Transgenic Research **4**, 277–287 (1995)

A new strategy of gene trapping in ES cells using 3′RACE

MICHIO YOSHIDA[1,2,3], TAKESHI YAGI[1]‡, YASUHIDE FURUTA[1,2], KENJI TAKAYANAGI[4], RYO KOMINAMI[4], NAOKI TAKEDA[1,2], TOMOYUKI TOKUNAGA[1]§, JOE CHIBA[3], YOJI IKAWA[1,5] and SHINICHI AIZAWA[1,2]*

[1]*Laboratory of Molecular Oncology, Tsukuba Life Science Center, Institute of Physical and Chemical Research, RIKEN, 3-1-1 Koyadai, Tsukuba, Ibaraki-305, Japan*
[2]*Department of Morphogenesis, Institute of Molecular Embryology and Genetics (IMEG), Kumamoto University School of Medicine, 2-2-1 Honjo, Kumamoto-860, Japan*
[3]*Department of Biological Science and Technology, Science University of Tokyo, 2641 Yamazaki, Noda, Chiba-278, Japan*
[4]*First Laboratory of Biochemistry, Niigata University School of Medicine, 1-757 Asahimachidori, Niigata-951, Japan*
[5]*Department of Biochemistry, Tokyo Medical and Dental University School of Medicine, Tokyo 113, Japan*

Received 29 July 1994; revised 18 October 1994; accepted 28 October 1994

"Gene trapping" in embryonic stem (ES) cells is a novel approach to identify a series of genes in mammals concomitant with the production of the corresponding mutant mice. However, this approach is currently unable to identify genes that are not expressed in ES cells. Here we describe a strategy to identify gene trapping clones which is not based on expression of a reporter gene. It uses the *neo*^{r} gene which lacks a polyadenylation signal and has a splice donor signal. Expression of the *neo*^{r} gene as fusion transcripts with the 3′ end containing the polyadenylation signal of tagged genes allows the identification of these clones by 3′ rapid amplification of the cDNA end in undifferentiated ES cells, even if the genes are not expressed in ES cells. Amplification was observed in about 25% of G418-resistant clones. Sequence analyses suggested the amplifications represent gene trapping events. The feasibility of this approach was further assessed by analysing one clone, PAT-12, in detail.

Keywords: gene trapping; ES cells; 3′RACE; mutant mice

Introduction

Gene targeting technology with mouse embryonic stem (ES) cells has opened a new era in reverse genetics in mammals (Bradley *et al.*, 1984; Mansour *et al.*, 1988). Along with transgenesis by injection of DNAs into zygotes (Brinster *et al.*, 1981), the genetic analysis of functions of a cloned gene can now be done basically at will in mammals. However, further efforts are required in conventional genetics in mammals to identify novel genes, particularly from their mutant phenotypes (Rossant and Joyner, 1989; Rossant and Hopkins, 1993). The experimental production of mutant animals, either by classical mutagenesis or by insertional mutagenesis of transgenes or retrovirus (Soriano *et al.*, 1987; Woychik *et al.*, 1990) has been hampered by difficulty in breeding mice: the frequency of gene mutations is low and mutants among subjects cannot be identified in advance. Only about 5% of retrovirus and 10% of transgene insertions

*To whom correspondence should be addressed at: Department of Morphogenesis, Institute of Molecular Embryology and Genetics (IMEG), Kumamoto University School of Medicine, 2-2-1 Honjo, Kumamoto-860, Japan (Fax: 81-96-366-4287).

‡Present address: Department of Neurobiology and Behavioral Genetics, National Institute for Physiological Sciences, Myodaiji, Okazaki 444, Japan.

§Present address: Laboratory of Reproductive Biochemistry, National Institute of Animal Industry, 2 Ikenodai, Kukizaki, Inashiki-gun, Ibaraki-305, Japan.

have been found to cause recessive phenotypes, according to laborious breeding studies (Jaenisch, 1988). The selection of random mutants for particular biological phenomena has been impossible. A method that can allow identification and screening or selection for mutations *in vitro* is needed urgently, and the "gene trapping" approach with mouse embryonic stem (ES) cells has the potential to answer this need (Gossler *et al.*, 1989; Friedrich and Soriano, 1991; Skarnes *et al.*, 1992). The most elegant current approach is the use of a vector containing a β-*geo* gene that lacks any transcriptional unit: β-*geo* is the fusion gene between β-galactosidase (β-gal) and neomycin phosphotransferase (*neo*^{r}) genes, the former serving as a marker of expression and the latter as a marker of drug resistance. The random integration of the vector gives mutant ES cells after G418 selection when the β-*geo* gene is inserted into genes expressed in undifferentiated ES cells. The expression patterns of the tagged genes during embryogenesis and maturation can be visualized with a chromogenic substrate X-gal (5-bromo-4-chloro-3-indoyl-β-D(−)-galactopyranoside) by staining chimaeras developing after injection of the ES clones into host embryos. The mutant mice can be produced from chimaeras to examine the phenotypes created by the insertions, and the tagged genes can be identified with sequences in the vector as a marker. The gene trapping approach in ES cells encounters difficulty, however, when identification is desired of those genes not expressed in undifferentiated ES cells. Screening of β-gal expression in differentiated cells either *in vitro* or subcutaneously in nude mice is not only time- and labour-consuming, but the repertoire of differentiated cells is limited.

This paper describes a strategy to identify gene trapping clones which is not based on expression of the reporter gene. The strategy is simply based on the use of the *neo*^{r} gene with a promoter active in ES cells but lacking a polyadenylation signal (pA); the splicing donor signal is placed behind the stop codon. Efficient expression, necessary to attain resistance on selection with the antibiotic G418 requires pA acquisition at the integration sites, thus enriching the frequency of gene trapping events among G418-resistant random integrations. More important is that expression of the *neo*^{r} gene as fusion transcripts with the 3′ end containing pA of the tagged genes allows the identification of these clones by rapid amplification of cDNA 3′ end (3′RACE; Frohman *et al.*, 1988) in undifferentiated ES cells, even if they are not expressed in ES cells. The strategy also greatly reduces the effort involved in characterizing the tagged genes at the molecular level. Here, we observed amplification in about 25% of the G418-resistant clones. Sequence analyses suggested the amplifications represent gene trapping events. The feasibility of this approach was validated by detailed analysis of the clone PAT-12.

Materials and methods

Gene-trapping vector

The *Xba* I–*Bam* HI 800 bp fragment of the *fyn* gene that contains a part of intron 1, splicing acceptor and a part of exon 2 was isolated from pGFYN3.0Bm (Yagi *et al.*, 1990), and joined to the *Nco* I–*Alu* I 36 bp fragment of chicken β-actin gene containing the translational initiation codon, by blunt end ligation after filling up the *Bam* HI and *Nco* I ends. The fragment was then fused to the *Nco* I–*Bam* HI 3.1 kb fragment of *lac*Z from pactβgal (Maekawa *et al.*, 1991) that has the SV40 polyA signal, by blunt end ligation after filling up the *Nco* I end: this yielded the *Xba* I–*Bam* HI *lac*Z segment. The sequences around the translational initiation codon ATG were as shown in Fig. 1.

The *Eco* RI–*Bam* HI fragment that contains the *neo*^{r} gene with *pgk-1* gene promoter was derived from pKJ2 (kindly provided by Dr Michael W. McBurney, University of Ottawa, Ottawa, Canada). The *Ssp* I–*Sal* I 1.7 kb fragment that contains a part of exon 3 (*Ssp* I site locates in exon 3), splicing donor and a part of intron 3 of the *fyn* gene was isolated from pGFYN4.7 (Yagi *et al.*, 1990), and was ligated to the *Bam* HI site behind the stop codon of the *neo*^{r} gene by blunt end ligation after filling up the *Bam* HI end; this yielded the *Eco* RI–*Sal* I *neo*^{r} segment.

