An Efficient Gene-Trap Method Using Poly A Trap Vectors and Characterization of Gene-Trap Events[1]

Hitoshi Niwa, Kimi Araki,[2] Shigemi Kimura, Shin-ichi Taniguchi,[3] Shoji Wakasugi, and Ken-ichi Yamamura

Department of Developmental Genetics, Institute of Molecular Embryology and Genetics, Kumamoto University School of Medicine, Kumamoto, Kumamoto 862

New trap vectors (U1 and U2) have been developed to trap genes in murine embryonic stem (ES) cells. The polyA addition signal of the neomycin phosphotransferase II (*neo*) gene was removed from these vectors so that they needed to trap an endogenous polyA signal for expression of the *neo* gene. The frequency of gene-trap events of these vectors was about five times higher than with the vector containing the polyA signal, and only one copy of the trap vector was integrated in most cases. Four out of five 5′-flanking regions of the integrated vector in ES cell lines were found to be novel endogenous promoters, suggesting that this method is efficient for trapping genes in ES cells. In two cases analyzed, large deletions or rearrangements spanning more than 10 kb were found in the 3′-flanking region of the trap vector introduced by electroporation. This result suggests that phenotypes observed in homozygotes with a mutated allele could be due to the disruption of a gene adjacent to the trapped gene, but not of the trapped gene.

One strategy for monitoring transcriptionally active regions of a genome involves use of an enhancer trap, based on the fact that transcription of a reporter gene containing a minimum promoter is activated by cellular enhancers. In *Drosophila*, the enhancer trap strategy has been used successfully in large-scale screening of developmentally regulated genes (*1, 2*). The β-galactosidase (*lacZ*) reporter gene provided a sensitive and easily assayable gene product to detect expression in whole embryos. A similar strategy was applied to mice, and transgenic mice carrying enhancer trap vectors were found to exhibit unique temporal and spatial patterns of *lacZ* expression (*3, 4*). To use this strategy effectively on a large scale in mice, *lacZ* reporter constructs were introduced into mouse embryonic stem (ES) cells. However, in mice the expression of a reporter gene is often influenced by the integration site, and an enhancer is sometimes located far from a coding region. These features make it difficult to identify and isolate the mouse endogenous gene. A second type of vector which was designed as a gene trap was developed to clone the promoter or exon sequences of the endogenous gene directly. The gene-trap vector contains a splice acceptor instead of a weak promoter in front of a *lacZ* gene (*5*). Thus, gene trap vectors are expected to generate spliced fusion transcripts between the reporter gene and the endogenous gene present at the site of integration (*6, 7*). In addition, all insertions of the gene trap vector may result in a mutation in the host

genes. Overall this gene-trap strategy in embryonic stem (ES) cells is a powerful method because it allows isolation of numerous unknown genes functioning in the early stages of development and morphogenesis, examination of the expression pattern of the endogenous gene, and the obtaining of insertional mutant mice (*5, 6, 8, 9*). This gene trap method in ES cells can thus be regarded as gene-targeting to unknown genes, as opposed to gene-targeting to known genes, which is known as homologous recombination. Previous reports have shown that the reporter constructs were integrated into endogenous loci, leading to specific patterns of expression during development and to disruption of the endogenous genes at the site of integration.

Although we can expect the 5′-flanking region of the trap vector to remain intact, the 3′-flanking region may be deleted or rearranged. This may make it difficult to clone the disrupted gene. Furthermore, it is difficult to deduce the connection between the disrupted gene and phenotype because a gene adjacent to the integration site may also be disrupted. Such deletions or rearrangements were not fully analyzed in the previous reports. In this study we examined the extent of deletions or rearrangements of the 3′-flanking regions of the gene-trap vectors. We also attempted to increase the frequency of trap events over random integrations and to select ES clones in which only a single gene is trapped or disrupted by the trap vector. The constructs used in our gene trap experiments consist of promoter-less *lacZ* and neomycin phosphotransferase (*neo*) expression unit without its own polyA addition signal sequence. We removed polyA addition signal sequence because the *neo* gene is expected to be expressed only when the trap vector is integrated in such a way as to utilize the polyA addition signal of the endogenous gene. The results demonstrate that a large deletion or rearrangement spanning more than 10 kb does occur during integration and that a gene-trap

Present addresses: [2] Department of Pathology, Centre Medical Universitaire, 1 Rue Michel Servet, CH1211, Geneve 4, Switzerland; and [3] First Department of Internal medicine, Tottori University School of Medicine, Yonago, Tottori 683.

vector without the polyA addition signal at the 3'-end of the *neo* gene enriches trap events in ES cells.

MATERIALS AND METHODS

Construction of Plasmids—We first constructed pAc-neoAn(−) and pAcneoAn(+) by replacing the *XhoI-MluI* sites of pMC1neo-polyA(−) and pMC1neo-polyA(+) (each purchased from Stratagene) with a 1.3-kb *XhoI-HindIII* fragment containing the chicken β-actin promoter region. Then, *BamHI* sites of the pMC1neo were converted to unique *XbaI* sites by using a *XbaI* linker. A 230-bp *PvuII-MnlI* fragment of CDM8 (*10*) containing the amber sup-pressor tRNA(*SupF*) gene was excised and the *MnlI* site was converted to blunt end by a standard procedure (*11*). An *EcoRI-AatII* fragment of pXSP64, a derivative of pSP64 (Promega) in which the *XbaI* site was removed, was excised and both sites were converted to blunt ends. Plasmid pXSP-SupF was constructed by ligating these two fragments. DNA fragments used in the gene trap were as follows. (i) A 4.0-kb *XbaI-BamHI* fragment excised from pAGS-*lacZ* (*12*) containing the rabbit β-globin splice acceptor sequence from the intron 1/exon 2 boundary, *lacZ*, and polyA addition signal sequence from SV40 largeT gene. The cohesive end of *BamHI* was converted to a flush end by standard procedures. (ii) A 2.2-kb *SmaI-BamHI* fragment from pXSP-SupF, in which the *BamHI* end was converted to a *XhoI* site by using a *XhoI* linker. (iii) A 2.0-kb *XhoI-XbaI* fragment from pAcneoAn(−) for U2 or pAcneoAn(+) for U2pA. These fragments were ligated to produce U2 and U2pA. U1 was constructed by removing the *HindIII-KpnI* fragment containing the translation start codon of *lacZ* from U2.

