JOURNAL OF VIROLOGY, Mar. 1991, p. 1507–1515
0022-538X/91/031507-09$02.00/0

Vol. 65, No. 3

Retrovirus Promoter-Trap Vector To Induce *lacZ* Gene Fusions in Mammalian Cells

SITA REDDY,* JAMES V. DEGREGORI, HARALD VON MELCHNER, AND H. EARL RULEY

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, 40 Ames Street, Cambridge, Massachusetts 02139

Received 8 June 1990/Accepted 4 December 1990

A retrovirus promoter-trap vector (U3LacZ) has been developed in which *Escherichia coli lacZ* coding sequences were inserted into the 3′ long terminal repeat (LTR) of an enhancerless Moloney murine leukemia virus. The U3LacZ virus contains the longest reported LTR (3.4 kbp); nevertheless, *lacZ* sequences did not interfere with the ability of the virus to transduce a neomycin resistance gene expressed from an internal promoter. Duplication of the LTR placed *lacZ* sequences in the 5′ LTR just 30 nucleotides from the flanking cellular DNA. Approximately 0.4% of integrated proviruses expressed β-galactosidase as judged by 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) staining, and individual clones expressing *lacZ* were isolated by fluorescence-activated cell sorting. In all clones examined, β-galactosidase expression resulted from the fusion of *lacZ* sequences to transcriptional promoters located in the flanking cellular DNA. Furthermore, by differential sorting of neomycin-resistant cell populations, clones were isolated in which *lacZ* expression was induced and repressed in growth-arrested and log phase cells, respectively.

Movable elements have been used to analyze transcriptionally active chromosomal regions in both prokaryotes (4) and eukaryotes (2, 3, 5, 6, 17, 29). These strategies involve (i) integration of a reporter gene or selectable marker into a large collection of chromosomal sites, (ii) selection of cell clones in which the transduced marker gene is constitutively or conditionally expressed, and (iii) characterization of nearby cellular sequences for the appropriate transcription-promoting activity. The process enables transcriptional promoters and displaced cellular genes to be isolated by first monitoring for patterns of marker gene expression.

Like other movable elements, retroviruses integrate at nearly random sites in the mammalian genome (8, 30, 35), and recombination involves the precise recognition and conservation of specific viral sequences (8, 35). Unfortunately, sequences within the long terminal repeats (LTRs), in particular, the viral enhancer, polyadenylation signal, and multiple initiation and termination codons, limit the ability of retroviruses to induce functional gene fusions. As a consequence, retroviruses have been used only to a limited extent, for example (i) as enhancer traps (2, 15) or (ii) as gene traps which rely on RNA splicing to remove intervening viral sequences (6, 15).

In view of these issues, we previously developed a new type of retrovirus vector (U3His) to induce functional gene fusions in mammalian cells. Coding sequences for a selectable marker (histidinol dehydrogenase, *hisD*) were inserted into the U3 region of an enhancerless Moloney murine leukemia virus. The elongated U3 region participated in the formation of LTRs such that virus integration placed *hisD* sequences in the 5′ LTR just 30 nucleotides (nt) from the flanking cellular DNA. Selection for histidinol resistance generated cell clones in which the *hisD* gene in the 5′ LTR was expressed on transcripts initiating nearby in the flanking cellular DNA (39). Analysis of upstream sequences, cloned after amplification by inverse polymerase chain reaction, revealed that the U3His vector usurps promoters that were active prior to integration and thus functions as a promoter trap (38).

Since the U3His virus can identify only those promoters that are active at the time of selection, the vector is not well suited for isolating regulated promoters. For this reason, we have investigated the properties of a promoter-trap vector (U3LacZ) containing the *Escherichia coli lacZ* gene inserted in U3. *lacZ* offers several advantages as a reporter of gene fusions. First, β-galactosidase expression is readily monitored by enzymatic assays and by histochemical staining. Second, viable cells expressing β-galactosidase can be isolated by flow cytometry (25). The U3LacZ virus also contains a neomycin resistance gene expressed from an internal promoter. Selection for neomycin resistance provides a measure of virus titers and allows provirus containing clones to be isolated regardless of β-galactosidase expression.

The present study shows that the U3LacZ virus functions as a promoter trap in a manner similar to U3His. Although the U3LacZ vector contains the longest reported LTR (3.4 kbp), *lacZ* sequences did not interfere with the ability of the virus to be passaged into recipient cells. Moreover, cell clones expressing β-galactosidase invariably contained proviruses in which the 5′ copy of *lacZ* was transcribed from upstream cellular promoters. In principle, the U3LacZ vector can be used to isolate regulated promoters and to study temporal and spatial patterns of gene expression in vivo (1, 13, 15, 21, 29). For example, in the present study, the U3LacZ vector was used to isolate cell lines in which *lacZ* expression was induced and repressed in growth-arrested and log-phase cells, respectively.

MATERIALS AND METHODS

Plasmids. pGgTKNeoU3LacZen(−) was derived from pGgTKneoU3His (39) by replacing the *hisD* gene with *E. coli lacZ* coding sequences obtained as a 3.1-kbp *Hin*dIII-*Xba*I fragment from pSDKLacZ (8a). pSDKLacZ contains a promoterless *lacZ* gene in which sequences encoding the first eight N-terminal amino acids have been replaced by a

* Corresponding author.

