

RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells

Yasumasa Ishida and Philip Leder*

Department of Genetics, Harvard Medical School, Howard Hughes Medical Institute, 200 Longwood Avenue, Boston, MA 02115, USA

Received June 1, 1999; Revised and Accepted October 5, 1999

ABSTRACT

Gene trapping is a form of insertional mutagenesis that causes disruption of gene function. Here we report the construction and extensive examination of a versatile retrovirus vector, RET (removable exon trap). The RET vector uses an improved poly A-trap strategy for the efficient identification of functional genes regardless of their expression status in target cells. A combination of a potentially very strong splice acceptor and an effective polyadenylation signal assures the complete disruption of the function of trapped genes. Inclusion of a promoterless GFP cDNA in the RET vector allows the expression pattern of the trapped gene to be easily monitored in living cells. Finally, because of loxP-containing LTRs at both ends, the integrated proviruses can be removed from the genome of infected cells by Cre-mediated homologous recombination. Hence, it is possible to attribute the mutant phenotype of gene-trapped cells directly to RET integration by inducing phenotypic reversion after provirus excision. The RET system can be used in conjunction with cell lines with functional heterozygosity, embryonic stem cells, lineage-committed cell lines that differentiate in response to specific inducing factors and other responsive cell lines that can be selected by virtue of their induced green fluorescence protein expression.

INTRODUCTION

Mutagenesis of cells or animals is a powerful genetic strategy for the elucidation of the molecular mechanisms of complex biological phenomena. While chemical mutagenesis is one of the most efficient and widely used procedures to create mutations (1–3), their molecular identification is usually time-consuming and labor-intensive (4–7). Insertional mutagenesis, in contrast, uses some form of a genetic tag to create mutations and subsequently recover the mutated genes. Although the number of mutations per cell which can be produced by insertional mutagenesis is not as high as that generated by chemical mutagenesis, the identification and recovery of mutated genes marked by genetic tags are relatively simple tasks (8–10).

Gene trapping is a form of insertional mutagenesis specifically designed to disrupt gene function by producing intragenic integration events (11). One type of gene-trap approach is based on the random integration of a gene-trap vector containing a promoterless marker gene (like that of β-gal, βgeo or Cre recombinase) (12–16). In a second step, expression of the marker gene confirms that the functional promoter of a cellular gene has, indeed, been trapped. Although any gene in the target cell, including transcriptionally inactive ones, can be identified with this strategy, the obvious disadvantage of this approach is that a large number of clones with random vector integrations have to be handled at the initial stages.

Other gene-trap strategies enrich for intragenic integration events from the very beginning. One commonly used method is the promoter/enhancer trap in which the expression of a promoterless selectable marker (like NEO, Hygro or βgeo) is dependent on the capture of an active transcriptional promoter/enhancer in the target cells (17–19). While this is a very reliable way to focus exclusively on functional genes, transcriptionally silent genes in the target cells are missed by this strategy. Another way to select for intragenic vector integration is to employ a polyadenylation (poly A) trap. In this case, an mRNA transcribed from a selectable marker gene lacking a poly A signal in a gene-trap vector is stabilized only when the gene-trap vector captures a cellular poly A signal (20–23). Since poly A trapping occurs independently of the expression of the target genes, any gene could potentially be identified at almost equal probability regardless of the relative abundance of its transcripts in target cells.

In view of the above, we constructed a versatile retrovirus vector (RET) for general gene-trap experiments *in vitro*. It employs an enhanced poly A-trap strategy for the stringent selection of intragenic provirus integration. Further, a very strong splice acceptor and an efficient poly A signal are included for the complete disruption of trapped genes, the expression patterns of which can easily be followed in living cells by analyzing green fluorescence protein (GFP) expression. Finally, it is possible to confirm that the mutant phenotype was created by provirus integration by removing the integrated provirus using Cre/loxP-mediated homologous recombination and restoring the integrity of the interrupted gene. Here we report the results of an extensive evaluation of

*To whom correspondence should be addressed. Tel: +1 617 432 7667; Fax: +1 617 432 7944; Email: leder@rascal.med.harvard.edu

the RET vector and discuss the potential application of the current gene-trap strategy.

MATERIALS AND METHODS

Plasmid construction

For pRET, the SupF sequence (a 208 bp *Bam*HI–*Bam*HI fragment) was replaced with a loxP signal in the U3 portion of the 3′ LTR of a gag$^+$ self-inactivating retrovirus vector pGen– (24). The following components were inserted into the unique *Xho*I site in the modified pGen– with loxP. (i) The gene-terminator cassette consisting of (a) the human bcl-2 gene intron 2/exon 3 splice acceptor (234 bp of intron portion and 199 bp of exon sequence containing multiple stop codons in all three reading frames) which was PCR-amplified from p18-4H containing a 4 kb fragment of the human bcl-2 gene (a gift from Y. Tsujimoto), (b) the IRES sequence identical to nt 260–845 of the 5′ untranslated region of encephalomyocarditis virus (EMCV-Rueckert) polyprotein RNA (25), (c) the coding region of the EGFP cDNA (Clontech, Palo Alto, CA) and (d) a poly A signal of the bovine growth hormone gene (26). (ii) The HSV (herpes simplex virus) tk (thymidine kinase) gene with the MC1 promoter (27) (an *Xho*I–*Hin*dIII fragment from pIC19R/MC1-TK, a gift from M. R. Capecchi). (iii) A NEO cassette consisting of (a) a short form (224 bp) of the RNA polymerase II gene promoter (26), the NEO coding region and the splice donor of the mouse hprt gene exon 8/intron 8 (17 and 157 bp of exon and intron portions, respectively), all of which are included in the 1.2 kb *Hin*cII–*Aat*II fragment of a NEO expression vector pKT3NP4 (a gift from K. Thomas and M. R. Capecchi) and (b) an mRNA instability signal (53 bp) derived from the 3′ untranslated region of the human GM-CSF cDNA (28). All of the type I–IV virus constructs (Fig. 2) lack the IRES/EGFP sequence in their gene-terminator cassettes. The type III and IV constructs possess, immediately downstream of the NEO cassette, a 1.2 kb *Ssp*I–*Eco*RV fragment of pKT3NP4 including the 3′ half of intron 8 and exon 9 (the last exon) of the mouse hprt gene. For pSAT, the entire coding region of the mouse IL-4 cDNA (PCR-generated) (29), a partially PCR-generated fragment of the mouse IgE receptor (FcεRI) α-chain gene (30) containing the 3′ end of exon 3 (28 bp), intron 3 (2.5 kb) and the 5′ portion of exon 4 (47 bp), and a *Bgl*II–*Xho*I fragment of pAPtag-1 (31) encoding the human placental alkaline phosphatase without secretion signal were co-ligated in-frame into pcDNA I Amp (Invitrogen, Carlsbad, CA) pre-digested with *Hin*dIII and *Xho*I. A multiple cloning site was synthetically generated at the *Eco*RI site in the middle of intron 3. The amino acid sequence of the junctional portions of the fusion protein is: IL-4 … QMDYS/GT (artificial)/KPVYL … IgE-R (α) … ILVHG (total 25 residues)/SSG (artificial)/IIPV … alkaline phosphatase. For pCAGGS-NLS/Cre, the coding region of pMC1-Cre (32) was PCR-amplified to generate a Kozak consensus sequence for the efficient translational initiation and transferred into the *Eco*RI site of pCAGGS containing a CMV enhancer and a chicken β-actin promoter (33).