Cells and Electroporation—ES MS-1 cells (derived from C57BL/6) were cultured using mitomycin-treated STO feeder cells in Dulbecco's Modified Eagle's Medium (Sigma) containing 15% fetal calf serum (Sera Lab.), 0.1 μM 2-mercaptoethanol, 110 μg/ml sodium pyruvate, 4.5 mg/ml D-glucose, and 1,000 units/ml recombinant murine

leukemia inhibitory factor (LIF). LIF was produced by transient expression of murine LIF cDNA using the vector pCAGGS-LIF (*13*) in BMT-10 cells (*14*). Murine LIF cDNA was generously provided by the Genetics Institute (Boston, U.S.A.). The cells were electroporated with the linearized DNA construct at 40 μg/ml using a Bio-Rad Gene Pulser, and selected with G418 at 150 μg/ml for 10 days. In the transfection experiment to test promoter activity, undifferentiated F9 embryonal carcinoma cell lines were used. These cells were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal calf serum and transfected by the calcium phosphate method. ES MS-1 cells were induced to differentiate by incubation in the presence of 100 nM retinoic acid for 4 days. Cells were fixed in 1% glutaraldehyde and stained with X-gal as described.

Production of ES Cell Chimeras—ES-D3 cells (derived from 129) (*15*) were cultured using LS-10 cells, subclones of STO cells, as feeder cells. C57BL/6 blastocysts were injected with 10 to 15 ES cells and transferred to the uterus of pseudopregnant ICR recipients as described previously (*16*). Chimeric males were mated with C57BL/6 mice and germline transmission was confirmed by the D3 agouti coat color and the inheritance of introduced trap vector DNA.

Cloning of the Flanking Mouse Genomic DNA—Genomic DNA flanking the integrated vector from each cell line was isolated by the plasmid rescue method. Each genomic DNA was digested with appropriate restriction enzymes and ligated at 10 μg/ml to obtain circular molecules. These circularized DNAs were transformed into *E. coli* JM109 by electroporation and plated on LB/agar plates using an ampicillin drug selection for the plasmids. Rescued plas-mids were analyzed by restriction mapping and DNA blotting, using standard procedures.

3'-RACE Analysis of neo Transcription—The RACE strategy with several modifications was used to amplify cDNA sequences 3' to *neo* gene (*17*). In brief, mRNA was reverse-transcribed using a dT17-adapter primer, 5'-GAC-TCGAGTCGACATCGATTTTTTTT-TTTTTTTTT-3'; and first amplification was subsequently performed using the adapter primer, 5'-GACTCGAGTCGACATCGAT-3', and the first *neo*-specific primer, 5'-GCGTTGGCTACCCG-TGATAT-3'. Then, second amplification was performed using the adapter primer and second *neo*-specific primer, 5'-CCATCGATTCGCAGCGCATCGCCT-3'. Amplified cDNA was cloned into pBluescript KS(−) (Stratagene) for sequencing.

Nucleic Acid Hybridizations—DNA (10 μg) digested with appropriate restriction enzymes was size-fractionated on a 0.7% agarose gel in Tris/borate buffer and transferred to a nylon membrane (HybondN+; Amersham). Hybridi-zation of the blots with probes labeled by random-priming was carried out at 65°C in a mixture containing 5×SSPE, 5×Denhaldt's solution, 0.5% SDS, and heat-denatured salmon sperm DNA (100 μg/ml). Blots were washed in 2× SSC for 10 min at room temperature and then in 0.2× SSC/0.1% SDS for 50 min at 65°C, and signals were detected by use of BAS 2000 (Fuji).

RESULTS

Effects of Removal of the PolyA Addition Signal of neo— ES MS-1 cells were transfected by electroporation with
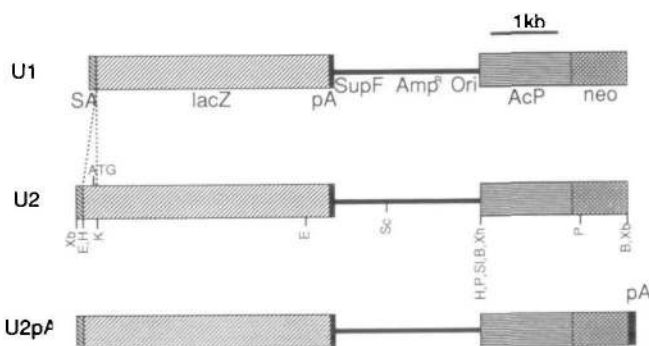

Fig. 1. **Construction of gene-trap vectors U1, U2, and U2pA.** U1 has neither the start codon of the reporter *lacZ* gene nor the polyA addition signal of the *neo* gene, U2 has only the start codon, and U3 has both. The restriction maps of each vector are shown. Abbrevia-tions: SA, splice acceptor site; *lacZ*, β-galactosidase gene; pA, polyA addition signal; SupF, t-RNA gene suppress umber mutation; Amp^R, ampicillin resistant gene; Ori, replication origin in *E. coli*; Acp, chicken β-actin promoter; *neo*, neomycin phosphotransferase II gene; ATG, initiation codon; B, *BamHI*; E, *EcoRI*; H, *HindIII*; P, *PstI*; Sc, *ScaI*; Sl, *SalI*; Xb, *XbaI*; Xh, *XhoI*. See "MATERIALS AND METHODS" for details.