The *Xba* I–*Bam* HI *lac*Z segment was placed between the *Xba* I and *Not* I sites of pBluescriptII SK(+) after ligating *Not* I linker to the *Bam* HI end, and the *Eco* RI–*Sal* I *neo*^{r} segment at the *Xho* I site after ligating *Xho* I linker to the *Eco* RI end, thus yielding the gene trapping vector pPAT.

ES cell culture and introduction of the vector

The ES cells used in the present study were TT2 derived from an F_1 embryo between C57BL/6 and CBA mice; their culture and introduction of the gene-trapping vector into these cells with G418 (200 μg ml^{-1}) were done as described (Yagi *et al.*, 1993a).

Generation of chimaeric and mutant mice

Chimaeric mice were generated as described (Yagi *et al.*, 1993a). Heterozygous F_1 progenies were obtained by mating male chimaeras with C57BL/6 females. Homozygous mutant offspring were generated by intercrossing heterozygotes. The genotype was determined by Southern blot analysis, and histological examinations were performed by fixing tissues or embryos with 4% formaldehyde in PBS, embedding them in paraffin, sectioning the paraffin blocks at 5 μm thickness and staining the sections with haematoxylin and eosin.

X-gal staining

To determine the β-Gal expression in undifferentiated cells, the ES clones or colonies cultured on dishes were

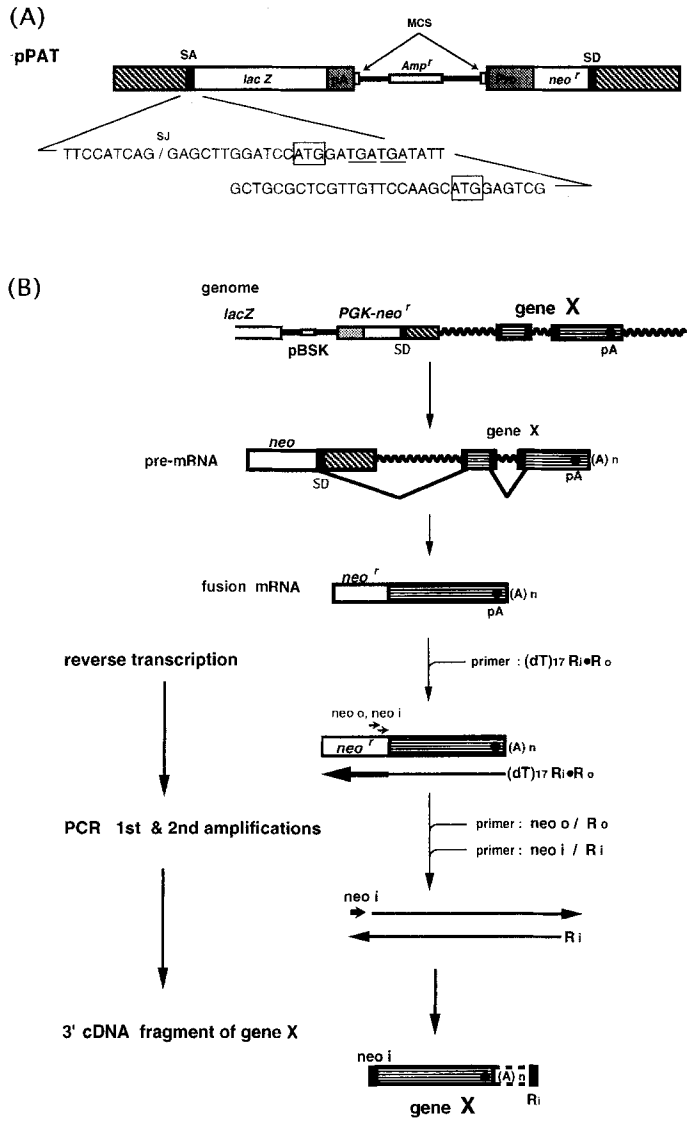

Fig. 1. Strategy for identification of gene-trapping clones. (A) Schematic representation of the gene-trapping vector pPAT. Cross-hatched boxes indicate the introns from the *fyn* gene; the box designated as SA: splicing acceptor sequences from the *fyn* gene; the box designated as pA: polyadenylation sequences from the SV40 large T gene; boxes designated as MCS: multiple cloning sites from pBluescriptII; the box designated as Pro: promoter sequences from phosphoglycerate kinase-1 (PGK) gene; the box designated as SD: splicing donor sequences from the *fyn* gene. The solid line with the insert of a small box designated as Amp^r represents the sequences from pBluescript-II. The nucleotide sequence around the splice junction (SJ) and the 5′ end of the *lacZ* gene is shown underneath; two ATG codons in open boxes are potential translational initiation codons for the *lacZ* gene, and stop codons underlined are placed out-of-frame. See Materials and methods for details. (B) The strategy for identification of gene-trapping clones by 3′RACE. Among the random recombinants of the gene-trapping vector, pPAT, after its introduction by electroporation into ES cells, the ES clones in which the vector is inserted into genes (gene X) will express the 3′ ends of the tagged genes as the fusion transcripts with the neo^r gene in ES cells directed by *pgk-1* gene promoter that is active in ES cells. To detect these fusion transcripts, total RNAs are prepared from each ES clone surviving after G418 selection. Reverse transcription is performed with $(dT)_{17}R_i{\cdot}R_o$ as the primer, followed by PCR amplification first with neo_o and R_o and then with neo_i and R_i as primers, respectively.

fixed in PBS solution containing 2% formaldehyde for 10 min at room temperature. To determine the β-gal expression in cells differentiated *in vivo*, about 5×10^6 ES cells suspended in PBS by trypsinization were transplanted under the skin of ICR nu/nu mice (Charles River, Japan). The grown cell mass was sliced and fixed in PBS solution containing 2% formaldehyde, 0.25% glutaraldehyde and 0.1% NP-40 for 1 h at room temperature. To examine β-Gal expression earlier than embryonic day 12, embryos were fixed and stained in whole mount. The embryos at later stages and tissues of adult mice were fixed and sliced with a vibratome (Microslicer, Dosaka EM Co., Ltd). Fixed samples were washed several times with PBS and stained in X-gal solution (X-gal (1 mg ml^{-1}), 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$ in 1 × PBS) at 30 °C for 12 h for the cell mass in nude mice or at 37 °C overnight for the others (Sanes *et al.*, 1986; Yagi *et al.*, 1993b).

3′RACE

The 3′RACE was done basically as described by Frohman *et al.* (1988). Total RNA (5 μg) was heated at 65 °C for 3 min, cooled on ice and incubated at 42 °C for 2 h in reverse transcription solution containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 1 mM DTT, 1 mM each of dNTPs, 10 units of RNase inhibitor, 125 nM $(dT)_{17}R_i{\cdot}R_o$ primer and 200 units of reverse transcriptase (GIBCO BRL). The reaction mixture was heated at 75 °C for 15 min and diluted 200-fold with TE (10 mM Tris·HCl pH 7.5, 1 mM EDTA), and 5 μl of the dilutions was added to 45 μl of the PCR reaction mixture; the final composition was 10 mM Tris·HCl (pH 9.0), 50 mM KCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 1 μM each of neo_o (5′ GCCTGCTTGCCGAATATCATGGTGGAAAAT 3′) and R_o primer. The PCR reaction mixture was denatured at 95 °C for 7 min, cooled to 75 °C, added with 2.5 units of *Taq* DNA

polymerase (Promega), overlaid with mineral oil, annealed at 50 °C for 2 min and incubated at 42 °C for 40 min. Amplification was carried out for 30 cycles in a thermal cycle (Perkin-Elmer/Cetus). Each cycle consisted of denaturation at 94 °C for 45 s, annealing at 50 °C for 25 s and extension at 72 °C for 3 min. The cycles were followed by a 15 min final extension at 72 °C. The PCR product was diluted 20-fold with TE, and using 1 μl the second PCR amplification was performed with the inner primers neo_i (5′ GGATCCCCTTCTTGACGAGTTCTTCTGA 3′) and R_i under the same conditions. One-fifth volume of the second PCR products was analysed by electrophoresis in 1% agarose gel and by staining with ethidium bromide. The sequences of primers R_o and R_i were described (Frohman *et al.*, 1988).