Retrovirology

For virus production, the ψ2 (34) or Phoenix (35) ecotropic packaging cell line was transiently transfected with a virus plasmid by the calcium phosphate method, and the virus-containing culture supernatant was harvested as described (36). Virus titer was measured as colony forming units after infection (standard method) and HAT selection (see below) of NIH 3T3 tk (-) cells (37).

Cell culture and manipulation

Wild-type and tk(–) NIH 3T3 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% calf serum, 2 mM L-glutamine, 100 U/ml of penicillin and 100 μg/ml of streptomycin. For HAT (or HT) selection, 100 μM of sodium hypoxanthine, 0.4 μM of aminopterin (omitted for HT) and 16 μM of thymidine were included in the culture medium for 8 days. For G418 and Gancyclovir (Ganc) selections, cells were cultured for 8–10 days in the presence of 1 mg/ml (as activity) of Geneticin (Gibco BRL, Gaithersburg, MD) and 10 μM of Cytovene (Syntex, Palo Alto, CA), respectively. After RET infection and G418 selection, GFP-positive cells were collected by an EPICS ALTRA cell sorter (Coulter, Miami, FL). Integrated proviruses were removed by transient transfection ($CaPO_4$ method) of the cells with pCAGGS-NLS/Cre followed by Ganc selection.

IL-4 and AP assay

Culture and DEAE-dextran-mediated transfection of COS7 cells were performed as described (36). Relative murine IL-4 levels were measured as raw OD_{450} values using the Intertest-4X ELISA kit (Genzyme, Cambridge, MA) according to the manufacturer's instructions. Relative AP levels were also measured as raw OD_{405} values as described (31).

3′ RACE and sequencing

RNA was isolated from each clone of infected NIH 3T3 cells (confluent in a 35 mm dish) using the Trizol reagent (Gibco BRL). First strand cDNA was synthesized in a 20 μl reaction from ~5 μg of total RNA with AD-poly(T) primer (5′-CGTAGCTCTAGACTCCGTGTCCAACT$_{20}$-3′) and SuperScript II reverse transcriptase (Gibco BRL) as described in the manufacturer's protocol. Two rounds of 3′ RACE was performed in a 50 μl reaction using the Advantage-GC cDNA polymerase mix (Clontech), first with 1 μl of the above cDNA mix (1/20 of the total reaction), NEO 1.5 primer (5′-GCGAATGGGCTGACCGCTTCCTCGTGC-3′) and AD primer (5′-CGTAGCTCTAGACTCCGTGTCCAAC-3′), and second with 1 μl of the first PCR mix (1/50 of the total reaction), NEO 2.0 primer (5′-TACGGTATCGCCGCTCCCGATTCGCAG-3′) and AD-plus primer (5′- CGTAGCTCTAGACTCCGTGTCCAACTTTT-3′). The PCR conditions were as follows: (i) 94°C, 3 min (1st round) or 1 min (2nd round), (ii) 8× (94°C, 40 s + 72°C, 4 min), (iii) 32× (94°C, 40 s + 66°C, 2 min + 72°C, 2 min) and (iv) 72°C, 4 min. After confirming the successful amplification of the DNA fragment(s) by agarose gel electrophoresis, the PCR primers were removed from 30 μl of the final 3′ RACE reaction mix by the QIAquick PCR purification kit (Qiagen, Valencia, CA), and the nucleotide sequence of the purified PCR fragment(s) was determined by direct sequencing with the NEO.SEQ primer (5′-TGACGAGTTCTTCTGAGGGGATCC-3′) containing a *Bam*HI recognition sequence at the end, which is unique to the NEO cassette of the RET vector, and an ABI 377 DNA sequencer (ABI, San Francisco, CA).