gene-trap vectors U1, U2, and U2pA. To examine the effect of the absence of the polyA addition signal sequence of the *neo* gene, we introduced U2 and U2pA plasmid DNA linearized with *Xba*I into ES cells by electroporation, selected with G418 for 10 days, and stained with X-gal. Of the total G418^r colonies, 7.7% stained blue when the polyA addition signal was removed, in contrast to 1.1% when the polyA addition signal was included (Table I-1). The result indicates that trapping of the endogenous polyA addition signal contributes to the enrichment of the gene trap by the reporter *lacZ* gene in the total G418^r colonies. When the polyA addition signal is removed from the *neo* gene, *lacZ* may be expressed under the control of the actin promoter of the *neo* gene, or the *neo* gene may be expressed using the polyA addition signal of the *lacZ* gene as a fusion transcript when more than 2 copies of the trap vector are integrated in tandem array. To estimate the frequency, trap vectors digested with *Sca*I were introduced, selected and stained. Of the total G418^r colonies, 1.3% in U2 and 0.7% in U2pA stained blue (Table I-1). On the other hand, ten times as many resistant colonies were obtained by the introduction of U2 linearized by *Sca*I rather than by *Xba*I. And the introduction of U2pA linearized by *Sca*I rather than *Xba*I produced three times as many colonies. These results suggest that the *neo-lacZ* fusion RNA was transcribed to produce fusion products with *neo* function but not with *lacZ* function, and that single-copy integration occurs in much higher frequency than multi-copy integration. In the case of U2pA, the polyA addition signal of the *neo* gene was located 43 bp upstream from the *Xba*I site and was far from the *Sca*I site. Thus, the polyA addition signal could be deleted during integration into the host genome, leading to the

reduction of colonies obtained by the introduction of U2pA.

Effect of the Presence or Absence of the Initiation Codon in the lacZ Gene—The U1 and U2 linearized by *Xba*I were introduced into ES cells and selected. There were three stop codons that fit in all reading frames in the region up to 30 bp upstream of the translation start codon of *lacZ* gene in U2 (Fig. 4A). Consequently all translation from fusion transcripts are expected to terminate at these stop codons and be reinitiated from the start codon of the *lacZ* gene, because such termination–reinitiation was previously reported to occur in mammalian cells (*18*). As the 144-bp region containing these stop codons and the start codon was excised from U2 to construct U1, the translation of *lacZ* occurs only when the *lacZ* gene is fused in-frame to the endogenous gene. In the intron, there are no stop codons and two ATG codons in the same frame. But these are out-frame to the *lacZ*. Thus, the *lacZ* is expressed as the fusion product when the *lacZ* is directly fused in-frame to an endogenous gene. Of the total G418^r colonies, 16.1% in U1 and 12.5% in U2 stained blue (Table I-2). When ES-D3 cells were used, 8.3% (2/24) and 33% (14/42) of the total G418^r colonies stained blue in U1 and U2, respectively. This suggests that there was no significant difference in the efficiency of gene trapping between U1 and U2, and that any variation of efficiency in these experiments was due to differences in experimental conditions for electroporation, as reported in previous gene-targeting experiments.

Cloning of Mouse DNA Flanking the Trap Vectors—We used sixteen trap lines obtained by the above experiments using U1 and U2 in the following analyses. Southern blots analysis of DNA from each cell line digested with *Bam*HI and *Pst*I showed that the vector was inserted into single site in all sixteen lines. To determine the copy numbers integrated, PCR analysis was done using a set of primers near the 5′- and 3′-ends of linearized vectors. No amplified DNA fragment appeared in fifteen of sixteen lines suggesting that a single copy of the vector was integrated into the

TABLE I. **Expression of reporter gene in the transfected ES cells.** 1×10^7 ES cells were electroporated with 32 μg of linearized vectors, selected in 150 μg/ml of G418 and stained with X-gal after selection for 10 days.

Expt. No.	Transfected vector	Linearized site	No. of G418^R colonies	No. of X-gal + colonies
1	U2	*Xba*I	117	9 (7.7%)
		*Sca*I	1188	15 (1.3%)
	U2pA	*Xba*I	585	3 (1.1%)
		*Sca*I	1635	12 (0.7%)
2	U1	*Xba*I	68	11 (16.1%)
	U2	*Xba*I	48	6 (12.5%)

TABLE II. **Characterization of rescued 5′-flanking region.**

Clone	Vector type	Expression of *lacZ*[a] Undif.	Diff.[b]	Rescued 5′-flanking region Length (kb)	Promoter activity[c]
AYU-1	U2	+++	−	2.1	+
AYU-2	U2	++	+	1.2	+
AYU-3	U1	−	+	3.0	−
AYU-4	U1	+	++	N.D.	N.D.
AYU-5	U2	+	−	0.6	+
AYU-6	U2	+	−	0.6	+