22-nt segment containing the Shine-Dalgarno (31, 33) and Kozak (19) consensus sequences.

Cells and viruses. Cell lines expressing packaging-defective ecotropic (ψ2) (23) and amphotropic (PAC317) (24) helper viruses were transfected with 10 μg of pGgTKNeoU3 LacZen(−) and selected in G418. Production of virus stocks and infection of NIH 3T3 cells were carried out as described previously (39).

Analysis of β-galactosidase activity. β-Galactosidase expression was monitored by staining fixed cells with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). Cells were typically stained for 4 h to prevent background staining resulting from lysosomal β-galactosidase (41). β-Galactosidase activity was quantified as described by Norton (26) using *o*-nitrophenyl-β-D-galactoside (ONPG) as a substrate. For fluorescence-activated cell sorter (FACS) analysis, U3LacZ-infected NIH 3T3 cells were treated with fluorescein di-β-D-galactopyranoside (FDG; Molecular Probes) and sorted according to fluorescence intensity into $LacZ^+$ and $LacZ^-$ cells by using a Becton-Dickinson FACS Star Plus cell sorter with a 530-nm band-pass filter as described previously (25).

Nucleic acid blot hybridizations. The structure of the proviruses in $LacZ^+$ and $LacZ^-$ NIH 3T3 cell lines was analyzed by Southern blot hybridization as previously described (39). Polyadenylated [poly(A)$^+$] RNA was prepared from $LacZ^+$ and $LacZ^-$ clones by oligo(dT)-Sepharose chromatography, fractionated in 1% formaldehyde–agarose gels, and transferred to Nitroplus 2000 hybridization membranes (Micron Separations Inc.) as described previously (27). ^{32}P-labeled probes were prepared (11) from restriction endonuclease fragments of pGgTKNeoU3LacZen(−) corresponding to *lacZ* (3-kbp *Nhe*I-*Eco*RI fragment), *neo* (1.2-kbp *Bgl*II-*Bam*HI fragment), or *gag* (1.1-kbp *Pvu*I-*Xho*I fragment).

RNase protection assay. Cellular RNA (30 μg) was hybridized at 55°C for 12 h to ^{32}P-labeled RNA probes as previously described (38, 39). Probes complementary to the provirus coding strand were prepared by using T3 RNA polymerase to transcribe a 643-nt *Bam*HI-*Hpa*I fragment of pGgTKNeoU3LacZen(−) cloned into Bluescript KS(−) (Stratagene). After hybridization, samples were digested with RNase A (5 μg/ml) and RNase T_1 (2 μg/ml) (Boehringer Mannheim Biochemicals) and processed for gel electrophoresis as previously described (38, 39). Protected fragments were separated on 6% polyacrylamide–8.3 M urea gels and visualized by autoradiography.

Immunoprecipitation. Cells (10^6) were placed in methionine-free Dulbecco minimal essential medium (Flow Laboratories)–1% fetal calf serum for 30 min and labeled for 10 min with [^{35}S]methionine (ICN Radiochemicals) at 0.1 mCi/ml. The cells were washed in PBS (140 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4) and lysed for 10 min at 4°C in 1 ml of Triton lysis buffer (150 mM NaCl, 50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.5], 5 mM EDTA, 1% Triton X-100) supplemented with antipain, leupeptin, aprotinin, chymostatin, and pepstatin (Sigma), each at 5 μg/ml. β-Galactosidase was immunoprecipitated from aliquots containing 5×10^6 trichloroacetic acid-precipitable counts by using an anti-β-galactosidase antibody (Promega). Immune complexes were collected by adding 15 μl of protein A-agarose beads (Repligen) for 1 h at 4°C. After three washes with lysis buffer supplemented with 0.1% sodium dodecyl sulfate (SDS), immunoprecipitated proteins were electrophoresed in SDS–8% polyacryamide gels (20) and visualized by fluorography.

RESULTS

***lacZ* sequences in U3 do not interfere with virus infectivity.** The U3LacZ vector pGgTKNeoU3LacZen(−) was derived from pGgTKneoU3His (39) by replacing the *hisD* gene with *E. coli lacZ* coding sequences and contains a *neo* gene under the transcriptional control of the herpes simplex virus type 2 thymidine kinase promoter to provide an independent measure of virus titers. Cell lines producing ecotropic and amphotropic U3LacZ viruses were generated by transfecting 10 μg of pGgTKNeoU3LacZen(−) into ψ2 (23) and PAC317 (24) lines, respectively. Titers of the U3LacZ virus were high (2×10^6 to 4×10^6 G418-resistant CFU/ml/10^7 ψ2 clones and 0.5×10^6 to 1×10^6 G418-resistant CFU/ml/10^7 PAC317 clones), suggesting that the extra 3.1-kbp *lacZ* sequence inserted into the LTR did not impair the ability of the virus to be passaged into recipient cells.