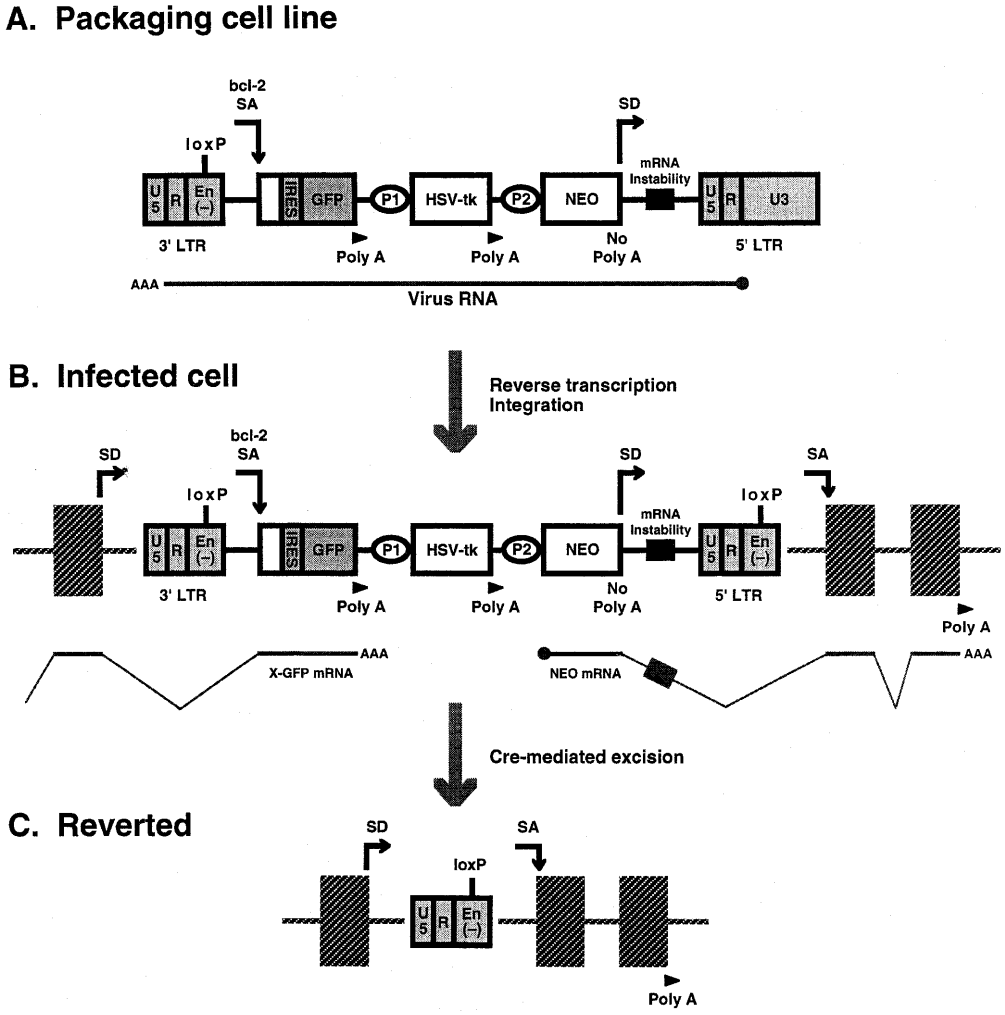

Figure 1. The RET gene-trap vector. (**A**) Structure of the RET vector in a packaging cell line. Note that the 5′ LTR of the RET vector has a wild-type sequence, whereas the 3′ LTR lacks the transcriptional enhancer in its U3 portion ('self-inactivating' retrovirus vector). Instead, the U3 portion of the 3′ LTR possesses a loxP signal. (**B**) Structure of the RET provirus in infected cells. The nucleotide sequence of the 3′ LTR is accurately copied to the 5′ LTR during reverse transcription and integration. Therefore, the 5′ LTR in infected cells also lacks the enhancer and possesses a loxP signal in the same orientation as the one in the 3′ LTR. The RET virus carries the following three elements in the reverse orientation relative to virus transcription: (i) a gene terminator cassette consisting of a strong splice acceptor (SA), multiple stop codons, an IRES sequence, a GFP cDNA and an efficient poly A signal; (ii) a complete and constitutive HSV tk cassette for virus titration and negative selection; and (iii) a unique NEO cassette for the enhanced poly A trap, containing a splice donor (SD) and an mRNA instability signal, but lacking a poly A signal. (**C**) Structure of residual provirus elements following Cre-mediated excision. Diagonally striped black rectangles represent exons of a trapped gene. En (–): enhancer deletion; P1: MC1 promoter; P2: RNA polymerase II promoter (short form); X-GFP mRNA: fusion transcript for the 5′-half of the trapped gene X and the GFP cDNA.

RESULTS AND DISCUSSION

Enhanced poly A trap

Instead of using an enhancer/promoter-trap strategy, the RET vector is designed to capture the poly A signal of cellular genes, regardless of their expression status in target cells (20–23). As shown in Figure 1, the NEO gene, driven by the constitutive, but moderately active RNA polymerase II promoter (short form) (26), lacks its own poly A signal, but encodes a splice donor signal followed by an mRNA instability signal derived from the 3′ untranslated region of the human GM-CSF cDNA (28). When provirus integration occurs outside a cellular gene, NEO mRNA fails to acquire a poly A tail and retains the mRNA instability signal which further destabilizes the NEO transcript, resulting in the G418-sensitivity of such infected cells. However, if the provirus is integrated into an intronic portion of a cellular gene in correct orientation (as shown in Fig. 1B), NEO mRNA acquires a poly A tail by using the poly A signal of the trapped gene. Furthermore, the mRNA instability signal is removed from the NEO transcript by pre-mRNA splicing between the splice donor of the NEO cassette and the splice acceptor of the trapped gene. These events result in the complete stabilization of NEO mRNA and, as a consequence, resistance of such infected cells to G418.

To assess each element of the above poly A-trap strategy, we made four types of RET vectors with differently constructed NEO cassettes (Fig. 2). The type I vector has exactly the same NEO cassette as the RET vector as shown in Figure 1, including the mRNA instability signal. The type II vector lacks the instability signal, but is otherwise identical to the type I

vector. The type III vector differs from type I only in that it has an additional exon sequence with a splice acceptor and a poly A signal between the mRNA instability signal and the 5′ LTR. The type IV vector lacks the instability signal, but is otherwise identical to type III.

Each of these viral constructs was tested by infecting NIH 3T3 tk(–) cells which were then HAT-selected. HAT medium selects for all infection events because the MC1 promoter-driven HSV tk gene is constitutively active in each construct (27). After identifying all the infected cells as HAT-resistant independent colonies, the culture medium was changed to one containing G418. For the type III and IV viruses, it was expected that most of the HAT-resistant colonies would survive G418 selection. This is because NEO mRNA derived from the type III and IV vectors should be stable, no matter where provirus integration occurs because of the downstream splicing event that loops out the instability signal and the addition of a poly A tail to the mRNA. In contrast, it is expected that only a fraction of HAT-resistant colonies carrying the type I or II viruses become G418-resistant, depending on where provirus integration occurs.