[a]Expression of the introduced reporter *lacZ* gene was determined by staining with X-gal for 12 h. The signs indicate: −, no staining; + to +++, increasing area of staining intensity. [b]Each gene trap ES cell line was stained before and after differentiation by retinoic acid. [c]Each rescued plasmid containing the 5′-flanking genomic region and the entire *lacZ* gene was introduced into F9-tk⁻ cells. After 3 days these transformants were stained with X-gal to determine the promoter activity of the rescued region. A plus sign indicates the presence of staining cells, a minus sign the absence of staining cells.
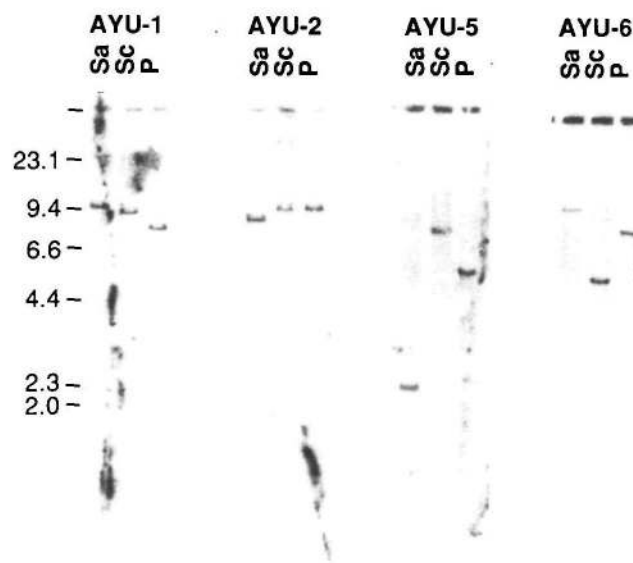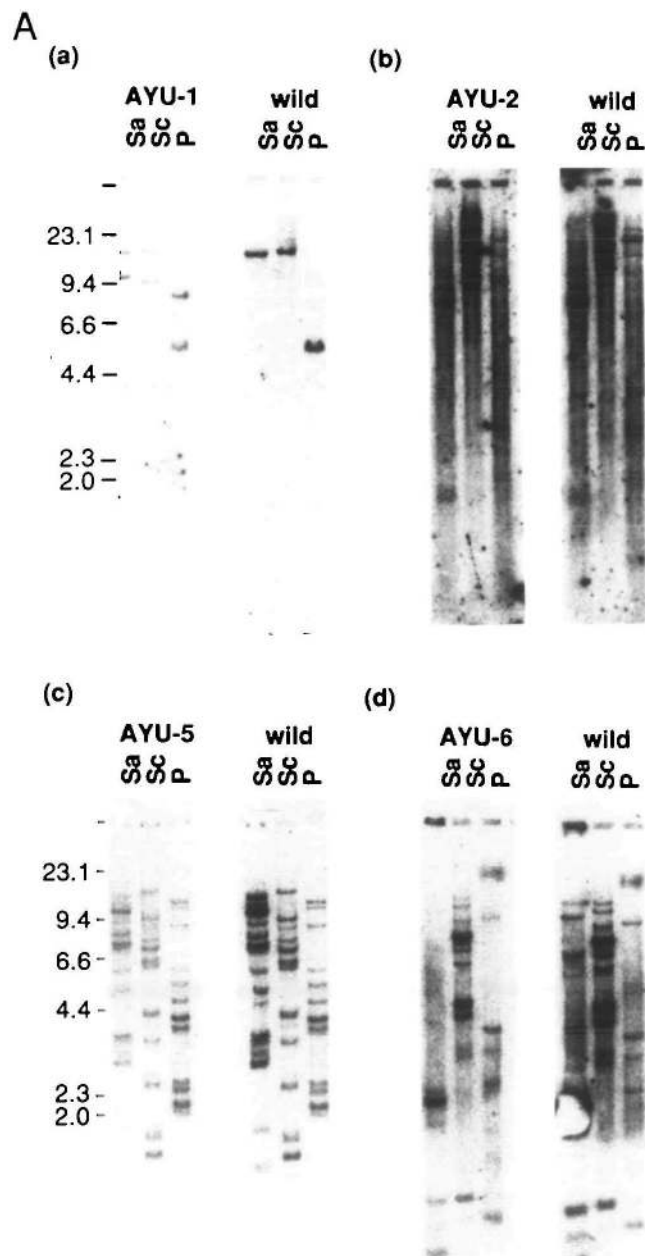

Fig. 2. **Southern blot analysis of the genome of four trapped ES cell lines.** Genomic DNAs (10 μg) from each line were digested with *Sac*I, *Sca*I, or *Pst*I, run on agarose gels, blotted, and hybridized to *lacZ* probe. A single band was detected in each case, so only a single copy of the vector was integrated into the unique site of the genome.

A

(a) (b)

(c) (d)

B

Fig. 3. (A) Southern blot analysis of the four trapped ES clones and normal ES cells. Genomic DNAs (10 μg) were digested with *Sac*I, *Sca*I, and *Pst*I, run on agarose gels, blotted, and hybridized with each rescued genomic fragment flanking the 5′-ends of the integrated vectors. With AYU-1 and AYU-2 probes, only single bands were detected in wild-type ES cells and new additional bands associated with the gene-trap insertion in trap lines. Multiple bands were detected in both wild-type and trapped ES cells with AYU-5 and AYU-6. (B) Restriction maps of the AYU-1 and AYU-2 loci in the wild and mutant alleles. All restriction sites were determined by Southern blot analysis of the genomic DNA from each trapped ES cell line and wild-type ES cells. Abbreviations: Bl, *Bln*I; Ps, *Pst*I; Sa, *Sac*I; Sc, *Sca*I; Sp, *Spe*I; Xb, *Xba*I; Xh, *Xho*I. The solid bars show the position of probes used in these analyses.