Most U3LacZ proviruses do not express β-galactosidase. NIH 3T3 cells were infected at a multiplicity of infection of 1 Neo^r CFU per cell, and after 14 days in selection, *lacZ* expression in Neo^r colonies was monitored by staining with X-Gal (Fig. 1). Approximately 0.6% of 6,770 Neo^r colonies examined expressed detectable β-galactosidase and stained blue with X-Gal, whereas Neo^r colonies isolated after infection with a control (U3His) virus failed to stain (Fig. 1).

To enrich for β-galactosidase-expressing cells, pools of Neo^r clones were treated with FDG and sorted according to fluorescence intensity into $LacZ^+$ and $LacZ^-$ cell populations (25). $LacZ^+$ cells were expanded in culture and resorted. The fluorescence profiles of U3LacZ-infected cells during the first and second FACS cycles are shown in Fig. 2C and D, respectively. By using the same sorting strategy with uninfected NIH 3T3 cells, no significant enrichment for cells with high endogenous β-galactosidase was seen (Fig. 2A and B), indicating that recovery of highly fluorescent cells depended on prior infection with U3LacZ. Approximately 70% of clones derived from $LacZ^+$ cells stained with X-Gal, whereas none of the $LacZ^-$ clones stained (Fig. 1). Thus, the overall enrichment through two FACS cycles was 117-fold. On average, $LacZ^+$ clones expressed approximately 10-fold-higher levels of β-galactosidase than $LacZ^-$ clones did (Fig. 3).

***lacZ* sequences in U3 are duplicated as part of the LTR.** Structures of the integrated U3LacZ proviruses in each of four $LacZ^+$ and $LacZ^-$ clones were analyzed by Southern blot hybridization. Cleavage of DNA from U3LacZ-infected clones with *Cla*I (Fig. 4B) generated a 6.6-kbp fragment, as expected for proviruses flanked by 3.1-kb *lacZ* sequences duplicated as part of U3. Variable-sized fragments in different clones originate from *Cla*I sites in the flanking cellular DNA extending to the *Cla*I site in *lacZ*.

Analysis of cellular DNAs digested with *Hin*dIII (an enzyme that does not cut within the provirus) revealed that each clone was an independent isolate (Fig. 4C). The number of proviruses in U3LacZ-infected Neo^r clones (half contained one provirus, whereas the other half contained two proviruses) compares favorably with the value expected, according to a Poisson distribution (10), following infection at a multiplicity of infection of 1 Neo^r CFU per cell and suggests that most proviruses are able to transduce G418 resistance.

***lacZ* expression is not an intrinsic property of the provirus.** Several experiments suggest that the ability to express β-galactosidase is acquired as a result of integration and is not an intrinsic property of the provirus. First, virus producer lines expressed low levels of β-galactosidase (Fig. 3),