As shown in Figure 2, only 13 and 26%, respectively, of the HAT-resistant colonies infected with the type I and II viruses survived G418 selection. By contrast, the majority (~80%) of the HAT-resistant colonies infected with the type III and IV viruses became G418-resistant. This data strongly suggests that resistance to G418 in cells infected with type I and II viruses depends on where provirus integration occurs. Moreover, the selectivity seems to be higher for the type I virus than for the type II virus, the latter lacking an mRNA instability signal (20–23). Another important point is the fact that there is no significant difference in the proportion of G418-resistant colonies for the type III and IV viruses. This indicates that the mRNA instability signal does not affect RNA stability if it is located in an intron. Therefore, once the RET virus is integrated into an intron of a gene in correct orientation (as shown in Fig. 1B), the mRNA instability signal of the NEO cassette does not destabilize the newly generated chimeric NEO pre-mRNA transcript.

To examine more directly the effectiveness of our strategy for the enrichment of intragenic provirus integration events, we infected NIH 3T3 cells with the RET retrovirus and determined the nucleotide sequences of the 3′ untranslated regions (UTR) of various NEO mRNAs obtained from independently selected G418-resistant NIH 3T3 clones. RNA was extracted from each resistant clone and cDNA fragments derived from the 3′ UTR of the chimeric NEO transcripts were amplified by two rounds of 3′ RACE (rapid amplification of cDNA ends) using NEO-specific 5′ primers and 3′ adapter primers. As seen in the examples shown in Table 1, the vector sequence immediately downstream from the splice donor of the NEO precursor mRNA was always precisely replaced by a non-vector sequence (n = 65). This indicates that the mRNA instability signal located downstream from the splice donor must be spliced out in order for the NEO gene to function properly. Also, a high (>90%) success rate for the 3′ RACE procedure suggests that the majority of the NEO mRNAs derived from G418-resistant cells have poly A (or poly A-like) structures at their 3′ ends. Although sequence information as to the exon–intron boundaries is not always available even for well-known genes, the nucleotide sequence of some 3′ RACE products

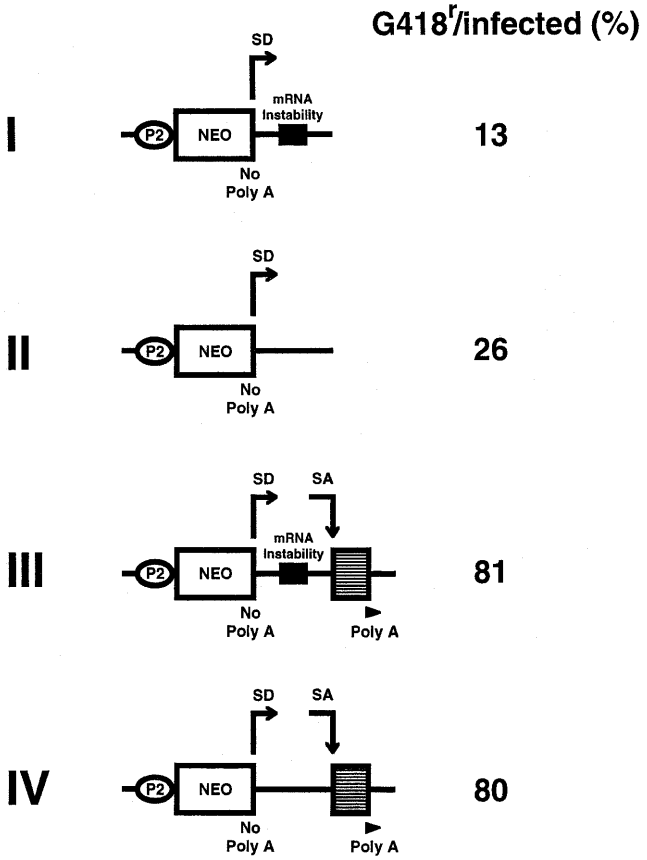

Figure 2. Enrichment for the intragenic provirus integration by the enhanced poly A trap. After infection and HAT selection of NIH 3T3 tk(–) cells, 400 HAT-resistant colonies for each virus construct were further grown in the medium containing G418 and HT. Percentages of the colonies which survived the G418 selection are shown. Striped rectangles represent exon 9 (last exon) of the mouse hprt gene.

confirmed that correct endogenous splice acceptors were used to accept the splice donor sequence from the NEO cassette (data not shown).

Taken together, these data suggest that our enhanced poly A-trap strategy works well in capturing functional cellular splice acceptors and poly A signals. On the other hand, splicing-out the instability signal coupled with acquisition of a poly A tail does not always result from trapping events in functional genes. As shown in Table 1, some repetitive genomic regions or reverse strands of known genes seem to behave like the 3′-most part of a functional gene (i.e. a splice acceptor followed by a poly A signal). This kind of ‘pseudo gene-trap’ seems to be particularly common among GFP-negative clones identified after G418 selection (see below).