The DNA fragments amplified by the 3′RACE were directly cloned into pBluescriptII SK(+) digested by *Eco* RV and tailed with dideoxythymidine (ddT) (Holton and Graham, 1990). The fragments were sequenced on both strands with the dideoxy chain termination method (Sanger *et al.*, 1977) using Sequenase (7-deaza-dGTP kit; USB).

Southern analysis

Genomic DNAs were extracted, digested with *Bgl* II or *Sst* I and hybridized with the *Bgl* II–*Xba* I fragment of *Pat-12* cDNA (probe B in Fig. 8) or *Pst* I fragment of the *neo*^{r} gene as described (Suda *et al.*, 1987).

Northern analysis

Total RNAs were extracted from tissues, embryos or ES cells by acid guanidinium thiocyanate-phenol-chloroform method (Chomczynski and Sacchi, 1987), and polyA(+) RNAs were isolated from them with Oligotex (Daiichi Chemical). PolyA(+) RNA 5 μg was electrophoresed on 1.2% agarose formaldehyde gels, transferred to nylon membrane and hybridized with the 450 bp DNA fragment cloned from the PAT-12 clone by 3′RACE, the *Bgl* II–*Xba* I fragment of *Pat-12* cDNA (probe B in Fig. 8) or the *Pst* I fragment of β-actin gene in solution containing 50% formamid, 5× SSPE, 5× Denhardt's reagent, 10% dextran sulphate and 1% SDS at 42 °C for 16 h, followed by washing in a solution of 2× SSPE and 2% SDS at 65 °C for 1.5 h.

RT-PCR

Total RNA 1 μg from kidneys and 0.1 μg of polyA(+) RNA from testes were reverse-transcribed with the primer AS1 (5′ GATTTGCAGTGTTGAATTACC 3′). The products were serially diluted 10-fold (10^{-1} – 10^{-3}), and subjected to 35 cycles of PCR in buffer containing 0.5 μM each of primer S1 (5′ AGCCAATCGTAAAAGACTGC 3′) and AS1. The PCR products were separated by electrophoresis in 1.5% agarose gel, transferred to a nitrocellulose filter and hybridized with the *Pvu* II fragment of *Pat-12* cDNA (probe A in Fig. 8).

Isolation of cDNA and sequence analysis

The cDNA library of TT2 ES cells (Horita *et al.*, 1992) was screened with the DNA fragment amplified by 3′RACE in PAT-12. Positive plaques were isolated by the standard method (Sambrook *et al.*, 1989), and the cDNA inserts were subcloned into pBluescriptII SK(+). The deletion mutants of *Pat-12* cDNA were made by unidirectional deletion method (Henikoff, 1984) and sequenced with the dideoxy chain termination method using a *Taq* dye primer cycle sequencing kit (ABI). The analyses were performed with an automatic DNA sequencer (ABI), and both strands were sequenced through the entire cDNA. The chromosomal localization was determined as described (Hayashi *et al.*, 1993).

Results

Strategy of gene-trapping

Figure 1A shows the gene-trapping vector, pPAT, used in the present study. It contained the *lac*Z gene as a marker of expression and the *neo*^{r} gene as a marker of selection. Intron 1 and the splicing acceptor of the *fyn* gene were placed in front of the *lac*Z gene so as to express *lac*Z as a fusion transcript with the 5′ end of the tagged gene; transcription would then represent the expression of the tagged gene directed by its promoter and enhancer. For a case in which the fusion occurred out-of-frame, the ATG was placed at the top of the *lac*Z gene in-frame and the stop codon was placed out-of-frame. These steps were as reported by Skarnes *et al.* (1992). The additional step in the present study was the deletion of the polyadenylation signal (pA) from the *neo*^{r} gene and placing of the splice donor (SD) signal after the stop codon of the *neo*^{r} gene; the SD used was that of exon 3 of the *fyn* gene (Yagi *et al.*, 1990). PolyA acquisition after splicing is required for the *neo*^{r} gene to be efficiently expressed, thereby enriching its integration into genes. Moreover, the 3′ ends of the mRNA of the tagged genes are expressed as fusion transcripts with the *neo*^{r} gene directed by *pgk-1* gene promoter which is active in ES cells. This allows the identification of the target gene in ES cells, even if the tagged gene is not expressed in these cells (Fig. 1B). Identification is facilitated by 3′RACE in ES cells; the reverse transcription (RT) reaction is first performed with the antisense primer to the polyA tail, $(dT)_{17}R_i{\cdot}R_o$, followed by polymerase chain reaction (PCR) first with outer primers neo_o and R_o and then with the inner primers neo_i and R_i.

First, the feasibility of the strategy was tested. Using a vector that contained the pA signal in the *neo*^{r} gene as control, the frequency of the random integrations and of gene-trapping clones among them was determined; the

frequency of the gene trapping events was manifested as the clones which expressed the marker *lacZ* gene in undifferentiated ES cells. In two independent experiments, the frequency of the G418-resistant clones decreased from 53.8 and 52.9 per 10^6 electroporated cells with the control vector to 21.8 and 17.8 per 10^6 electroporated cells with pPAT; thus there was about 2.5–3-fold decrease in the number of recombinants. Among these, 2.8% and 2.7% of the clones expressed β-gal with pPAT in contrast to 1.2% and 1.1% with the control vector. Thus, the enrichment of the gene-trapping events among G418-resistant clones was comparable with the decrease in the number of G418-resistant clones.

Forty-nine G418-resistant clones were then isolated after the introduction of pPAT and subjected to further analysis. Total RNAs were prepared from each clone to perform 3′RACE. As shown in Fig. 2, amplification was detected in 12 clones (25%). Independently, these 49 clones were analysed for β-gal expression; this was done by staining undifferentiated ES cells and the differentiated cell mass produced by transplanting the ES clones subcutaneously into nude mice with a chromogenic substrate, X-gal (referred to hereafter as "*in vitro* and *in vivo* assays" respectively). The analysis identified four clones (8%) expressing β-gal, two in undifferentiated ES cells and two in the differentiated cell mass, only two of which gave amplification by 3′RACE. Among 45 clones in which β-gal expression could not be detected, 10 clones gave the amplification (Table 1).

Table 1. Frequency of gene-trapping clones

	lacZ (+)			
	ES cells[a]	*Differentiated*[b]	lacZ (−)	*Total*
3′RACE (+)[c]	0	2	10	12 (24%)
3′RACE (−)	2	0	35	37

[a]Number of ES clones that expressed β-gal in undifferentiated ES cells.
[b]Number of clones that expressed β-gal in differentiated cell mass that were produced by transplanting ES cells sub-cutaneously in nude mice.
[c]Number of clones that gave significant amplification of DNAs by 3′RACE.