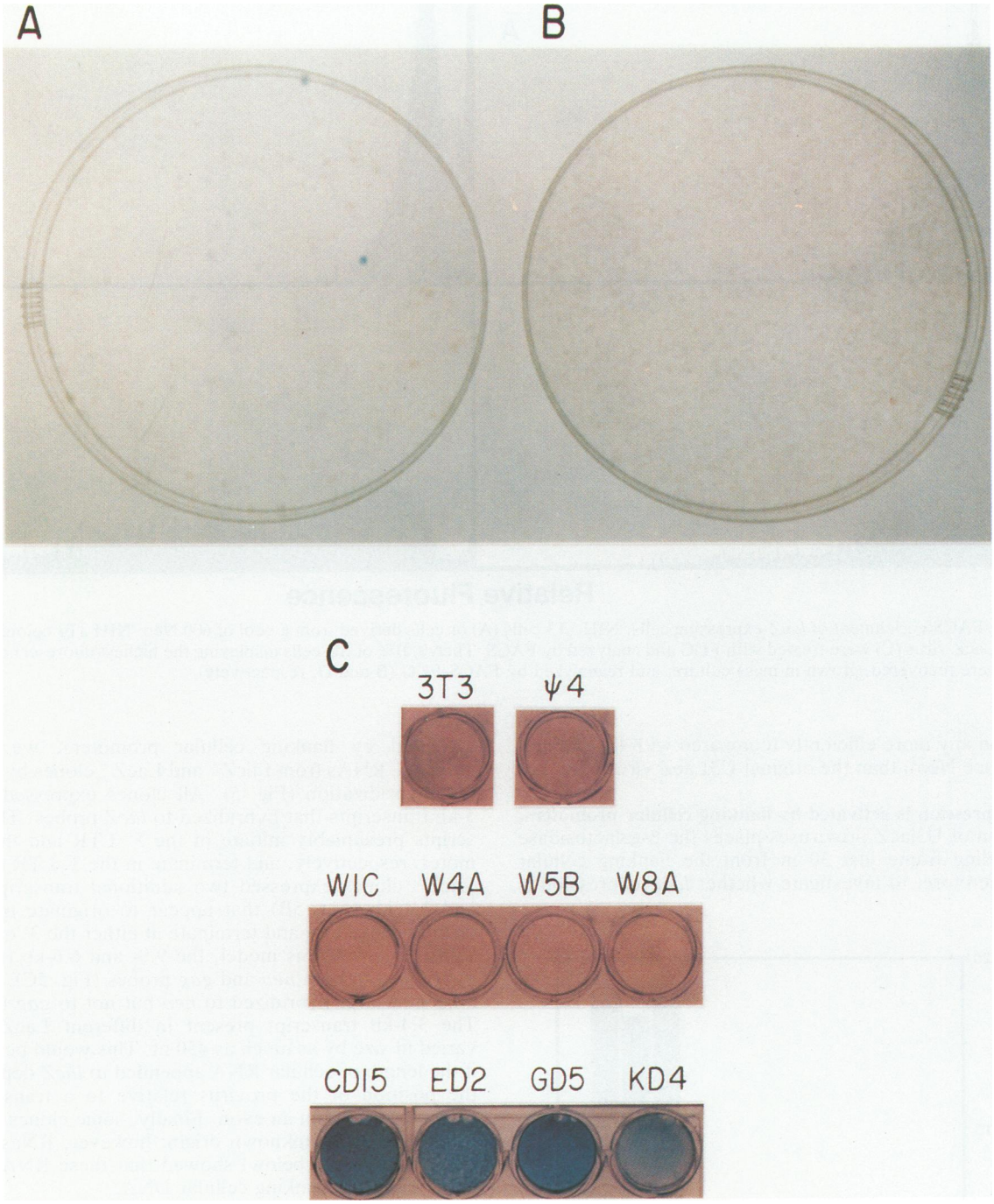

FIG. 1. Expression of β-galactosidase in NIH 3T3 cells infected with U3LacZ viruses. NIH 3T3 cells were infected with U3LacZ (A) or U3His (B) viruses, selected in G418 for 14 days, and stained with X-Gal. $LacZ^+$ (blue) colonies were observed only in cultures infected with U3LacZ. (C) *lacZ* expression in individual cell clones recovered after treatment with FDG and sorting by FACS: 3T3, parental NIH 3T3 cells; ψ4, U3LacZ ψ2 producer cells; W1C, W4A, W5B, and W8A, $LacZ^-$ clones recovered after sorting for low fluorescence; CD15, ED2, GD5, and KD4, $LacZ^+$ clones recovered after sorting for high fluorescence.

indicating that translation of 3′ *lacZ* sequences did not occur at a significant level. Second, lack of β-galactosidase expression in $LacZ^-$ clones was phenotypically stable. Thus, all four $LacZ^-$ clones failed to produce $LacZ^+$ colonies when plated at clonal densities (frequency less than 10^{-4}). Loss of β-galactosidase expression occurred more frequently in $LacZ^+$ clones: 0.2 to 1% of colonies derived from four $LacZ^+$ cell lines failed to stain with X-Gal. Third, β-galactosidase expression was not apparently activated by mutations, since the proviruses in $LacZ^+$ lines lacked gross structural rearrangements (Fig. 4). Finally, the viruses rescued from $LacZ^+$ clones did not transduce β-galactosidase

