous mice for each gene insertion; genotype results from such breedings are indicated. (B) Since no viable homozygous mice could be produced for insertion 01-0032, timed pregnancies were performed to isolate homozygous embryos at E8.5. Pictures of the embryos (not to scale) and three-primer genotyping PCR results are represented (T, transposon allele; +, wild-type allele). (C) The predicted amino acid sequence of the predicted gene disrupted by insertion 01-0032 is shown; amino acid residues that are underlined and in boldface type indicate sequence motifs characteristic of the mitochondrial carrier protein family.

locus) may affect this rate. However, in this work the amplification of novel insertion sites in mice or cells with a donor locus consisting of a multicopy concatomer of transposons, rather than of single-copy elements as in Luo *et al.* (1998) and Fischer *et al.* (2001), proved to be difficult and necessitated the use of blocking primers to inhibit amplification of transgene vector sequences. The efficiency of these blocking primers is unknown and thus amplification of novel insertions linked to the donor locus could have been compromised in these mice by the presence of the concatomer, biasing toward a lower observed incidence of local transposition. Despite this potential bias, 45% (13/29) of the insertion sites cloned from mice that retained the transposon conatomer were mapped to chromosome 9, while mice that lacked transgene sequences yielded insertions 40% (6/15) of the time on chromosome 9 (data not shown). In addition, the efficiency of cloning insertions in the presence and absence of the concatomer was essentially the same. Thus, it would appear that we were equally successful in cloning local transposon insertions in both groups and that no bias exists.

Analysis of the sequence flanking each insertion site did reveal a tendency for SB transposase to select TA dinucleotides that occurred within AT-rich regions (Figure 5). Although this difference was statistically significant, the AT content flanking the transposon insertion sites was only 10% higher than that of the sequence flanking randomly selected TA dinucleotides. The SB transposase also appears to prefer the consensus ANNTANNT. In this regard, SB transposase seems to be more similar to Tc1 than to Tc3 in its insertion site preference (van Luenen and Plasterk 1994). However, the AT content of the sequence flanking transposon insertion sites is still significantly higher even if the −3 and +3 positions are excluded from the analysis ($P = 0.021$). Therefore the preferred consensus site does not entirely account for the increase in AT content in the sequence around transposon insertion sites. AT richness and sequence preferences at −3 and +3 are both preserved even when insertions mapped to chromosome 9 are excluded from the data set (data not shown). This indicates that our results are not biased by any altered sequence content in the 9A2-3 region of the mouse genome. Other published transposon insertion junction sequences tend to support our observations of the preferred consensus site (data not shown; Ivics *et al.* 1997; Luo *et al.* 1998; Yant *et al.* 2000; Fischer *et al.* 2001; Horie *et al.* 2001; Dupuy *et al.* 2002). Recent data from cloned sequences flanking SB-mediated transposition events in HeLa cells selected in zeocin show a strong tendency of transposons to insert within an AT-repeat: ATATATAT with the center TA as the site of insertion (Vigdal *et al.* 2002). The data presented herein are consistent with this trend, but the TA appears to be the only sequence absolutely necessary for transposon insertion in both data sets. Additional work will be required to determine if this insertion tendency will significantly reduce the mutagenicity of the *Sleeping Beauty* transposon system. However, our data reveal that SB transposons can insert into a variety of genes in regions all over the genome.

The transposon insertion sites do not appear to differ notably from randomly selected TA dinucleotides in their position relative to transcribed sequence. We did not expect to see a significant difference in the number of SB insertions within genes *vs.* random TAs when we compared the two groups, and this was verified by our data. It is important to note that transcribed regions are most likely underrepresented in the Celera mouse genome assembly. Many predicted gene transcripts lack a complete open reading frame, and therefore many of

the transposon insertions that occurred near genes may have actually occurred within the transcription unit. In fact, two of the insertions not mapped within genes (01-0001 and 01-0020) appear to be within introns of specific EST clones (data not shown). It is thus evident that informatics issues confuse the number of gene insertions detected in our analysis. The gene insertions noted here are an estimate of the total number of transcription units actually disrupted by transposons in this screen. All 44 germline transposon insertions were reanalyzed using a variety of gene-calling algorithms. As mentioned, we have mapped the transposon insertions using the public version of the mouse genome assembly (MGSC V3) using our insertion mapping and annotation pipeline (IMAP), which automatically maps insertion sites (Roberg-Perez *et al.* 2003). Using IMAP we have been able to map all 44 transposons to specific chromosomal and nucleotide positions on the public version of the mouse genome (Table 1). Of the 37 insertions that were successfully mapped to a chromosome and nucleotide position with the Celera system, all but one is mapped to the same chromosome using IMAP. Furthermore, the nucleotide positions assigned for these insertions never differed >8 Mb between IMAP and Celera assignments with an average difference of 4 Mb (K. Roberg-Perez, unpublished data). Consistency between the assemblies is further indicated by the colinearity of assigned insertion positions, with the exception of two adjacent insertions in the chromosome 9 cluster. These results confirm and extend our initial mapping work using the Celera mouse genome assembly.