Analysis of amplified DNA by 3′RACE

To examine the nature of amplifications by 3′RACE in ten clones that did not give rise to β-gal expression in the above assays, six amplified DNAs were arbitrarily cloned and sequenced. All of the amplified DNAs contained 39 bp sequences that precisely coincided with the upstream region of splicing junction in the *neo*^{r} gene of the vector; the sequences downstream of the splicing junction were different from those in the vector (Fig. 3). Five out of six amplified DNAs had polyadenylation signals. One of the amplified sequences was that of the *lacZ* gene, suggesting tandem integration of the vector in this clone (PAT-17); the splicing acceptance occurred within the *lacZ* gene. One amplification (PAT-8.2) coincided with the long terminal repeat sequences of ETn3, the mouse early transposon (Brulet *et al.*, 1985). The sequences of the other four amplified DNAs did not match with any sequences reported in a survey with the GenBank; these gene trapping clones were PAT-5, PAT-7, PAT-8.1 and PAT-12. We assumed that these amplified DNA fragments represent 3′ sequences of the tagged genes. To confirm gene-trapping events in these clones, β-gal expression was examined in chimaeric embryos at 10 and 18 days post coitus after injecting the cells into 8-cell ICR embryos (referred to hereafter as "chimaera assay"). The

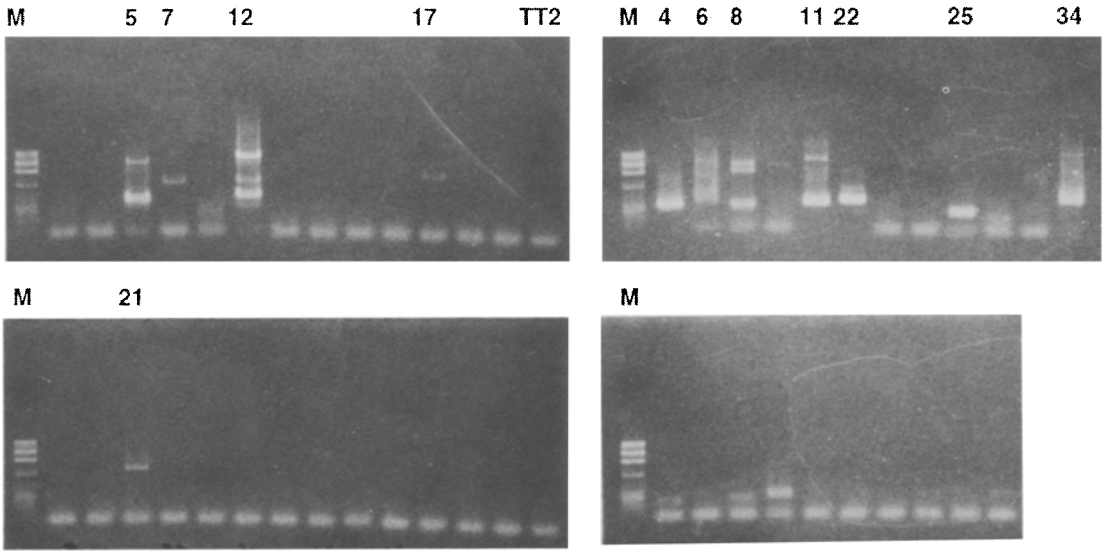

Fig. 2. Identification of gene-trapping clones by 3′RACE. The 3′RACE was performed on 49 ES clones independently isolated after G418 selection, and the 2nd PCR products were analysed by electrophoresis and EtBr staining. Significant amplification was observed in the 12 clones numbered. M, ΦX174 DNA digested with *Hae* III; TT2, parental ES cells.

Fig. 3. Sequence analysis of amplified DNAs. The sequence around the splicing junction downstream of the *neo*^{r} gene of the vector is shown above, and the sequences of amplified DNAs are shown below. The capital letters corresponding to the cross-hatched box are sequences of the third exon of the *fyn* gene in the vector (cf. Fig. 1), and the small letters corresponding to the dotted box are the sequences of the third intron of the *fyn* gene. The underlined gt indicates splicing donor site. Six amplifications by 3′RACE in Fig. 2 were arbitrarily chosen, cloned into pBluscriptII SK(+) and sequenced. The sequences matching those in the vector were dotted, and the sequences around the splice junction and at the 3′ terminus are shown along with the size of the amplified DNAs. The assignment of each gene from its entire sequence is indicated at right.

PAT-5, PAT-7 and PAT-8.1 clones yielded β-gal activity in this assay, whereas the PAT-12 clone did not. The nature of the amplification in PAT-12 clone was then further examined by isolating the tagged gene (*Pat-12*) as a clone that might contradict this approach. The details of the β-gal positive clones in the *in vitro*, *in vivo* and/or chimaera assays will be reported elsewhere.

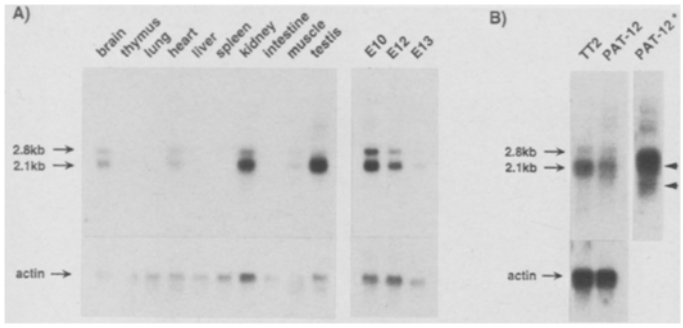

Fig. 4. Northern blot analysis of *Pat-12* expression (A) in adult tissues and embryos and (B) in parent ES (TT2) and PAT-12 ES cells. PolyA(+) RNAs were prepared from each tissue of adult mice, embryos between 10 and 13 days (E10–E13) and ES cells, and hybridized with the 450 bp DNA cloned from the PAT-12 clone by 3′RACE, except PAT-12* lane. The same blot was rehybridized with a probe for β-actin gene. Similar expression was observed with *Bgl* II–*Xba* I fragment of *Pat-12* cDNA (probe B shown in Fig. 8) as a probe. In PAT-12 ES cells, the fusion transcripts between the *neo*^{r} gene and the 3′ end of the *Pat-12* gene were also detected as shown in PAT-12* lane with a probe for *neo*^{r} gene (arrowheads).

DNA amplification in PAT-12 clone by 3′RACE represents gene trapping event

Using the amplified DNA fragment, northern analysis was first performed in adult tissues and embryos at various developmental stages to confirm that the amplified *Pat-12* DNA fragment is indeed expressed. In adult tissues it was expressed in brain, heart, kidney and testis, and two transcripts of about 2.8 kb and 1.8 kb were present in these tissues (Fig. 4A). In embryos, expression was intense in mid-gestation embryos, but it was low in embryos at later stages (Fig. 4A). Unexpectedly, it was also expressed in undifferentiated ES cells, through β-gal activity was not detected in PAT-12 ES cells; no RNA bands that hybridized with *lac*Z were detected (see below).

cDNA was then isolated from the cDNA library of the ES cells using the amplified Pat-12 DNA fragment as the probe. About 2.8 kb of cDNA was isolated. To examine whether the gene-trapping vector was indeed inserted into this locus, Southern analysis was done with this cDNA as a probe (probe B in Fig. 8), by digesting with *Bgl* II, which does not cut the vector, or with *Sst* I, which cuts the vector; the presence of restriction fragment

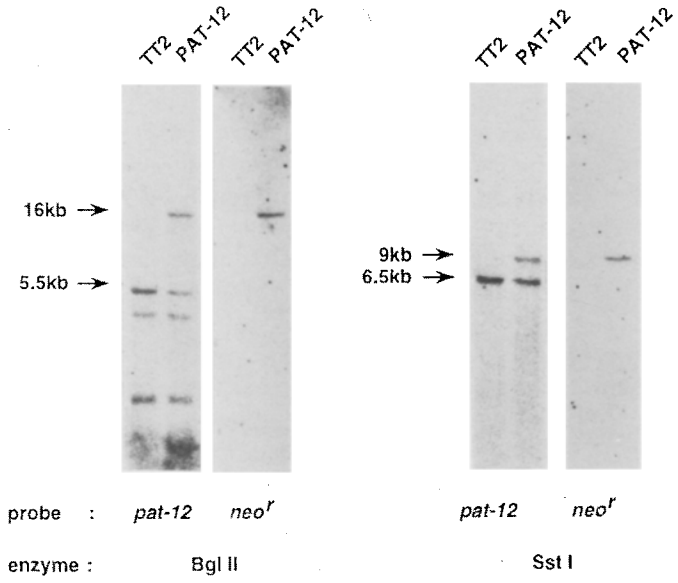

Fig. 5. Southern blot analysis on *Pat-12* locus. The genomic DNAs were prepared from the parental TT2 ES cells and PAT-12 ES cells, digested with *Bgl* II or *Sst* I and hybridized with the *Bgl* II–*Xba* I fragment of *Pat-12* cDNA (probe B shown in Fig. 8). The same blot was rehybridized with a probe for *neo*^{r} gene.

length polymorphism (RFLP) strongly suggests that the isolated cDNA represents the locus of the integration of the vector (Fig. 5). As expected, PAT-12 ES clone had one hybridizing band that was not present in the parental ES cells, and the size of this band (16 kb by *Bgl* II or 9 kb by *Sst* I) was roughly the same as the sum of endogenous band (5.5 kb or 6.5 kb) and the size of the vector (10.5 kb or 2.5 kb). In addition, the band was hybridized with the *neo*^{r} gene (Fig. 5). Furthermore, expression of this gene in the PAT-12 clone was reduced to half that of parental ES cells, as shown by northern blot analysis (Fig. 4B). All these results are consistent with the expectation that the DNA identified by 3′RACE in PAT-12 clone represents the gene at the site of integration.