genes. We introduced plasmids rescued from five lines into undifferentiated F9 cells by the calcium phosphate method and stained these transfectants with X-gal after 3 days. Four of the five lines showed expression of reporter genes (Table II), suggesting that promoter elements were retained in rescued fragments and these were sufficient for expression in F9 cells. These four flanking regions were used to probe Southern blots of DNA from each of the four cell lines and non-transfected ES MS-1 cells. In the case of AYU-1 and AYU-2, these probes detected only one or two bands in normal ES MS-1 DNA and appropriate additional bands in each cell line DNA, suggesting that these DNA fragments were derived from cellular genes. On the other hand, the probes of AYU-5 and AYU-6 detected more than ten bands in normal DNA. This suggests that they were derived from gene families or contain repetitive sequences such as retrotransposon. These four flanking genomic DNAs were sequenced and searched for consensus promoter motifs, and in all cases typical promoter motifs were found. In the flanking region of AYU-1, a typical TATA box, an NF-1 binding motif containing inverted repeats, an Sp-1 binding site, and two typical octamer binding sites were observed. We are most interested in this clone because of the motifs in this region and the strong expression of the reporter gene in the undifferentiated state followed by dramatic reduction of expression after differentiation. We are analyzing this in more detail. The sequence of this region will be reported separately. In the AYU-2 flanking region, there are two GC boxes, GGGCGG, at only 50-bp upstream from the integration site, and the AYU-6 flanking region also contains GC box-like motifs

genome in these lines (data not shown). We rescued a genomic DNA fragment spanning about 600 to 3,000 bp flanking the 5′-end of the inserted vector by the plasmid rescue method from five cell lines, named AYU-1, AYU-2, AYU-3, AYU-5, and AYU-6, that showed expression of *lacZ* before or after induction of differentiation (Table II). Subcellular localization of X-gal staining was cytoplasmic (diffuse) in all six cases. The rescue failed in one line, AYU-4, in which many copies of the vector were integrated.

Identification of Cellular Promoters in Flanking Regions—The plasmid rescue method allows isolation of flanking regions connected to intact reporter genes, and the promoter activities of these regions can readily be estimated by re-introducing the rescued plasmid into appropriate cell lines and detecting transient expression of reporter
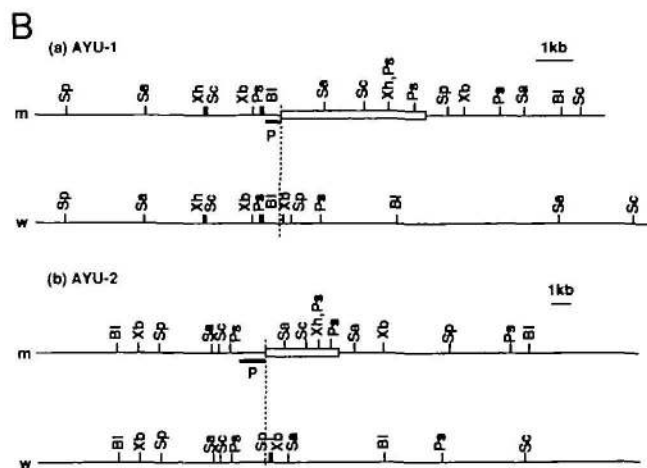
(A) β-globin/lacZ junction

ctctagagcctctgctaaccatgttcatgccttcttcttttcctacagCTCCTGGGCA

ACGTGCTGGTTGTTGTGCTGTCTCATCATTTTGGCAAAGAATTCCAAGCTTGGGATCTC
                                       HindIII

TATAATCTCGCGCAACCTATTTTCCCCTCGAACACTTTTTAAGCCGTAGATAAACAGGC

TGGGACACTTCACATGAGCGAAAAATACATCGTCACCTGGGACATGTTGCAGATCCATG

CACGTAAACTCGCAAGCCGACTGATGCCTTCTGAACAATGGAAAGGCATTATTGCCGTA

AGCCGTGGCGGTCTGGTACCGGTGGGTGAAGACCA
                  KpnI

(B) AYU-2

TTGTGCTGTGTGCTGCGCGCGATGGTTCCGTACGCAGCTCGATGGCTGATTGAGCCTAA

GCTGCGCTACTGGTAGGTTGAAAACACGTTAGACGAAACGCAGGTCTCATGCTGTGTGC

TGAAGATCTGGAGTCACCCTTCTGCAAGGCACCTGGGATTGAGCCGGCCGCTCCCTCAT

CTCGCGACGCTTCCCTGCTTACCGGCCCCACCCCCACCTTTTTTTTTTTTTCCCTACGAC

CTGCGTCTGCGTCCCACCCCCACcctctgctaaccatgttcatgccttcttcttttcct

acag

Fig. 4. **Nucleotide sequence of the gene-trap splice acceptor and the rescued endogenous promoter regions from three gene-trap cell lines.** (A) Nucleotide sequence of the β-globin/*lacZ* junction region of the U2 gene-trap vector. The intron region of β-globin containing the splice acceptor site is shown in lowercase letters and other sequences in uppercase letters. The *lacZ*-coding sequence was fused to the second exon of β-globin, 40 codons long, derived from *E. coli gpt* gene, which contains prokaryotic promoter and 28 codons from *E. coli trpS* gene. The initiation codon of *lacZ* is boxed, and three stop codons between the splice acceptor and the initiation codon are underlined. In U1, the sequence between the two arrows in the figure is deleted. Sequences of 5′-flanking genomic regions and junctions of integrated vectors rescued from the AYU-2 (B), AYU-5 (C), and AYU-6 (D) cell lines are shown. Partial sequence of the AYU-2 clone, 1.1 kb in length, and entire sequences of AYU-5 and AYU-6 are shown. The β-globin sequence of the trap vector is shown in lowercase letters and the genomic sequence in uppercase letters. Predicted TATA-box sequence is boxed and GC box-like sequences are underlined.