Disruption of trapped genes

The RET vector contains two transcriptional terminators (poly A signals): the bovine growth hormone (BGH) gene poly A signal in the gene terminator cassette [one of the most efficient thus far analyzed (26)] and a ‘fail-safe’ terminator in the HSV tk gene (Fig. 1). Transcription of the trapped cellular gene should be terminated at either one of these poly A signals. Even if some read-through occurs at these sites, a potentially strong splice acceptor in the gene terminator cassette should

Table 1. Examples of the nucleotide sequence of cDNAs for the fusion transcripts between the NEO cassette and the trapped genes

RET vector sequence around NEO splice donor	Accession no.	Comment
exon intron GGGATTTGAAT GTAAGTAATGCTTTTTTT.....'ATTTA'...		
GFP-positive clones		
GGGATTTGAAT GTCAGAGNGGGGCTGGNAACAACTGGGCCA...	AA798990	β-tubulin
GGGATTTGAAT GGCAAGAAATACAAGCCCCTAGACCTGCGA...	AA058275	L35 ribosomal protein
GGGATTTGAAT GTACAACTCTAAGAGAAGACACTGGNGGAG...	AA518802	L39 ribosomal protein
GGGATTTGAAT TTGGACTAAGCGACCTTGAGNGGATTCGAN...	L31609	S29 ribosomal protein
GGGATTTGAAT GGCNTGTTGAGTCGGCTGCATCTAGAGANC...	W83911	EST, embryo (13.5–14.5 d.p.c.)-derived
GFP-negative clones	Y11148	LUN protein (lung-specific transcription factor)
GGGATTTGAAT TNGGACTGCCCNGTTACCAGCACCACTCCA...	AF013604	Serotonin transporter
GGGATTTGAAT CGCATTATTAAAAGTATCACTCCTGAAACA...	M84487	Vascular cell adhesion molecule-1 (VCAM-1)
GGGATTTGAAT GAAAAGAACATAACNGGGACTATTTTTCGC...	AI325997	Interferon-inducible protein 1–8 U
GGGATTTGAAT TCCGGGNCCGGAAGANGGGGGGCGACATGA...	AA269939	Skeletal muscle myosin heavy chain
GGGATTTGAAT GAGGAACAATCCAATGTCAACCTGGCCAAA...		
GGGATTTGAAT GTTGACTTTGGATTCGCTAANAAGATNGGC...	L12460	Cyclic GMP-dependent protein kinase II
GGGATTTGAAT ACACGAGCAATTTGTCCTCTGAGAGGCTCC...	AA230373	EST, kidney-derived
GGGATTTGAAT GATCACTTATACTCATTTAAATAATCGGAT...	K00131	B2 repeat sequence
GGGATTTGAAT GNCTCTTGACNCCCTNTGGGGNNNCACAAA...	D13544	Reverse strand (primase small subunit)
GGGATTTGAAT GCTATAGACATGANGGCTGAAACTTCAATG...		No homolog in databases
GGGATTTGAAT TTTGTGTGTTTCAGAGCAAATGGACCAGAC...		No homolog in databases

cDNAs of the fusion transcripts were amplified by NEO-specific 3′ RACE, and the nucleotide sequence of the most upstream portion of the PCR products was determined by direct sequencing with a NEO primer. 'ATTTA' in the intronic portion of the RET vector sequence is the mRNA instability signal derived from the 3′ untranslated region of the human GM-CSF cDNA.

disrupt the correct splicing between the two exons of the trapped cellular gene (Fig. 1). This particular splice acceptor is derived from the human bcl-2 gene (intron 2/exon 3) that physiologically splices across the 370 kb intron 2 of the bcl-2 gene (38). No alternative splicing of this intron has been observed despite its enormous length, suggesting that this splice acceptor is relatively efficient. In fact, the structure of this bcl-2 splice acceptor is unique in that it has an exceptionally long pyrimidine tract (~100 bp) in front of the essential AG residues (data not shown). This is a potential advantage over gene-trap vectors that utilize a splice acceptor derived from the mouse engrailed-2 gene (12,14–16,18,22). Creating null alleles with these latter vectors has proven difficult, since splicing around the gene trap insertion usually occurs and results in the restoration of the undisrupted wild-type mRNA (39–42).

To test further the efficiency of the RET gene terminator cassette, we made a plasmid (pSAT) which allowed us to analyze the combined effect of acquiring a splice acceptor and a poly A signal (Fig. 3). If there is no disrupting element in the intron of this construct, splicing between the splice donor and acceptor of the FcεRI α-chain gene connects the mouse interleukin-4 (IL-4) and human secreted alkaline phosphatase (AP) cDNAs, resulting in the secretion of an IL-4–AP fusion protein from cells transfected with the plasmid. However, if the intron carries a strong disruptive sequence, only IL-4, and not AP, will be detected in the supernatant of transfected cells. As shown in Table 2, transfection of COS7 cells with pSAT itself produced both IL-4 and AP activities in the supernatant. Insertion of a minimal disruption unit (the bcl-2 gene splice acceptor and the growth hormone gene poly A signal) into the intron appears to have completely abrogated the splicing and/or transcription because only IL-4, and not AP, is detected in the supernatant of the transfected cells (Table 2). As a control,

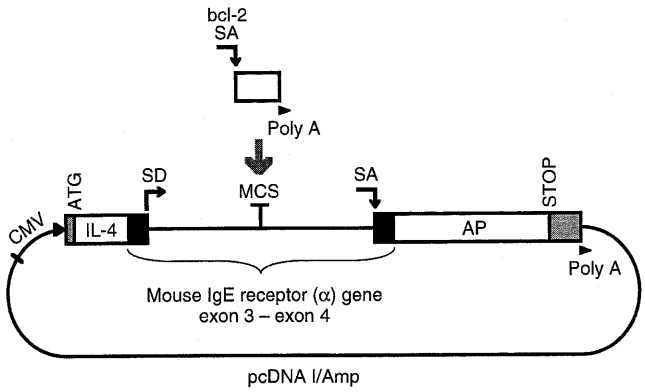

Figure 3. pSAT, the splice acceptor test plasmid. Coding region of the mouse IL-4 cDNA, a fragment of the mouse IgE receptor (FcεRI) α-chain gene (covering the 3′ half of exon 3, intron 3 and the 5′ half of exon 4) and the human placental alkaline phosphatase (AP) cDNA were connected in-frame to encode a fusion protein of IL-4 and AP. MCS: multiple cloning site.

we see that insertion of the minimal disruption unit in the reverse orientation restores downstream AP activity (Table 2). This indicates that the minimal disruption unit is highly effective, does not 'leak', and acts in an orientation-dependent, sequence-specific manner. The high performance of this disruption cassette is very important for gene trap experiments because leaky disruption of trapped genes often complicates analysis of the phenotype of gene-trapped cells (39–42).