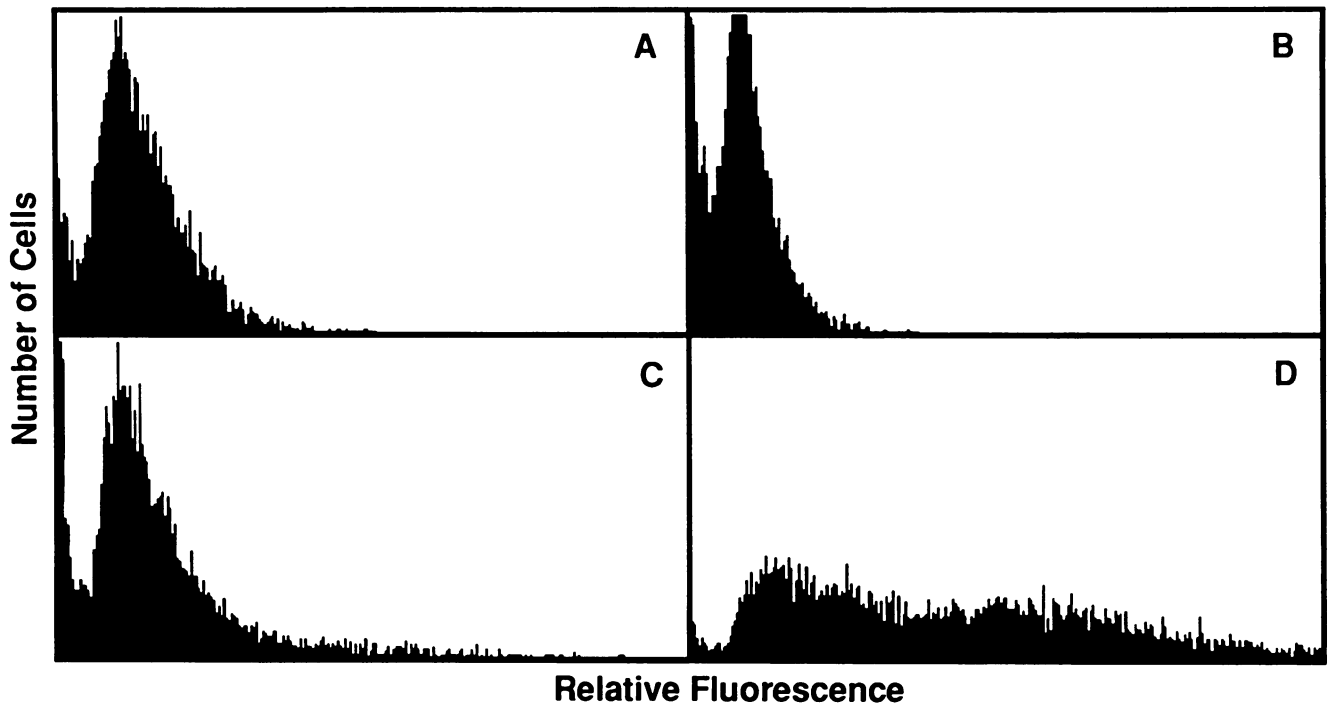

FIG. 2. FACS enrichment of *lacZ*-expressing cells. NIH 3T3 cells (A) or cells derived from a pool of 600 Neo^r NIH 3T3 colonies induced by the U3LacZ virus (C) were treated with FDG and analyzed by FACS. Then 0.01% of the cells displaying the highest fluorescence in panels A and C were recovered, grown in mass culture, and reanalyzed by FACS-FDG (B and D, respectively).

expression any more efficiently (compared with their ability to transduce Neo^r) than the original U3LacZ virus (data not shown).

***lacZ* expression is activated by flanking cellular promoters.** Integration of U3lacZ proviruses places the β-galactosidase open reading frame just 30 nt from the flanking cellular DNA. Therefore, to investigate whether *lacZ* expression is activated by flanking cellular promoters, we analyzed poly(A)$^+$ RNAs from LacZ$^+$ and LacZ$^-$ clones by Northern blot hybridization (Fig. 5). All clones expressed 6.6- and 5-kb transcripts that hybridized to *lacZ* probes. These transcripts presumably initiate in the 5′ LTR and the *tk* promoter, respectively, and terminate in the 3′ LTR (Fig. 5A). LacZ$^+$ clones expressed two additional transcripts of 9.9 and 3.4 kb (Fig. 5B) that appear to originate in flanking cellular sequences and terminate at either the 3′ or 5′ LTR. Consistent with this model, the 9.9- and 6.6-kb transcripts hybridized to both *neo* and *gag* probes (Fig. 5C), while the 5-kb transcript hybridized to *neo* but not to *gag* (Fig. 5D). The 3.4-kb transcript present in different LacZ$^+$ clones varied in size by as much as 450 nt. This would be expected if the length of cellular RNA appended to *lacZ* depended on the position of the provirus relative to a transcriptional promoter or within an exon. Finally, some clones exhibited RNA species of unknown origin; however, RNase protection analysis (see below) showed that these RNAs are not initiating in the flanking cellular DNA.

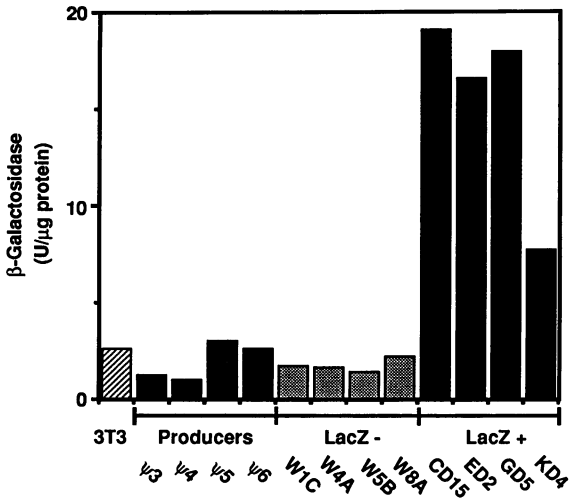

FIG. 3. β-Galactosidase expression in NIH 3T3 cell clones. Abbreviations: 3T3, parental NIH 3T3 cells; ψ3, ψ4, ψ5, ψ6, clones of ψ2 cells producing the U3LacZ virus; W1C, W4A, W5B, and W8A, LacZ$^-$ clones; CD15, ED2, GD5, and KD4, LacZ$^+$ clones. One unit (U) of β-galactosidase is equivalent to 1 nM ONPG hydrolyzed per min at 37°C.

To determine whether transcripts in LacZ$^+$ clones initiated within the flanking cellular DNA, we analyzed total RNA by an RNase protection assay with probes complementary to the provirus coding strand. RNA from all clones protected a 643-nt fragment from polycistronic transcripts containing both *neo* and 3′ *lacZ* sequences. However only RNA from LacZ$^+$ clones protected a 501-nt fragment, the size expected for hybrid cell-virus transcripts extending through the 5′ end of the LTR (Fig. 6).