(C) AYU-5

CTGCAGCACACTCCCTTTTCCTTTGGTCTGTGACCTAGCTCTTGTCTGCACTGTTTTAC

CTGACAGTATACAGTAAGCTGTTCCTCTGGTTCTCATCCTGCCAAGTATATTTCCACAC

ACAAACCCCTGCCACAAAGGTTTTGTCCAGATGCTGCCAGTTAATAAATTTATTCATAG

GCTACAGCCGGGTCCAAGGAGAGACTAGGTCCACAAAGCAGCACACAGATGGCCCCCTA

TGCCTGGCTGACGCTTACCATCCATGCGCTGGAAAGACCTACCACACAGTCATTCTTTG

GTAAGAAGCTTGCTCCCCAGGCTTTATGGGAGATGACTTCGTACTTACCATGTCTGAAG

TTCTGATCTTGCCTGATTTACCCTCCCAGCACTCACTTAAGACACCTAAGAAAGCAACA

CATAGAAACACTCAGGTCACCTGAGTTAAAGAGAGGAACTCAAGGGCTTAGTACCCGGA

AGAGCCACCTCCTCTTCTGACTGGCAGCTCCGTGTGGAGTTTCTGATTGGCTACTGAGT

CTTGAATGACATgagcctctgctaaccatgttcatgccttcttcttttcctacag

(D) AYU-6

CTGCAGTCCACCTTCAGATAAATTCTACTTGGAGTGCCTTACTATCTCCTAAGCTTTAG

TAAATCCTTTACATCATGAATCCTGCACTTTTATATCACAGCTTAATATTGGATTGCTA

CCCAGCACATGAGATGTCTTTTACAACCTCCACAAAATTAACTCCAAATGGATTGTCCA

CCTTAACGTCAAATGCTAAGAACAGAACGTGTAGAAACTAGGACTCTCTCAGTTTGGGA

TCCTGACAAAGATCTTTCTACACACAGCCCAAATGCTCTGCACAAGATGCAGCCCCCCA

CAGTGAGGTTATTCAGAGGCTACAGGTGGGTCCCCTAGTGGCTAGGGCCTCAGGGGACC

TCACACGTGGCCGCCGGTTCCCAGTCAGTCCCCGGAAAGACCCACCACACCAGTCGCGC

TGCTCAGACCCCTGCTTCTCCCCATTGTTGGCAGTACCCTGTGCTCACCATGTCGCAAC

TTCGGAGCTTGCCTGAGCACAGTCCACAGAATCCAGCGGCAGGACAAGGAGCACAGCAC

ACAGGCCCATTCAAGCCTGCTCAGCGACTAATTGAAATTCGCGCCAAGGCATCCGGAAA

AACCCCGCCTTCCCTTCTGACTGGCAGGTCAGCCTAAAGGCTCTGATTGGCCAATAAGtc

tgctaaccatgttcatgccttcttcttttcctacag

near the integration site. In AYU-5, a TATA box and GC box-like motifs are found in the middle of the rescued 600-bp DNA fragment. A search of the GenBank data base (Version 18) for these four nucleotide sequences indicated that these sequences were novel. In AYU-1 and AYU-2, we also cloned 3′-flanking regions by the plasmid rescue method after digestion with *EcoRI*.

Transcription of the Rescued Sequences—In Northern blot analysis using these 5′ rescued regions as probes, the AYU-3 and AYU-5 flanking regions detected endogenous transcripts. In AYU-3, transcripts of approximately 1.8-kb were detected in almost all adult tissues; and in AYU-6, 2.5-kb transcripts were detected only in undifferentiated ES cells (data not shown). In AYU-1, transcripts of about 10-kb were detected in adult brain, duodenum, and spleen, and small 8-kb transcripts were detected in testis using the 3′-rescued region as a probe. In AYU-2, 5-kb transcripts were detected in almost all adult tissues (data not shown). The characteristic sequence of the promoter region and the ubiquitous expression of endogenous transcripts suggests that the trap vector may be integrated into a housekeeping gene in AYU-2.

Rearrangement and Deletion of Mouse Genome Near the Integration Sites—The expression of reporter gene in the trap vector means that the 5′-flanking region remains

intact. However, deletions or rearrangements in the 3′-flanking region were not examined in previous reports. This is important because phenotypic consequences could be due to the disruption of other gene(s) located in the 3′-flanking region of the trapped gene or in another region rearranged during the integration of the trap vector. We addressed this question by examining the 3′-flanking region of AYU-1 and AYU-2. In the case of AYU-1 and AYU-2, the rescued genomic DNA fragments detected single bands on genomic Southern blots in normal ES MS-1 cells and additional unique bands in each respective trap line (Fig. 3A), and the *lacZ* and *neo* probes also detected single bands in each respective trap line (Fig. 2). Thus, we made restriction maps of wild and mutated alleles near the integration sites (Fig. 3B) and compared them with each other. The results suggest that there were large deletions or rearrangements extending over more than 10 kb around the integration sites in these two cases. In contrast, deletions of only 2 to 20 bp were found in the 5′- or 3′-ends of the introduced vectors in all six cases, as revealed by sequencing of each rescued plasmid (Figs. 4 and 5).