Monitoring the expression of trapped genes in living cells

As a further 'fail-safe' provision, the gene terminator cassette of the RET vector contains multiple stop codons in all three reading frames in the sequence 3′ to the bcl-2 splice acceptor.

It thereafter encodes an internal ribosome entry site (IRES) sequence (25), a GFP cDNA and a BGH poly A signal (Fig. 1). The GFP serves as a visible marker of transcriptional activity of the trapped gene. RNA splicing between an exon of the trapped gene and the gene terminator cassette generates a fusion transcript consisting of the 5′ part of the trapped gene and the GFP cDNA. Ribosomes translating the fusion transcript would be expected to dissociate from it at one of the multiple stop codons and to re-associate at the IRES portion, resulting in bi-cistronic expression of the first half of the trapped gene and GFP. Note also that neither of the two LTRs of the integrated RET provirus has transcriptional enhancers, thereby minimizing the possibility of insertional activation of trapped or adjacent genes (9).

Table 2. Quantitative assessment of the efficiency of the minimal gene-disruption cassette

Introduced plasmid	IL-4 ELISA (OD_{450})	AP activity (OD_{405})
No DNA	0.182 ± 0.012	0.839 ± 0.012
pSAT-Empty	2.433 ± 0.143	3.473 ± 0.123
pSAT-BCL2.SA/GH.pA Fwd	2.412 ± 0.058	0.821 ± 0.018
pSAT-BCL2.SA/GH.pA Rev	2.404 ± 0.088	2.812 ± 0.154

COS7 cells were transiently transfected with one of the indicated plasmids, and the AP activity and the IL-4 amount in the culture supernatant were measured as described in Materials and Methods. Average OD values ± standard deviations of three independent transfection experiments are shown. GH, growth hormone; pA, poly A signal; Fwd, forward; Rev, reverse.

To examine whether the level of GFP expression correlates well with that of the expected expression behavior of a trapped gene, a mixed population of gene-trapped NIH 3T3 cells were first separated by a fluorescence-activated cell sorter into GFP-positive and -negative subgroups. After isolating independent clones from each subgroup, the cDNA sequences of the trapped genes were determined (Table 1). Overall, the level of GFP expression seems to correlate well with that expected of the trapped gene. The GFP-positive subgroup contained clones in which ubiquitously and highly expressed genes like that of β-tubulin or L35 ribosomal protein were trapped. In contrast, the GFP-negative subgroup included clones corresponding to genes that are not expressed (or only very weakly expressed) in NIH 3T3 cells (Table 1). In a similar experiment, the immunoglobulin κ chain gene was found to be trapped by the RET vector in a GFP-negative mouse embryonic stem (ES) cell clone (data not shown), further supporting the notion that any gene can be trapped by the RET vector regardless of its expression pattern in the target cells.

From the strongly GFP-positive subgroup, all five independent clones analyzed thus far had ‘trapped’ an active gene [i.e. four known genes and one expressed sequence tag (EST)] (Table 1). In the case of 60 independent GFP-negative clones analyzed, 15 (25%) trapped apparently functional genes (seven known genes and eight ESTs). The balance included nine clones (15%) in which either repetitive DNA sequences or reverse strands of known genes were trapped and 36 clones (60%) in which the DNA sequences matched with ‘no significant hit’ in databases (Table 1). Obviously, this group of ‘no hit’ DNA sequences is likely to include novel functional genes as well as non-functional genomic sequences.

In the case of NIH 3T3 cells, ~10% become clearly GFP-positive after RET infection and G418 selection (data not shown) and almost all of these GFP-positive cells presumably involve the trapping of functional genes. In addition, according to the above estimate, at least 25% of the remaining GFP-negative cells (90% of total) are also expected to have undergone an authentic gene trap event (90 × 0.25 = 23% of total). Taken together, as a minimum, 33% (10% GFP-positive and 23% GFP-negative) of the total RET-infected/G418-selected cells seem to have trapped functional genes. This is likely to be an underestimate given that a proportion of the ‘no hit’ sequences will be identified as expressed genes in the future.

Gene trapping has been used successfully to identify genes whose expression is induced upon *in vitro* differentiation of ES cells (16,43,44), growth-factor deprivation from a hematopoietic precursor cell line (13) and T-cell activation (14). Our RET vector system could obviously be used for similar purposes with significant advantage. In the RET system, it is not necessary to isolate all the cells with randomly integrated gene-trap vectors and analyze further for the induced expression of a marker gene included in the vector. The RET vector first selects for intragenic integration events, thereby minimizing the number of clones to be handled in detailed gene induction analyses in later stages. Moreover, the RET system provides for rapid isolation of a cDNA fragment corresponding to an inducible gene by using 3′ RACE on chimeric NEO mRNA which is constitutively synthesized from an RNA *pol*II promoter (Table 1). The RET vector can also be used to disrupt genes in ES cells whose expression is restricted to differentiated peripheral tissues. Such trapping events can certainly be enriched by focusing on GFP-negative ES clones after RET infection and G418 selection.

Cre/loxP-mediated excision of the integrated RET provirus and restoration of the function of the trapped gene

When a mutant phenotype is observed after a gene-trap experiment, it is desirable to be able to confirm that the resulting phenotype was really brought about by provirus insertion. This is most rigorously done by removing the integrated provirus from the host genome and scoring for phenotypic reversion. As shown in Figure 1, the integrated RET virus acquires an additional loxP site in its 5′ LTR in the same orientation as the loxP signal in the 3′ LTR. To demonstrate the ability to excise integrated RET proviruses, a mixed population of strongly GFP-positive NIH 3T3 cells in which genes had been trapped by the RET virus were transiently transfected with an expression vector encoding the Cre recombinase and selected under Ganc (to select for loss of the HSV tk gene). Cells from multiple Ganc-resistant colonies were mixed and subjected to a flow cytometric analysis (Fig. 4). No GFP-positive cells were detected among the Ganc-resistant cells. Furthermore, the fluorescence intensity of the virus-removed cells was indistinguishable from that of the uninfected parental NIH 3T3 cells (Fig. 4), indicating that the integrated RET proviruses were efficiently excised by the Cre recombinase.