Pat-12 *is an unknown gene*

Figure 6 gives the complete nucleotide sequence of the total 2781 base length. An open reading frame of 183 amino acid residues is framed by untranslated sequences of 963 nucleotides at 5′ end and 1269 nucleotides at 3′ end. A polyadenylation signal (AATAAA) and a polyadenylation-like signal (AATTAAA) are located in the 3′ untranslated region; the size between them is about 1.0 kb and roughly coincides with the difference between two transcripts (2.8 kb and 1.8 kb). The sequence around the putative translational initiation codon ATG satisfactorily matched the Kozak consensus sequence. The *Pat-12* DNA fragment amplified by 3′RACE completely coincided with the cDNA at the position from 1435 to

```
ACGAAAGGTTAAAGGAAGAACTTGCAAAAGCTCAGAGGGAATTGAAGTTAAAAGATGAAG
AATGTGAGAGACTTTCGAAAGTCCGGGATCAACTTGGGCAGGAGTTGGAAGAACTCACAG
CTAGTTTGTTTGAGGAAGCTCACAAGATGGTGAGGGAAGCGAACGTCAAGCAAGCCACTG
CAGAGAAGCAGCTGAAGGAAGCCCAGGGGAAAGTGAGTCGGGGCCTTCGCAGTGGAGCGG
GGAGACCGTGGGGAGAGCGGTCATGAGTTCCTAGCGCCCTTGGTGGCTGTCCTGACCCCT
CGGAGAAGTGTGCTAATGTGGAAGCTGCAAATAGGCACTAACTATGTAACCTTGACTCGG
GTGTGTAGTGTGTCACAGCCAGCTTGCTTATCACCATGGAAATCATGTAGAAGGATGTTG
GAGTCATTTTTAAAAGTTAATATTTATTCTCTTAATTCACTAGGGGGGAAAAAGAGACCT
AAGTAGATTTTATTATTTTTAGGAGTGAACTAAAGGTAATCGTAAGCTGATACATTTTAAT
ATTTTGAAAATTCATACTAATGTTAGTATCTTTGACCATTTAAAATTTACCTTTTGACAT
TATTGGTAGTTGAATCCAATCCATATGGAAACCCAAAAATCTTGGATTGCTTTCCCTTCA
TTACTTCATAGAAAAATAACAAATTAAACTTATTTTCTGTTCTTGAAATTCGCCAAGTTG
CTAGCATTGGCCACATCTGTCACAGCGCATGCTCACGGTGTGACGGCCACATCTGCCGTA
GAGCATGCTGTCGTGTGACGCTGCTGCTTTCTTTTCAAGATTGATGTACTTCAGGCTGAA
GTGGCCGCTCTGAAGACACTCGTACTGTCCAGTTCTCCAACATCGCCTACGCAGGAGCCT
CTGGCGGCTGCAAAGACCCCCTTTAAAAGGGGGCACACAAGGAACAAAAGTACAAGCAGC
GCCATGGGTGGGAGCCACCAGGACTTGAGTGTGATCCAGCCAATCGTAAAAGACTGCAAA
   M  G  G  S  H  Q  D  L  S  V  I  Q  P  I  V  K  D  C  K
GAGGCTGATTTATCGCTGTATAATGAGTTCAGATCTTGGAAGGATGAGCCCACGATGGAC
E  A  D  L  S  L  Y  N  E  F  R  S  W  K  D  E  P  T  M  D
AGAACGTGTCCTTTCCTGGACAAAATCTACCAGGAGGACATCTTTCCATGTTTAACATTT
R  T  C  P  F  L  D  K  I  Y  Q  E  D  I  F  P  C  L  T  F
GCAAAGAGTGAGTTGGCTTCAGCTGTTCTGGAGGCTGTGGAGAACAACACTCTAAGCATC
A  K  S  E  L  A  S  A  V  L  E  A  V  E  N  N  T  L  S  I
GAACCAGTGGGGTTGCAGCCCATCCGCTTCGTGAAGCTTCCGCTGTCGAGTGCGGAGGGC
E  P  V  G  L  Q  P  I  R  F  V  K  L  P  L  S  S  A  E  G
CAAAAAAGTGTGCGCTCACTGGCCAGAGTAACCTGCAGGCACAGGATTAGGCTCGGGGAC
Q  K  S  V  R  S  L  A  R  V  T  C  R  H  R  I  R  L  G  D
TCCAGCTGCTGTTACTACATCTCTCCTTTCTGCAGATACAGGATCACTTCTGTATGTAAC
S  S  C  C  Y  Y  I  S  P  F  C  R  Y  R  I  T  S  V  C  N
TTTTTTACATACATTCGATACATTCAGCAGGGACTCGTGAAACAGCAGGATGTCGATCAG
F  F  T  Y  I  R  Y  I  Q  Q  G  L  V  K  Q  Q  D  V  D  Q
ATGTTTTGGGAAGTTATGCAGTTGAGAAAGGAGATGTCGTTGGCAAAGCTGGGATATTTC
M  F  W  E  V  M  Q  L  R  K  E  M  S  L  A  K  L  G  Y  F
AAAGAGGAACTCTGATGCCCTGTGTGCGGGGCCGTGCGTGCTGCCCCTCCCTGGACCACT
K  E  E  L  *
GAAGGATGACGGAACATTTTTATGGTTACTCAATATTTATTTCTAAGGAGAGGCCACAGA
CTTCCCTCCTCTTGCACACTACACTGAGAAGTGCTGTGATTTCTTTTAAGCTGACTGAGC
TGTATCCCTTAGTGGAGCACACTATGAAGAATTGCAGGCGCTAGTGATGGTAATTCAACA
CTGCAAATCAAAACCACCAAATAAATATCAGACGGAGTGCGTGTGTCTTTCTTAGCTCTG
ACTCGGCATAGAGAGTTGCAGAAGACAATAGTAGTCTTCCCTTTGTTCCCCTGTGAGAGG
CAAGGAAGGAGCCAGGTGGACAGGACTCCTTCCTGTCTGTGTGTATGCTGACATCCATAC
GAAAGGTTTTATTTATTATTATTCTGTCCGCTGTAATGTAGTGAATCCCTTGATCATAAT
GGTTTGCTAAGTGCAATAAATGAAGAACAGTGCGCGTGCTTGAGCATCCTGGCTTCATGT
TAGCACTGAGCATCATCATGGAATAGAAGGAGTGTTAAGATTAATGGATGAGGGCGTGGT
CTAGACGGATGGGATCTGCCAGCTCTGAAGCTGATACTCACGGAGAGACCACCAACATCA
TTCGCACAAGTAAAATAGGTTCGTGTGCAGACATAGTGTAGGGGGGGCAAGGAAAACAGC
CTCGTTGACTCCACTATGATCACTATGCTCACGTTTCTGGAGAATAGCCTTGGACTTCAG
ATATTTCCTACGTATATATGCACAATTCTGAATCAAAATGGCCTCCTTTATTAGGTGTGT
GTGTGTGTGTGTGTGCGCGCGCTCGTGTGTGTGCGCATGCTCGTGGATGCCTCT
CACTCAGGACTGCTGCACACAGACCTGGAATTGGAGTTGCAGACCGTTGTCAGTGCTGTC
TCTCCAGCTCATCTCTCTTGCTTTGAAAGTAAATCGATGGAGCATTGCTGCCGGTAAACC
TTCACATCTCAAAAATGATTTTTCTACATGTTACGTGTTTGTGTGAGGGTGTGTCTGCGT
GTGTGGCAGTCACATGACAGCTTTTGGGAATCTGTTCTCTCCTTCTGTATGGGTTTTCGT
GATCAAACAGACTCAGGACTTCGGGGTTGGTAGCAATATCCTTATATGCGCTGAGCCATC
TCTCTGGTGCGGATCTCCACATCCTAACCCTTGTGTGTTAATTGCAATTAAATATAATTT
AATTGTCTCTAAAAAAAAAAA
```