LacZ$^+$ clones do not express fusion proteins. Sequence analysis revealed that the translational reading frame upstream of *lacZ* is not blocked by termination codons (data not shown). Consequently, β-galactosidase expression could result when the provirus integrates next to a promoter or in

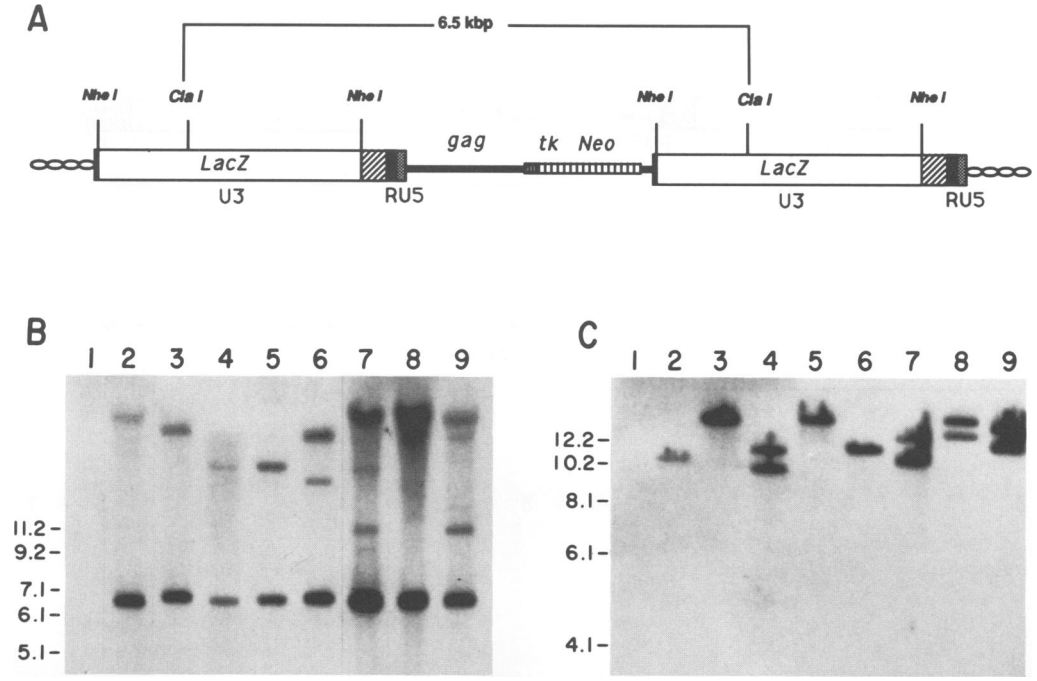

FIG. 4. Southern blot analysis of U3LacZ proviruses in NIH 3T3 cell lines. (A) Structure of the U3LacZ provirus. (B and C) Cell DNAs were digested with *Cla*I (panel B) or *Hin*dIII (panel C), fractionated on 1% agarose gels, blotted, and hybridized to *lacZ* probes as follows: lane 1, NIH 3T3; lanes 2 to 5 LacZ$^-$ clones W1C, W4A, W5B, and W8A; lanes 6 to 9, LacZ$^+$ clones CD15, ED2, GD5, and KD4.

the appropriate translational reading frame in an exon. However, the electrophoretic mobilities of β-galactosidase proteins immunoprecipitated from LacZ$^+$ clones were similar to native *E. coli* β-galactosidase, indicating that most activating gene fusions did not append cotranslated sequences to the 5′ end of *lacZ* (Fig. 7).

Isolation of cell clones with inducible *lacZ* expression. In principle, differential sorting of neomycin-resistant clones provides a means to isolate cells in which *lacZ* is fused to regulated promoters. We were particularly interested in isolating cells in which *lacZ* expression could be induced by changes in the growth state of the cell. BALB 3T3 cells were infected with the U3LacZ virus (one Neo^r CFU per cell), and a collection of 2×10^4 Neo^r clones was selected, grown in mass culture, and arrested at confluence in 2% calf serum for 10 days, and cells displaying the highest fluorescence (highest 0.1 percentile; quiescent:LacZ$^+$) were selected by FACS-FDG. Quiescent:LacZ$^+$ cells were expanded in culture and cells displaying the lowest fluorescence (lowest 30 percentile; log:LacZ$^-$) were sorted from exponentially growing cell cultures. Finally, quiescent:LacZ$^+$/log:LacZ$^-$ cells were growth arrested, resorted for LacZ$^+$ cells (highest 0.1 percentile), and plated at clonal densities. A total of 96 clones were expanded and tested individually for differential β-galactosidase expression; 30 expressed a quiescent:LacZ$^+$ phenotype as judged by X-Gal staining, and in 2 of the 10 clones chosen for further analysis, *lacZ* expression was induced and repressed in growth-arrested and log-phase cells, respectively (Fig. 8). In contrast, *lacZ* was not significantly induced in three randomly selected log:LacZ$^-$ clones (Fig. 8 and data not shown).

DISCUSSION

A retrovirus promoter-trap vector has been described in which the *E. coli lacZ* gene provides a marker for instances in which the virus integrates near cellular promoters. At 3.4 kbp, the U3LacZ virus contains the largest reported LTR, considerably longer than the normal Moloney murine leukemia virus LTR (0.59 kbp) or the 2.7 kbp of the nondefective Moloney murine leukemia virus LTR containing adenosine deaminase sequences (14). Since U3LacZ titers were quite high, the 3.1-kbp *lacZ* sequence affected neither the generation of LTRs during reverse transcription nor integration of the provirus. Although the length of foreign sequences tolerated by LTRs is unknown, we suspect that only packaging constraints of the virion will limit the maximum length of sequences tolerated within U3.