As shown in Figure 1, one enhancer-negative LTR remains on the host chromosome after the excision of the RET provirus. However, the effect of this LTR on the transcription

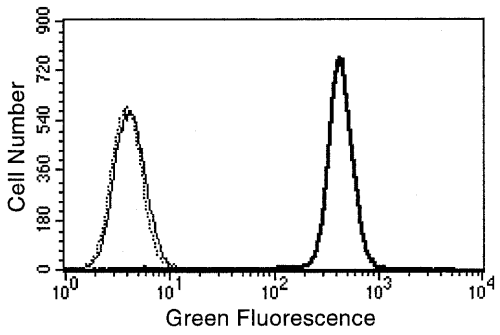

Figure 4. Cre/loxP-mediated removal of the integrated RET proviruses. A mixed population of the twice-sorted NIH 3T3 cells strongly expressing GFP were transiently transfected with a Cre-expressing plasmid and selected against the HSV tk gene using Ganc, and green fluorescence and cell number were determined in a flow cytometer. Thin line: uninfected NIH 3T3 cells; thick line: infected, G418-selected and twice-sorted NIH 3T3 cells strongly positive for GFP; dotted line: Cre-transfected and Ganc-selected NIH 3T3 cells which had been strongly positive for GFP.

and splicing of the cellular gene should be minimal for the following reasons. (i) Since the single LTR left after the excision lacks the transcriptional enhancer, it should not affect the promoter of the previously trapped cellular gene. (ii) The structure of the single LTR is exactly the same as that of the 5′ LTR located downstream of the mRNA instability signal in infected cells (Fig. 1B and C). Since this 5′ LTR does not interfere with splicing between the NEO splice donor and the splice acceptor of the trapped gene (see chimeric NEO transcripts in Table 1), the same tolerance for splicing around the remaining LTR would be expected after virus removal. Thus, the trapped cellular genes should recover their function following excision of the integrated RET proviruses.

Cells cultured *in vitro* sometimes show mutant phenotypes after disruption of only a single allele of an autosomal gene (45–48). This is believed to be the consequence of functional heterozygosity of a fraction of the genes in cultured cells. In fact, many genes in Chinese hamster ovary-K1 (CHO-K1) cells have been shown to be functionally heterozygous, probably due to abnormal imprinting of DNA in the cell line (1). Therefore, it would be interesting to carry out insertional mutagenesis experiments using the RET vector and retrovirally infectable CHO-K1 cells looking for some mutant phenotype. The strong disruptive ability and removability of the RET vector would be crucial to this kind of insertional mutagenesis *in vitro.*

CONCLUSIONS

We have described a versatile retrovirus vector, RET, which should be useful in a wide range of gene-trap experiments. Unlike other useful poly A-trap vectors, the RET virus contains a constitutively active selectable marker, the HSV tk cassette, which makes virus titration possible and subsequent infection experiments more quantitative. The enhanced poly A-trap strategy involving an mRNA instability signal seems to be more efficient and selective than conventional ones. Moreover, the gene-disruption cassette appears to be 'leak proof' as evidenced by trials in a transient transfection assay system. In addition, the ease of monitoring the expression patterns of trapped genes using GFP in living cells allows us to apply this vector to a variety of experimental systems. Finally, the Cre recombinase-mediated removability of the disrupting elements makes this system a particularly powerful genetic tool in which the mutant phenotype can be reverted with restoration of gene function.

ACKNOWLEDGEMENTS

We are grateful to all the members of the Leder Laboratory and Connie Cepko for advice, discussions and encouragement. We thank P. Soriano, Y. Tsujimoto, T. F. Lane, C. Deng, K. Thomas, M. R. Capecchi and K. Rajewsky for plasmids, G. P. Nolan and A. D. Miller for cell lines and Juanita Campos-Torres, Cathie Daugherty, Montserrat Michelman and Shawn C. Fields-Berry for expert technical assistance. We also thank Jeff Ecsedy, Jennifer Michaelson, Boris Reizis and Robert Weiss for their critical reading of the manuscript and Yaoqi Wang for expert FACS technique. Y.I. was initially supported in part by the Uehara Memorial Foundation (Tokyo, Japan) and Kyoto University Faculty of Medicine.