PolyA Signal of neo Transcript—As stated earlier, the frequency of gene-trap events is higher in the trap vector without the polyA signal of the *neo* gene than in that with it. To examine whether this is due to the utilization of the
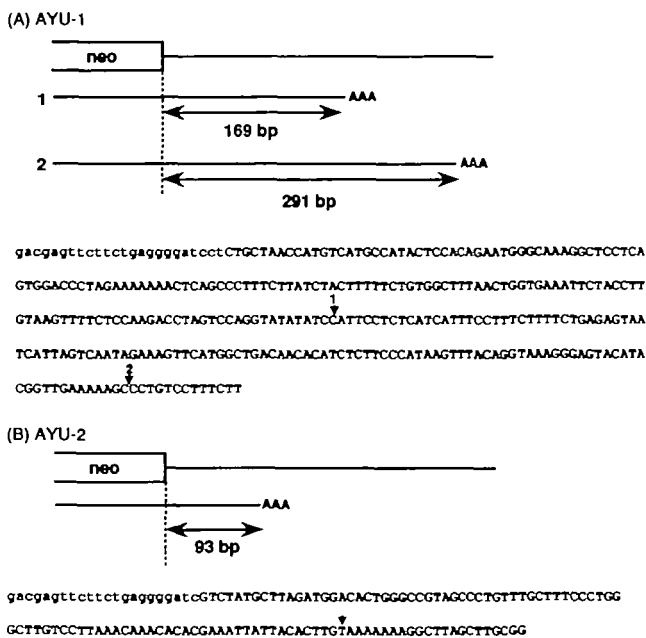
(A) AYU-1

gacgagttcttctgaggggatcctCTGCTAACCATGTCATGCCATACTCCACAGAATGGGCAAAGGCTCCTCA

GTGGACCCTAGAAAAAAACTCAGCCCTTTCTTATCTACTTTTTCTGTGGCTTTAACTGGTGAAATTCTACCTT

GTAAGTTTTCTCCAAGACCTAGTCCAGGTATATATCCATTCCTCTCATCATTTCCTTTCTTTTCTGAGAGTAA

TCATTAGTCAATAGAAAGTTCATGGCTGACAACACATCTCTTCCCATAAGTTTACAGGTAAAGGGAGTACATA

CGGTTGAAAAAGCCCTGTCCTTTCTT

(B) AYU-2

gacgagttcttctgaggggatcGTCTATGCTTAGATGGACACTGGGCCGTAGCCCTGTTTGCTTTCCCTGG

GCTTGTCCTTAAACAAACACACGAAATTATTACACTTGTAAAAAAAGGCTTAGCTTGCGG

Fig. 5. Nucleotide sequence of the fusion transcripts of the neo genes lacking their own polyA addition signal. The genomic regions downstream of the 3'-ends of the integrated vectors are cloned from AYU-1(A) and AYU-2(B) cell line and sequenced partially. The regions of the neo gene are shown in lowercase letters and the flanking genomic sequence in uppercase letters. The polyA addition sites found in cDNA of the fusion transcripts cloned by the 3'-RACE procedure are marked by a vertical arrow. In the case of AYU-1, two cDNA fragments of different length, 169 and 291 bp, were isolated. In these clones the polyA tail is added at different sites of the genomic region, suggesting that each transcript may use different genomic sequences as the polyA addition signal. No completely matched sequences to the consensus of the mammalian polyA addition signal were found in upstream regions of polyA addition sites in these cases.

polyA signal of an endogenous gene, we cloned partial cDNA of neo transcripts by the 3'-RACE method from five trap clones (AYU-1, AYU-2, AYU-4, AYU-5, and AYU-6). In the agarose gel electrophoresis of the products from second cycle PCR, we detected two bands of approximately 350 and 250 bp in AYU-1, two bands of approximately 330 and 310 bp in AYU-2, and one band of approximately 300 bp in AYU-5, and these bands were recovered, digested with ClaI, and subcloned into the ClaI site of pBluescript KS(−) for sequencing. Sequence analysis revealed that three of the five clones had the identical sequence to the 3'-end of the neo gene at one end and polyA at the other end, and unknown sequence between them. The unknown sequences of two neo transcripts in AYU-1 were identical to the 3'-flanking genomic region of the integrated vector of AYU-1 cloned by the plasmid rescue method. So these transcripts were read through the neo gene and 3'-flanking genome, but there were no clear sequences completely matching the consensus sequence of the mammalian polyA addition signal, AATAAA, in this genomic region. Some similar sequences were found in this region but we could not conclude which sequence functions as the polyA addition of the neo transcript. We obtained the same result in the case of AYU-2. No definite band appeared in AYU-4 and 6; and in these cases, the neo gene may be transcribed through the 3'-flanking genome or fused to an exon of endogenous gene

by splicing, leading to the production of long mRNA of which we could not amplify the 3'-region by the 3'-RACE method.

Germ-Line Transmission of Genes Trapped by PolyA Trap Vectors—Unfortunately the trap clones derived from ES MS-1 cells could not contribute to germ cells in chimeric mice. We also carried out the gene-trap experiment using the ES D3 cells. In this experiment, the cell lines established from the clones expressing the reporter genes in undifferentiated states were injected into C57BL/6 blastocysts and chimeric mice were produced. Nine out of 12 chimeric mice showed germ-line transmission. These offspring are now breeding to homozygosity to test whether any phenotype will appear.

DISCUSSION

The selective disruption of genes expressed in undifferentiated embryonic stem cells has been reported (5, 6, 8, 9). However, isolation of developmentally regulated genes that are silent in the undifferentiated state remains a formidable task. To avoid laborious and time-consuming work, we developed a new method to increase the gene-trap frequency by removing the polyA signal from the gene-trap vector.