Although provirus transcripts normally extend through the site of polyadenylation in the 5′ LTR, transcripts initiating in the flanking cellular DNA terminate at both the 5′ and 3′ LTRs. This result is consistent with the observation that U3 sequences are necessary for RNA processing (9). Approximately 1 in 200 of all Neo^r colonies induced by U3LacZ expressed detectable levels of β-galactosidase. This ratio reflects the fact that the *neo* gene is expressed from an autonomous promoter, whereas *lacZ* expression requires integration near a cellular promoter. Compared with the U3His virus (39), the U3lacZ vector is 10-fold more efficient at inducing functional gene fusions. Two mechanisms may contribute to the relative ease of activating *lacZ* expression. First, as judged by Northern (RNA) blot analysis, the level of β-galactosidase required for detection appears to be lower than the level of histidinol dehydrogenase required for histidinol resistance. This would allow *lacZ* to identify

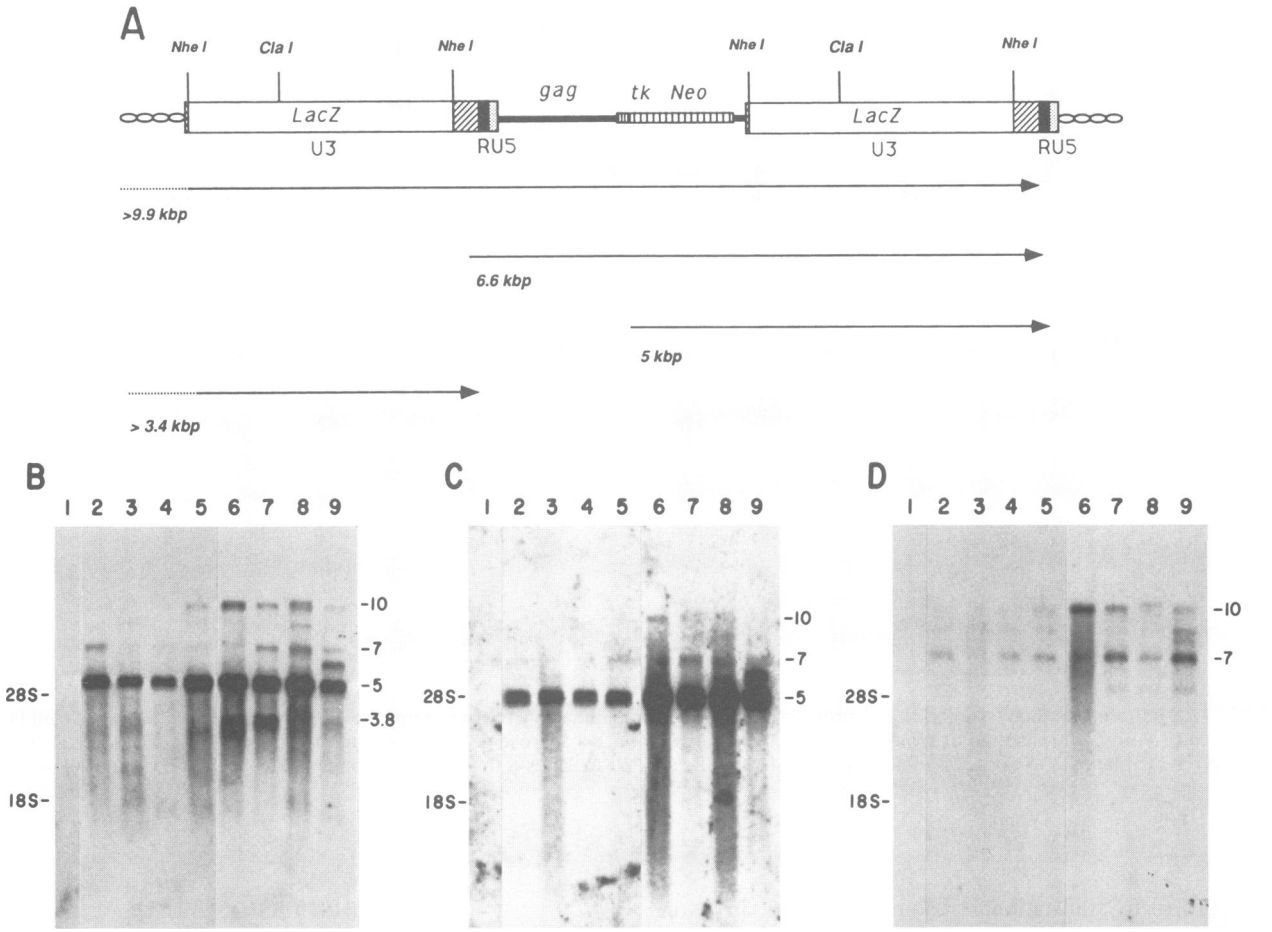

FIG. 5. Northern blot analysis of provirus transcripts. (A) U3LacZ provirus structure and predicted transcripts. (B to D) Poly(A)$^+$-selected RNA (2.5 μg) was fractionated on 1% formaldehyde-agarose gels, transferred to nylon membranes, and hybridized to *lacZ*- (panel B); *neo*- (panel C), or *gag* (panel D)-specific probes as follows: lane 1, NIH 3T3; lanes 2 to 5, LacZ$^-$ clones W1C, W4A, W5B, and W8A; lanes 6 to 9, LacZ$^+$ clones CD15, ED2, GD5, and KD4.