Fig. 6. Nucleotide sequence of *Pat-12* cDNA and deduced amino acid sequence. The deduced 183 amino acid sequence is shown in single letters under the nucleotide sequence. The amplified DNA by 3′RACE corresponded to the position from 1434 to 1713 (boxed), and the gene trapping vector was concluded to be inserted prior to this sequence. Polyadenylation and polyadenylation-like signals are underlined. Primers used for chromosomal mapping are indicated by arrows.

1714 that was a part of open reading frame and the 3′ untranslated region, and the first polyadenylation signal was located in the 3′ end of this region. A search of the GenBank, however, did not reveal any genes that were

significantly homologous to this cDNA nor any structural motifs at either the level of nucleotide or amino acid sequence.

Chromosomal localization

Chromosomal location of the *Pat-12* gene was determined by a segregation analysis using 131 intersubspecific backcross mice between C57BL/6 and MSM strains; MSM is an inbred strain derived from a Japanese wild mouse, *Mus musculus molossinus* (Bonhomme and Guenet, 1989). A polymorphism was found at 3′ non-translated region containing the GT repeat (see legends of Figs 6 and 7). Two primers were synthesized and subjected to PCR and gel electrophoresis analysis. MSM gave a slower migrating band than C57BL/6 and, accordingly, each backcross mouse showed either C57BL/6 or MSM type (data not shown). The strain distribution pattern (SDP) was compared to that of 116 microsatellite loci, which had been determined using 67 published markers (Cornell *et al.*, 1991; Dietrich *et al.*, 1992) and 49 newly isolated markers (unpublished data). Figure 7 summarizes the result of typing of the backcross mice. A clear linkage of the *Pat-12* locus was found with two markers, D10Nds2 and D10Mit14 on chromosome 10. The most likely order of loci indicated by the haplotype analysis was:

$$\text{Cen}-\text{D10Nds2} - 7.6\ \text{cM} \pm 2.3 - \textit{Pat-12}$$
$$- 5.3\ \text{cM} \pm 2.0 - \text{D10Mit14} \ldots.$$

No mouse mutants have been mapped yet around the *Pat-12* locus (Lyon and Searle, 1989).

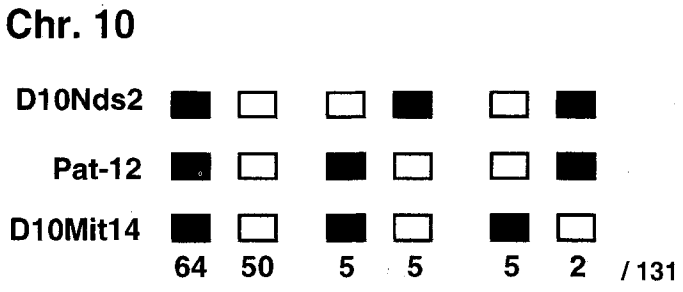

Fig. 7. Distribution of the haplotypes for 131 progeny from inter-subspecific backcross mice obtained by mating (C57BL/6 × MSM) F_1 females to MSM males. The loci followed in the cross are indicated on the left. The filled squares represent the C57BL/6 allele, and the open squares the MSM allele. Each column represents the chromosome identified in the progeny. The number of progeny carrying each type of chromosome is listed at the bottom; the *Patl2* locus is located 5.3 cM proximal to the D10Mit14 locus on chromosome 10 (LOD score: 27.6). The primer sequences used are 5′-CTGAATCAAAATGGC-CTCCTT-3′ and 5′-TGAGAGGCATCCACGAGCAT-3′, which span a 95 bp region. PCR products of D10Nds2, *Pat-12*, and D10Mit14 were analysed by acrylamide gel electrophoresis.

Mutant mice

Mutant mice were generated to further assess the nature of the integration event in PAT-12 clone. The ES cells were injected into 37 eight-cell stage ICR embryos to generate chimaeras; ten male chimaeras were obtained, seven of which were mated and produced ES derived-agouti progeny at 100% frequency. Heterozygous F_1 progenies were generated by mating male chimaeras with C57BL/6 females and identified by Southern blot analysis. The genotype analysis of offspring by the intercross among heterozygotes showed the presence of offspring which had lost the 5.5 kb endogenous band and had only the 16 kb tagged band when digested with *Bgl* II (Fig. 8A). The frequency of homozygous mutant offspring was 12 out of 49, being about 24% in line with Mendelian expectations. Homozygous mice were apparently normal and homozygous progenies were born normally from them; no abnormality was apparent histologically in highly expressing tissues.

The insertion interrupted the transcription

Since the insertion took place at the 3′ end of the cDNA, the apparent normality of the homozygous mutants raised the question whether the tagged gene was indeed disrupted by the insertion of the vector. Northern blot analysis of homozygously mutant mice indicated that no 2.8 kb or 1.8 kb transcripts were present in any tissues of homozygous mice (Fig. 8B). Expression was reduced in tissues of heterozygous mice. The transcripts in testis of homozygously mutant mice, 2.3 and 1.3 kb, were fusion products between the *neo*^{r} gene and the 3′ end of the tagged genes, since these were detected by the *neo*^{r} gene (cf. Fig. 4B) but not by a probe at the 5′ side of cDNA (data not shown); the *pgk-1* gene promoter of the *neo*^{r} gene has high transcriptional activity in this tissue (Boer *et al.*, 1990). No transcripts that hybridized with *lac*Z gene were detected in any tissues of homozygous mice (data not shown).

To further confirm the absence of transcripts in homozygously mutant mice, reverse transcription polymerase chain reaction (RT-PCR) was performed in kidney and testis (Fig. 8). When the PCR was performed with primers S-1 and AS-1, which span the insertion site of the vector, RNAs from wild-type mice yielded a 720 bp DNA product, while RNAs from homozygous mutants did not yield this amplification product. Similarly, when the PCR was performed with primers S-1 and AS-2 (which correspond to upstream of the integration site), a 380 bp amplification product was not detected in RNA from homozygous mutants; expression dropped to about half in heterozygotes. These results indicate that the insertion of the vector somehow disrupted the expression of the tagged gene, *Pat-12*, and consequently that of the *lac*Z gene.