Furthermore, it is clear that the gene trap and poly(A) trap of the transposon function as predicted to effectively disrupt endogenous wild-type gene expression. Stop codons within all three frames of the gene-trap transposon spliced into the upstream exons of an endogenous gene are predicted to cause a truncated protein product to be generated upon translation. The resultant mutant transcript should demonstrate expression patterns identical to the full-length endogenous transcript, but when translated may lack the function of the wild-type protein product of the gene. We have demonstrated the ability of our transposon gene trap to efficiently mutate a gene upon intronic insertion, eliciting a mutant phenotype, and are currently assessing the remaining insertions for their effects at the sequence and phenotype level.

Taken together, these results suggest that random *in vivo* germline transposon-tagged mutagenesis is a feasible approach to functional genomics in the mouse. Given the transposition frequency we have obtained in the mouse male germline (Dupuy *et al.* 2001), along with the frequency of gene insertion predicted here, the *Sleeping Beauty* system does not appear to be efficient enough to perform a genome-wide mutagenesis screen without a substantial increase in transposition frequency. To rival ENU mutagenesis, which is thought to mutate ~30 genes per gamete in treated males (Justice *et al.* 2000), we will need to produce between 135 and 270 insertions per gamete. This number accounts for roughly one-third of the insertions in transcription units ($30 \times 3 = 90$), subtracts all local transposition events ($90 \times 1.5 = 135$), and can be calculated for a gene trap that functions in both orientations (135) or in only one orientation ($135 \times 2 = 270$). Of course, rapid mutant gene identification in the downstream part of any screen compensates for the reduced mutagenicity of SB transposon mobilization. Nevertheless, we are attempting to improve transposition frequency 10-fold using improved transposase and transposon transgenes. Like ENU, our preliminary data suggest that SB transposon insertions will be found to cause hypomorphic alleles in some cases. Unlike ENU, SB transposon vectors could be engineered to express useful reporter molecules such as GFP, β-galactosidase, or the Cre recombinase in the temporal and spatial pattern of the disrupted endogenous gene.

It should be possible to utilize the local transposition phenomenon that we observed to focus transposon mutagenesis into defined regions of the genome of high biological interest. Used in this way, saturation mutagenesis could be performed in a 5- to 15-Mb region surrounding a transposon concatomer array. In the mouse this corresponds to ~2–7 cM. In these experiments, we observed two new insertions per gamete with approximately one-third of those attributed to local transposition. Thus, it will be feasible to achieve a 1× coverage of a 10-Mb region, with transposon insertions every 20 kb, in as few as 750 mice. In addition, mobilization of transposons within the germline of mice could be utilized for chromosome engineering, mobilizing border elements, and to further our understanding of gene clusters.

We thank the University of Minnesota Mouse Genetics Laboratory for their assistance and Steve Buganski for his mouse husbandry. We also thank Dr. William Shawlot for his input and assistance with the E8.5 embryos and Craig Eckfeldt for his guidance with quantitative real-time PCR. This work was supported by the Arnold and Mabel Beckman Foundation and the National Institutes of Health (NIDA R01DA14764).

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Clamp, M., D. Andrews, D. Barker, P. Bevan, G. Cameron *et al.*, 2003 Ensembl 2002: accommodating comparative genomics. Nucleic Acids Res. **31:** 38–42.

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.

Dupuy, A. J., K. Clark, C. M. Carlson, S. Fritz, A. E. Davidson *et al.*, 2002 Mammalian germ-line transgenesis by transposition. Proc. Natl. Acad. Sci. USA **99:** 4495–4499.

Eide, D., and P. Anderson, 1985 Transposition of Tc1 in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA **82:** 1756–1760.

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.

Gregory, S. G., M. Sekhon, J. Schein, S. Zhao, K. Osoegawa *et al.*, 2002 A physical map of the mouse genome. Nature **418:** 743–750.

Hartl, D., 2001 Discovery of the transposable element mariner. Genetics **157:** 471–476.