In previous reports of the gene trap using ES cells, deletions or rearrangements near the integration sites were not fully characterized, despite their particular importance in considering the relationship between the rearrangements and the mutated phenotypes. In two cases we demonstrated the existence of large deletions and/or rearrangements in the genome around the integration sites. Under such circumstances we could not conclude that abnormal phenotypes are due to the disruption of the trapped gene, because a gene adjacent to the trapped gene may also be disrupted. To solve this problem, it is necessary to use other methods of vector introduction or to construct a new vector. Retrovirus vector is shown to introduce only a single copy of the exogenous DNA into the host genome with small risk of DNA rearrangement. However, it is often laborious to obtain a high titer of recombinant retroviral vectors. Thus, we tried to develop a new trap vector which is designed to trap a polyA signal in addition to a promoter or an exon. By deleting the polyA signal from the neo gene, we can expect to select trap clones with minimal deletions or rearrangements and to enrich the trap events over random integrations. However, two problems may arise by using this vector. One is the possibility of expression of the neo gene by use of the polyA signal of the lacZ gene; the other is the possibility of expression of the lacZ gene under the actin promoter of the neo gene when more than two copies of the trap vector are integrated. The results suggest that the former could occur but that the latter did not. As only one copy of the trap vector is integrated in most cases, it is not necessary to take into consideration the expression of the neo gene by use of the polyA signal of the lacZ gene. In any case, an approximately fivefold enrichment of the trap events over random integrations was observed with our trap method, although deletions or rearrangements spanning more than 10 kb still existed.

Skarnes et al. demonstrated that the splice acceptor site upstream of the reporter gene was used properly and

efficiently to create fusion gene products (*6*). But this was not found in our experiment, due to the loss of a consensus sequence required for proper splicing from the vector upon integration into the host genome, as confirmed by sequence analysis of the integrated vectors. It was found that a short sequence of from 2 to 20 bp was deleted from the ends of the integrated vectors and that the deleted region often contained the splice acceptor consensus sequence. Thus, it is necessary to add a longer sequence in front of the splice acceptor to retain the function. This small deletion may be the reason for integration of the trap vector downstream of the promoter region rather than the exon. In four out of five trap clones analyzed, trap vectors were integrated just downstream to the novel endogenous promoters. Macleod *et al.* reported a similar phenomenon with a promoterless *neo* gene (*19*). In this case, most insertions were shown to be in close proximity to CpG islands. This is consistent with our results. They also showed short regions of homology between the genomic target DNA and the construct ends, although we could not confirm this. Retrovirus vectors were also shown to have a tendency to be integrated close to the DNase I hypersensitive site resulting in the promoter trap (*20, 21*). The advantage of the promoter trap will be the creation of null alleles in most cases, because it is unlikely that a fusion protein will retain much of its original activity, and expression of the normal message from the tagged gene could only arise if there is alternative splicing bypassing the promoter trap.

The advantage of the polyA trap method in gene-trap experiments will become apparent when it is combined with an appropriate *in vitro* differentiation system, directed to specific pathways. With such a combination, we should be able to prescreen clones of interest and reduce the amount of work required for generation of chimeric mice using every ES clone. For example, the ES cells spontaneously start to differentiate in suspension culture in the absence of the leukemia inhibitory factor (LIF) and form an embryoid body resembling an early post-implantation embryo (*15*). At least the visceral endoderm, the primitive ectoderm and the mesodermal precursors are differentiated in this system, so that we can analyze the genes involved in the differentiation of these cell lineages.

Systematic isolation of developmental control genes *via* gene-trap events will greatly facilitate molecular understanding of mammalian development.

REFERENCES

1. Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., & Gehring, W.J. (1989) *Genes Dev.* 3, 1288-1300
2. O'Kane, C.J. & Gehring, W.J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 9123-9127
3. Allen, N.D., Cran, D.G., Barton, S.C., Hettle, S., Reik, W., & Surani, M.A. (1988) *Nature* 333, 852-855
4. Kothary, R., Clapoff, S., Brown, A., Campbell, R., Paterson, A., & Rossant, J. (1988) *Nature* 335, 435-437
5. Gossler, A., Joyner, A.L., Rossant, J., & Skarnes, W.C. (1989) *Science* 244, 463-465
6. Skarnes, W.C., Auerbach, B.A., & Joyner, A.L. (1992) *Genes Dev.* 6, 903-918
7. Reddy, S., DeGeogi, J.V., Melchner von, H., & Ruley, H.E. (1991) *J. Virol.* 65, 1507-1515
8. Friedrich, G. Soriano, P. (1991) *Genes Dev.* 5, 1513-1523
9. von Melchner, H., DeGregori, J.V., Rayburn, H., Reddy, S., Friedel, C., Ruley, H.E. (1992) *Genes'Dev.* 6, 919-927
10. Seed, B. & Aruffo, A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3365-3369
11. Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989) *Molecular Cloning*; 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
12. Miyazaki, J., Takaki, S., Araki, K., Tashiro, F., Tominaga, A., Takatsu, K., & Yamamura, K. (1989) *Gene* 79, 269-277
13. Niwa, H., Yamamura, K., & Miyazaki, J. (1991) *Gene* 108, 193-200
14. Gerard, R.D. & Gulzman, Y. (1985) *Mol. Cell. Biol.* 5, 3231-3240
15. Doetchman, T.C., Eistetter, H., Katz, M., Schmidt, W., & Kemler, R. (1985) *J. Embryol. Exp. Morphol.* 87, 27-45
16. Hogan, B., Constantini, F., & Lacy, E. (1986) *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Laboratory, Cold Spring Harbor, New York
17. Frohman, M.A., Dush, M.K., & Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8998-9002
18. Peabody, D.S. & Berg, P. (1986) *Mol. Cell. Biol.* 6, 2695-2703
19. Macleod, D., Lovell-Badge, R., Jones, S., & Jackson, J. (1991) *Nucleic Acids Res.* 19, 17-23
20. Rohdewohld, H., Weiher, H., Reik, W., Jaenish, R., & Breindl, M. (1987) *J. Virol.* 61, 336-343
21. Vijata, S., Steffen, D.L., & Robinson, H.L. (1986) *J. Virol.* 60, 683-692