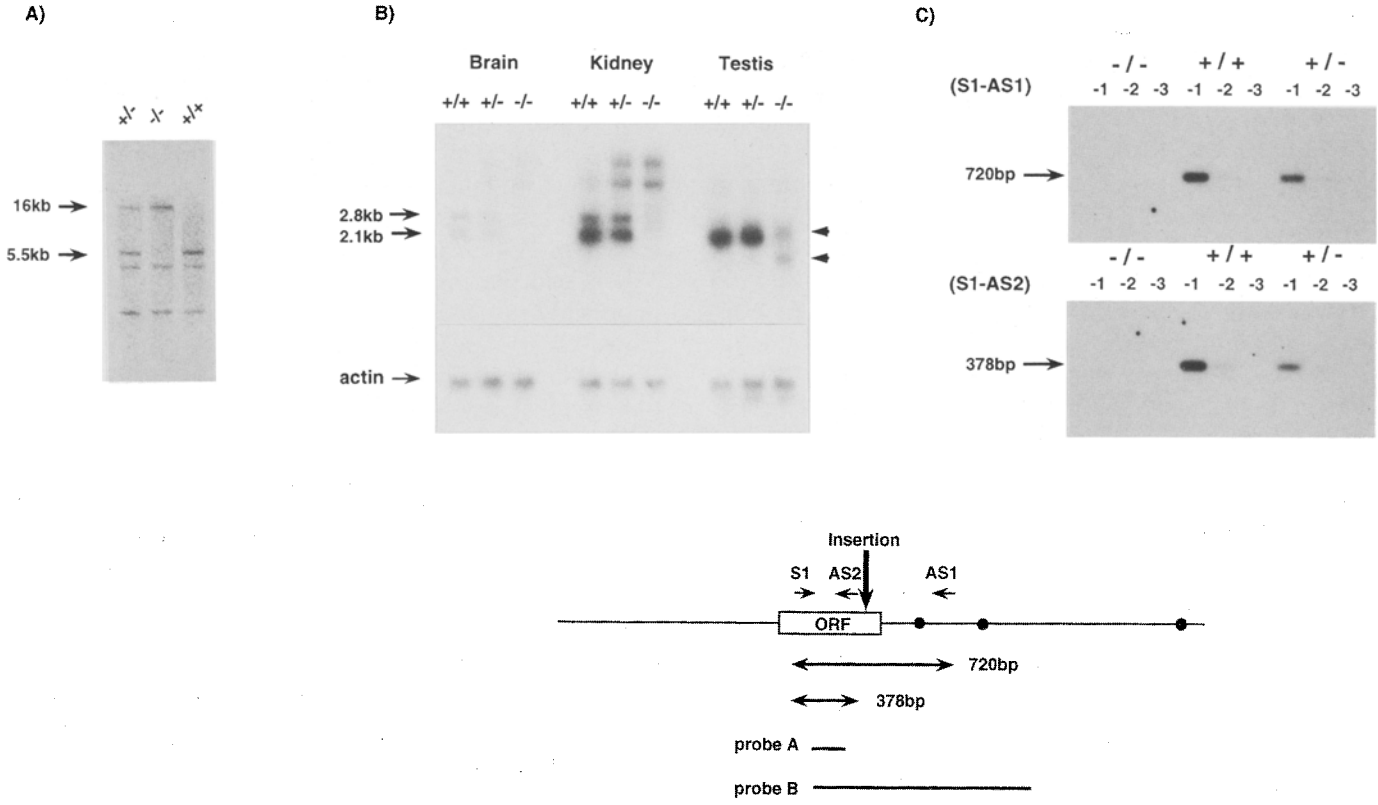

Fig. 8. Analysis of *Pat-12* mutant mice. (A) Genomic Southern blot analysis of wild type, heterozygous and homozygously mutant offspring. Genomic DNAs were digested with *Bgl* II and hybridized with the probe B. (B) Northern blot analysis of *Pat-12* expression in adult mutant offspring. PolyA(+) RNAs were prepared from brains, kidneys and testes that expressed *Pat-12* at high levels in wild type mice (cf Fig. 4A), and hybridized with the probe B. The same blot was rehybridized with a probe for β-actin gene. The transcripts in testes of homozygously mutant mice indicated by arrowheads are the fusion transcripts between the *neo*^{r} gene and the 3′ end of *Pat-12* gene since this was hybridized with a probe for *neo*^{r} gene (cf. Fig. 4B). High molecular weight transcripts in kidneys remain to be clarified. (C) The RT-PCR analysis of *Pat-12* expression in mutant mice. The reverse-transcribed products were made from RNAs of kidneys and testes. The products were diluted in one-fold (−1), 10-fold (−2) and 100-fold (−3), and amplified with AS1 and S1 or AS2 and S1 as primers by PCR. The amplified products were hybridized with probe A. The primers and hybridization probes used are indicated below.

Discussion

The final goal in the development of gene trapping technology is to identify a series of genes with functions associated with particular developmental or biological phenomena. Those genes that are expressed in ES cells and may play roles in early development can be identified by β-gal expression in ES cells (Gossler *et al.*, 1989; Friedrich and Soriano, 1991; Skarnes *et al.*, 1992). Genes that are up-regulated or down-regulated in response to the signal of ES cell differentiation *in vitro*, such as those affected by retinoic acid, can also be identified. However, it has been only accidentally and rarely that a gene that is expressed at a site and at a stage of particular interest has been identified. Recent advances in the chimaerizing potential of ES cells have made possible the efficient production of chimaeras that are composed almost exclusively of ES-derived cells, with only a minimal contribution of host embryo-derived cells (Nagy *et al.*, 1993; Yagi *et al.*, 1993a). This allows screening of the gene-trapping clones directly for β-gal expression in developing chimaeric embryos at a particular site or stage of interest. A prerequisite for this approach is the ability to identify the gene trapping clones among random recombinants. The majority of random integrations are considered to be those outside genes. The development of negative selection for such clones would be elegant, but such selection is not available. An alternative is the development of a convenient method to identify the ES clones in which genes are trapped by the vector, and this is what we have sought here.

The strategy described here allows broader identification of gene trapping clones than the current approach by expression of the reporter gene. The clones showing PCR amplification will be directly assessed for β-gal expression at a particular stage of interest in chimaeric embryos or mice. At the same time, the fragments amplified by 3′RACE can be used to characterize the β-gal expression by northern and *in situ* hybridization analysis in normal

embryos or mice. Among 49 clones examined, the number of gene trapping clones identified by β-gal expression was two in the *in vitro* assay and two in the *in vivo* assay; this number was increased to 12 using the present approach.

The validity of the method depends on the extent of spurious results obtained with the PCR amplification. Two of the above 12 clones showed β-gal expression by *in vitro* and *in vivo* assays. On six of ten clones that did not have β-gal activity in these assays, amplified fragments were sequenced, and two were spurious: integration into transposon-like elements and tandem integration. These are inevitable as a minor fraction of the integration events not only by the present method, but also by other methods; with the present method, these clones can be excluded from further analysis by sequencing amplified fragments. Sequence analyses suggested four to represent gene trapping events, and three of these indeed showed β-gal expression in chimaera assay. The present approach has the potential to identify the gene trapping clones that do not express β-gal activity either because of: (1) integration of *lacZ* out of the frame with the tagged gene, (2) disruption of transcription or translation, (3) destabilization of RNA or protein product, or (4) structural effects on β-gal activity in fusion proteins. The PAT-12 clone was one such example. These results suggest the potential usefulness of the present approach, but obviously study on a larger scale is required to demonstrate unequivocally the validity of the present approach.

The critical question is that the present work selects for integrations taking place in the 3′ regions of the genes trapped; many of these may be not mutants, though PAT-12 was. Only two were identified by the present approach among four clones expressing β-gal by *in vitro* and *in vivo* assays. This is possibly due to our technical problems in the size of the RT-PCR products; integration which occurred at the 5′ end of the gene may be excluded. The average size of mRNAs is about 2–3 kb and that of untranslated regions is about 300 bp. Recent progress in the technology, however, has made it possible to produce PCR targets up to 10 kb in length efficiently (Nielson *et al.*, 1994).

The PAT-12 clone was analysed here as a potentially spurious clone that might contradict the present approach. In this clone, however, the vector was integrated at the 3′ end of a novel gene that maps to a distal part of chromosome 10 and is expressed during midgestation embryogenesis and in several adult tissues. The gene was expressed in undifferentiated ES cells, though the PAT-12 ES cells did not express the β-gal activity. This was attributed to the absence of *lacZ* transcripts. Mutant mice were generated, but no phenotype resulting from the mutation was apparent. The integration at the end of the gene raised the question whether the integration did not disrupt the gene. However, no transcripts were found in homozygous mutant mice, possibly due to inhibition of transcription or unstabilization of mRNAs by the insertion of the vector. The gene may be a member of a novel family of genes that play important roles in overlap, as suggested in mutant mice produced by gene-targeting (Soriano *et al.*, 1991; Saga *et al.*, 1992; Shull *et al.*, 1992; Erickson, 1993; Nada *et al.*, 1993).