Heng, H. H., and L. C. Tsui, 1993 Modes of DAPI banding and simultaneous in situ hybridization. Chromosoma **102:** 325–332.

Heng, H. H., J. Squire and L. C. Tsui, 1992 High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. Proc. Natl. Acad. Sci. USA **89:** 9509–9513.

Horie, K., A. Kuroiwa, M. Ikawa, M. Okabe, G. Kondoh *et al.*, 2001 Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. Proc. Natl. Acad. Sci. USA **98:** 9191–9196.

Hrabe de Angelis, M., H. Flaswinkel, H. Fuchs, B. Rathkolb, D. Soewarto *et al.*, 2000 Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat. Genet. **25:** 444–447.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen *et al.*, 2002 The Ensembl genome database project. Nucleic Acids Res. **30:** 38–41.

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.

Izsvak, Z., Z. Ivics and R. H. Plasterk, 2000 Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. J. Mol. Biol. **302:** 93–102.

Jenkins, N. A., N. G. Copeland, B. A. Taylor, H. G. Bedigian and B. K. Lee, 1982 Ecotropic murine leukemia virus DNA content of normal and lymphomatous tissues of BXH-2 recombinant inbred mice. J. Virol. **42:** 379–388.

Justice, M. J., D. A. Carpenter, J. Favor, A. Neuhauser-Klaus, M. Hrabe de Angelis *et al.*, 2000 Effects of ENU dosage on mouse strains. Mamm. Genome **11:** 484–488.

Kissenberth, W. C., N. T. Brettingen, J. K. Lohse and E. P. Sandgren, 1999 Ubiquitous expression of marker transgenes in mice and rats. Dev. Biol. **214:** 128–138.

Liao, G. C., E. J. Rehm and G. M. Rubin, 2000 Insertion site preferences of the P transposable element in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **97:** 3347–3351.

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.

Medhora, M., K. Maruyama and D. L. Hartl, 1991 Molecular and functional analysis of the mariner mutator element Mos1 in Drosophila. Genetics **128:** 311–318.

Niwa, H., K. Yamamura and J. Miyazaki, 1991 Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene **108:** 193–200.

Nolan, P. M., J. Peters, M. Strivens, D. Rogers, J. Hagan *et al.*, 2000 A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat. Genet. **25:** 440–443.

O'Hare, K., and G. M. Rubin, 1983 Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell **34:** 25–35.

Palmieri, F., 1994 Mitochondrial carrier proteins. FEBS Lett. **346:** 48–54.

Pelicic, V., S. Morelle, D. Lampe and X. Nassif, 2000 Mutagenesis of Neisseria meningitidis by in vitro transposition of Himar1 mariner. J. Bacteriol. **182:** 5391–5398.

Plasterk, R. H., Z. Izsvak and Z. Ivics, 1999 Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet. **15:** 326–332.

Roberg-Perez, K., C. M. Carlson and D. A. Largaespada, 2003 MTID: a database of *Sleeping Beauty* transposon insertions in mice. Nucleic Acids Res. **31:** 78–81.

Schaffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge *et al.*, 2001 Improving PSI-BLAST protein database search sensitivity with composition-based statistics and other refinements. Nucleic Acids Res. **29:** 2994–3005.

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.

Spradling, A. C., D. M. Stern, I. Kiss, J. Roote, T. Laverty *et al.*, 1995 Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc. Natl. Acad. Sci. USA **92:** 10824–10830.

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz *et al.*, 2002 The Bioperl toolkit: Perl modules for the life sciences. Genome Res. **12:** 1611–1618.

Tower, J., G. H. Karpen, N. Craig and A. C. Spradling, 1993 Preferential transposition of Drosophila *P* elements to nearby chromosomal sites. Genetics **133:** 347–359.

Townley, D. J., B. J. Avery, B. Rosen and W. C. Skarnes, 1997 Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. Genome Res. **7:** 293–298.

van Luenen, H. G., and R. H. Plasterk, 1994 Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans. Nucleic Acids Res. **22:** 262–269.

van Luenen, H. G., S. D. Colloms and R. H. Plasterk, 1993 Mobilization of quiet, endogenous Tc3 transposons of Caenorhabditis elegans by forced expression of Tc3 transposase. EMBO J. **12:** 2513–2520.

Vigdal, T. J., C. D. Kaufman, Z. Izsvák, D. F. Voytas and Z. Ivics, 2002 Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. J. Mol. Biol. **323:** 441–452.

Yant, S. R., L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak *et al.*, 2000 Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat. Genet. **25:** 35–41.

Communicating editor: C. Kozak