REFERENCES

1. Siminovitch,L. (1976) *Cell*, **7**, 1–11.
2. Schimenti,J. and Bucan,M. (1998) *Genome Res.*, **8**, 698–710.
3. Brown,S.D. and Nolan,P.M. (1998) *Hum. Mol. Genet.*, **7**, 1627–1633.
4. Antoch,M.P., Song,E.J., Chang,A.M., Vitaterna,M.H., Zhao,Y., Wilsbacher,L.D., Sangoram,A.M., King,D.P., Pinto,L.H. and Takahashi,J.S. (1997) *Cell*, **89**, 655–667.
5. King,D.P., Zhao,Y., Sangoram,A.M., Wilsbacher,L.D., Tanaka,M., Antoch,M.P., Steeves,T.D., Vitaterna,M.H., Kornhauser,J.M., Lowrey,P.L., Turek,F.W. and Takahashi,J.S. (1997) *Cell*, **89**, 641–653.
6. Li,Z., Otevrel,T., Gao,Y., Cheng,H.L., Seed,B., Stamato,T.D., Taccioli,G.E. and Alt,F.W. (1995) *Cell*, **83**, 1079–1089.
7. Tsukamoto,T., Miura,S., Nakai,T., Yokota,S., Shimozawa,N., Suzuki,Y., Orii,T., Fujiki,Y., Sakai,F., Bogaki,A. *et al*. (1995) *Nature Genet.*, **11**, 395–401.
8. Favor,J. and Morawetz,C. (1992) *Mutat. Res.*, **284**, 53–74.
9. Jonkers,J. and Berns,A. (1996) *Biochim. Biophys. Acta*, **1287**, 29–57.
10. Woychik,R.P. and Alagramam,K. (1998) *Int. J. Dev. Biol.*, **42**, 1009–1017.
11. Evans,M.J. (1998) *Dev. Dyn.*, **212**, 167–169.
12. Gossler,A., Joyner,A.L., Rossant,J. and Skarnes,W.C. (1989) *Science*, **244**, 463–465.
13. Russ,A.P., Friedel,C., Ballas,K., Kalina,U., Zahn,D., Strebhardt,K. and von Melchner,H. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 15279–15284.
14. Whitney,M., Rockenstein,E., Cantin,G., Knapp,T., Zlokarnik,G., Sanders,P., Durick,K., Craig,F.F. and Negulescu,P.A. (1998) *Nature Biotechnol.*, **16**, 1329–1333.
15. Wurst,W., Rossant,J., Prideaux,V., Kownacka,M., Joyner,A., Hill,D.P., Guillemot,F., Gasca,S., Cado,D., Auerbach,A. *et al*. (1995) *Genetics*, **139**, 889–899.
16. Forrester,L.M., Nagy,A., Sam,M., Watt,A., Stevenson,L., Bernstein,A., Joyner,A.L. and Wurst,W. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 1677–1682.
17. von Melchner,H. and Ruley,H.E. (1989) *J. Virol.*, **63**, 3227–3333.
18. Skarnes,W.C., Moss,J.E., Hurtley,S.M. and Beddington,R.S. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 6592–6596.
19. Friedrich,G. and Soriano,P. (1991) *Genes Dev.*, **5**, 1513–1523.
20. Yoshida,M., Yagi,T., Furuta,Y., Takayanagi,K., Kominami,R., Takeda,N., Tokunaga,T., Chiba,J., Ikawa,Y. and Aizawa,S. (1995) *Transgenic Res.*, **4**, 277–287.
21. Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L., Person,C. and Sands,A.T. (1998) *Nature*, **392**, 608–611.
22. Salminen,M., Meyer,B.I. and Gruss,P. (1998) *Dev. Dyn.*, **212**, 326–333.
23. Niwa,H., Araki,K., Kimura,S., Taniguchi,S., Wakasugi,S. and Yamamura,K. (1993) *J. Biochem. (Tokyo)*, **113**, 343–349.
24. Soriano,P., Friedrich,G. and Lawinger,P. (1991) *J. Virol.*, **65**, 2314–2319.

25. Duke,G.M., Hoffman,M.A. and Palmenberg,A.C. (1992) *J. Virol.*, **66**,1602–1609.
26. Soriano,P., Montgomery,C., Geske,R. and Bradley,A. (1991) *Cell*, **64**, 693–702.
27. Thomas,K.R. and Capecchi,M.R. (1987) *Cell*, **51**, 503–512.
28. Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
29. Noma,Y., Sideras,P., Naito,T., Bergstedt-Lindquist,S., Azuma,C., Severinson,E., Tanabe,T., Kinashi,T., Matsuda,F., Yaoita,Y. and Honjo,T. (1986) *Nature*, **319**, 640–646.
30. Ye,Z.S., Kinet,J.P. and Paul,W.E. (1992) *J. Immunol.*, **149**, 897–900.
31. Flanagan,J.G. and Leder,P. (1990) *Cell*, **63**, 185–194.
32. Gu,H., Zou,Y.R. and Rajewsky,K. (1993) *Cell*, **73**, 1155–1164.
33. Niwa,H., Yamamura,K. and Miyazaki,J. (1991) *Gene*, **108**, 193–199.
34. Mann,R., Mulligan,R.C. and Baltimore,D. (1983) *Cell*, **33**, 153–159.
35. Grignani,F., Kinsella,T., Mencarelli,A., Valtieri,M., Riganelli,D., Lanfrancone,L., Peschle,C., Nolan,G.P. and Pelicci,P.G. (1998) *Cancer Res.*, **58**, 14–19.
36. Ausubel,F., Brent,R., Kingston,R., Moore,D., Seidman,J., Smith,J. and Struhl,K. (1987–1999) *Current Protocols in Molecular Biology*. John Wiley and Sons, New York, NY.
37. Wei,C.M., Gibson,M., Spear,P.G. and Scolnick,E.M. (1981) *J. Virol.*, **39**, 935–944.
38. Seto,M., Jaeger,U., Hockett,R.D., Graninger,W., Bennett,S., Goldman,P. and Korsmeyer,S.J. (1988) *EMBO J.*, **7**, 123–131.
39. Voss,A.K., Thomas,T. and Gruss,P. (1998) *Dev. Dyn.*, **212**, 258–266.
40. Serafini,T., Colamarino,S.A., Leonardo,E.D., Wang,H., Beddington,R., Skarnes,W.C. and Tessier-Lavigne,M. (1996) *Cell*, **87**, 1001–1014.
41. Sam,M., Wurst,W., Kluppel,M., Jin,O., Heng,H. and Bernstein,A. (1998) *Dev. Dyn.*, **212**, 304–317.
42. McClive,P., Pall,G., Newton,K., Lee,M., Mullins,J. and Forrester,L. (1998) *Dev. Dyn.*, **212**, 267–276.
43. Thorey,I.S., Muth,K., Russ,A.P., Otte,J., Reffelmann,A. and von Melchner,H. (1998) *Mol. Cell. Biol.*, **18**, 3081–3088.
44. Stanford,W.L., Caruana,G., Vallis,K.A., Inamdar,M., Hidaka,M., Bautch,V.L. and Bernstein,A. (1998) *Blood*, **92**, 4622–4631.
45. Straus,D.B. and Weiss,A. (1992) *Cell*, **70**, 585–593.
46. Jasin,M., Elledge,S.J., Davis,R.W. and Berg,P. (1990) *Genes Dev.*, **4**, 157–166.
47. Chang,W., Hubbard,S.C., Friedel,C. and Ruley,H.E. (1993) *Virology*, **193**, 737–747.
48. Aoe,T., Ohno,H., Sonta,S. and Saito,T. (1993) *Int. Immunol.*, **5**, 725–733.