It is said that mammals encode about 10^5 genes; it is impossible to make a mutant mice library of this order. However, it is possible to make the library as a mutant ES cell library; recent increases in the germline differentiating efficiency of ES cells has made this practical (McMahon and Bradley, 1990; Li *et al.*, 1992; Nagy *et al.*, 1993; Yagi *et al.*, 1993a). The collection of gene trapping clones with sequence information at the 3′ end of cDNA will greatly benefit mammalian genetics. Identification of genes by a variety of methods including the cDNA and genome project can be linked directly to mutant mice through this library, instead of making the mice one by one, using gene targeting.

Acknowledgement

This work was supported in part by grants-in-aid from the Science and Technology Agency and the Ministry of Education, Science and Culture of Japan.

References

Boer, P.H., Potten, H.C., Adra, H., Jardine, K., Mullhofer, G. and McBurney, M.W. (1990) Polymorphisms in the coding and noncoding regions of murine *Pgk-1* alleles. *Biochem. Genet.* **28**, 299–308.

Bonhomme, F. and Guenet, J.L. (1989) The wild house mouse and its relatives. In Lyon M.F. and Searle, A.G. eds, *Genetic Variants and Strains of the Laboratory Mouse* p. 658. New York: Oxford University Press.

Bradley, A., Evans, M., Kaufman, M.H. and Robertson, E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255–6.

Brinster, R.L., Chen, H.Y., Trumbauer, M., Senear, A.W., Warren, R. and Palmiter, R.D. (1981) Somatic expression of herpes thymidine kinase. *Cell* **27**, 223–31.

Brulet, P., Condamine, H. and Jacob, F. (1985) Spatial distribution of transcripts of the long repeated ET_n sequence during early mouse embryogenesis. *Proc. Natl Acad. Sci. USA* **82**, 2054–8.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–9.

Cornell, R.J., Aitman, T.J., Hearne, C.M. and Todd, J.A. (1991) The generation of a library of PCR-analysed microsatellite variants for genetic mapping of the mouse genome. *Genomics* **10**, 874–81.

Dietrich, W., Katz, H., Lincoln, S.E., Shin, H.S., Friedman,

J., Dracopoli, N. and Lander, E.S. (1992) A genetic map of the mouse suitable for typing intraspesific crosses. *Genetics* **131**, 423–47.

Erickson, H.P. (1993) Gene knockout of c-*src*, transforming growth factor β1, and tenascin suggest superflous, nonfunctional expression of proteins. *J. Cell Biol.* **120**, 1079–81.

Friedrich, G. and Soriano, P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–23.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA* **85**, 8993–9002.

Gossler, A., Joyner, A.L., Rossant, J. and Skarnes, W.C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–5.

Hayashi, T., Ohtsuka, H., Kuwabara, K., Mafune, Y., Miyashita, N., Moriwaki, K., Takahashi, Y. and Kominami, R. (1993) A variant family of mouse minor satellite located on the centromeric region of chromosome 2. *Genomics* **17**, 490–2.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–9.

Holton, T.A. and Graham, M.W. (1990) A simple and efficient method for direct cloning of PCR product using ddT-tailed vectors. *Nucl. Acids Res.* **19**, 1156.

Horita, K., Yagi, T., Kohmura, N., Tomooka, Y., Ikawa, Y. and Aizawa, S. (1992) A novel tyrosine kinase, hyk, expressed in murine embryonic stem cells. *Biochem. Biophys. Res. Comm.* **189**, 1747–53.

Jaenisch, R. (1988) Transgenic animals. *Science* **240**, 1468–74.

Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–26.

Lyon, M.F. and Searle, A.G. (1989) *Genetic Variants and Strains of the Laboratory Mouse*. 2nd edition. New York: Oxford University Press.

Maekawa, T., Matsuda, S., Fujisawa, J., Yoshida, M. and Ishii, S. (1991) Cyclic AMP response element-binding protein, CRE-BP1, mediates the E1A-induced but not the Tax-induced transactivation. *Oncogene* **6**, 627–32.

Mansour, S.L., Thomas, K.R. and Cappechi, M.R. (1988) Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–52.

McMahon, A.P. and Bradley, A. (1990) The *Wnt-1* (*int-1*) proto-oncogene is required for development of large region of the mouse brain. *Cell* **62**, 1073–85.

Nada, S., Yagi, T., Takeda, H., Tokunaga, T., Nakagawa, H., Ikawa, Y., Okada, M. and Aizawa, S. (1993) Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell* **73**, 1125–35.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J.C. (1993) Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8.

Nielson, K.B., Schoettlin, W., Bauer, J.C. and Mathur, E. (1994) *Taq* extender™ PCR additive for improved length, yield and reliability of PCR products. *Strategies* **7**, 27.

Rossant, J. and Joyner, A.L. (1989) Towards a molecular genetic analysis of mammalian development. *Trends Genet.* **5**, 277–83.

Rossant, J. and Hopkins, N. (1992) Of fin and fur: mutational analysis of vertebrate embryonic development. *Genes Dev.* **6**, 1–13.

Saga, Y., Yagi, T., Ikawa, Y., Sakakura, T. and Aizawa, S. (1992) Mice develop normally without tenascin. *Genes Dev.* **6**, 1821–31.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, 2nd ed., Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sanger, F., Niklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–7.

Sanes, J.R., Rubenstein, J.L.R. and Nicolas, J.F. (1986) Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. *EMBO J.* **5**, 3133–42.

Shull, M.M., Ormsby, I., Kier, A.B., Pawlowski, S., Diebold, R.J., Yin, M., Allen, R., Sidman, C., Proetzel, G., Calvin, D., Annunziata, N. and Doetschman, T. (1992) Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature* **359**, 693–9.

Skarnes, W.C., Auerbach, B.A. and Joyner, A.L. (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**, 903–18.

Soriano, P., Gridley, T. and Jaenisch, R. (1987) Retroviruses and insertional mutagenesis: proviral integration at the Mov34 locus leads to early embryonic death. *Genes Dev.* **1**, 366–75.

Soriano, P., Montgomery, C., Geske, R. and Bradley, A. (1991) Targeted disruption of the c-*src* proto-oncogene leads to osteopetrosis in mice. *Cell* **64**, 693–702.

Suda, Y., Aizawa, S., Hirai, S., Inoue, T., Furuta, Y., Suzuki, M., Horihashi, S. and Ikawa, Y. (1987) Driven by the same Ig enhancer and SV40 T promoter *ras* induced lung adenomatous tumors, *myc* induced pre-B cell lymphomas and SV40 large T a variety of tumors in transgenic mice. *EMBO J.* **6**, 4055–65.

Yagi, T., Ikawa, Y., Yoshida, K., Shigetani, Y., Takeda, N., Mabuchi, I., Yamamoto, T. and Aizawa, S. (1990) Homologous recombination at c-*fyn* locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. *Proc. Natl Acad. Sci. USA* **87**, 9918–22.

Yagi, T., Tokunaga, T., Fruta, Y., Nada, S., Yoshida, M., Tsukada, T., Saga, Y., Takeda, N., Ikawa, Y. and Aizawa, S. (1993a) A novel ES cell line, TT2, with high germline-differentiating potency. *Anal. Biochem.* **214**, 70–6.

Yagi, T., Shigetani, Y., Okado, N., Tokunaga, T., Ikawa, Y. and Aizawa, S. (1993b) Regional localization of Fyn in adult brain; studies with mice in which *fyn* gene was replaced by *lacZ*. *Oncogene* **8**, 3343–51.

Woychik, R.P., Maas, R.L., Zeller, R., Vogt, T.F. and Leder, P. (1990) "Formins": proteins deduced from the alternative transcripts of the *limb deformity* gene. *Nature* **346**, 850–3.