weaker promoters, assuming no offsetting differences in protein or RNA stability. Second, cellular sequences appended to *lacZ* transcripts were longer than hybrid *hisD* transcripts, suggesting that the size of each site capable of activating U3LacZ is larger than for U3His. This implies that detectable levels of *lacZ* can be expressed even when the initiating AUG codon for β-galactosidase is downstream of other initiation codons (18). Finally, unlike *hisD*, *lacZ* coding sequences are not blocked by upstream termination codons; therefore, expression of β-galactosidase fusion proteins (7) could result from integrations in exons. However, β-galactosidase fusion proteins were not detected in any of the four LacZ$^+$ clones examined. The reasons for this are not known, but similar results were obtained with gene-trap retroviruses, suggesting that *lacZ* activation preferentially involves integrations near the 5′ end of genes (6).

It is possible to estimate the maximum number of sites which permit *lacZ* expression in NIH 3T3 cells. The total integration target for *neo* transduction cannot exceed the size of the genome (3×10^9 nt), implying that the target permitting *lacZ* expression is smaller than 1.5×10^7 nt. Dividing the total integration target by the average size of each integration site (i.e., 450 nt [the average length of cellular sequence appended to *lacZ* as estimated by Northern blot analysis]) yields a maximum of 3×10^4 sites capable of promoting detectable levels of β-galactosidase. This value falls within the range of active genes (2×10^4 to 4×10^4) as estimated by RNA renaturation kinetics (22).

This calculation assumes (i) that all cells in which U3LacZ integrates will also acquire neomycin resistance, (ii) that most of the genome is a target for integration, and (iii) that retroviruses integrate randomly. These assumptions, although unproven, appear largely valid. Thus, although *neo* expression relies on the relatively weak *tk* promoter (40), the number of proviruses in Neo^r clones follow a distribution appropriate for the multiplicity of infection. Although several studies suggest that integration is biased for transcriptionally active chromatin (6, 28, 37), another suggests that 80% of all integrations are random and involve most of the genome (30). Likewise, the low efficiency of insertional mutagenesis (at expressed hemizygous loci) in cultured cells is consistent with a large integration target (16, 36). Even the apparently high proportion (5%) of integration events that induce mutations in mice need not imply that integration is restricted, since 3×10^4 phenotypically significant genes, averaging 5,000 nt in size, would occupy 5% of the genome (32, 34, 42).

In summary the U3LacZ promoter-trap virus provides an

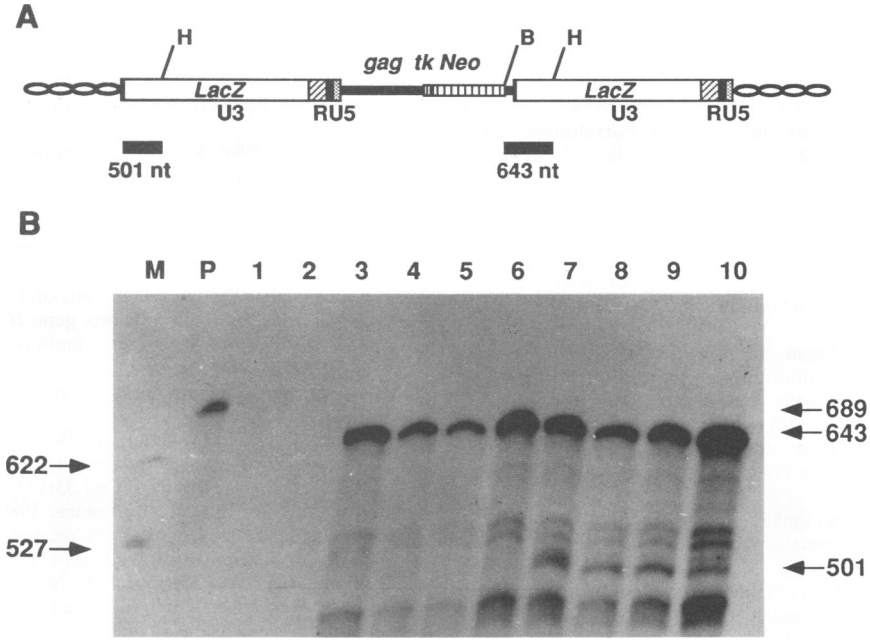

FIG. 6. RNase protection analysis of provirus transcripts. (A) Transcripts extending through the 5′ and 3′ LTRs should protect 501 and 643 nt, respectively, of a 689-nt probe that spans the 643-nt region between the provirus *Bam*HI (B) and *Hpa*I (H) sites. (B) The following RNAs were analyzed: yeast tRNA (lane 1); NIH 3T3 cells (lane 2); LacZ$^-$ clones W1C, W4A, W5B, and W8A (lanes 3 to 6); and LacZ$^+$ clones CD15, ED2, GD5, and KD4 (lanes 7 to 10). Lanes corresponding to end-labeled DNA molecular weight markers (M) and undigested probe (P) and the positions of molecular weight markers (left) and mobilities of expected fragments (right) are also indicated.

efficient means of inducing functional *lacZ* gene fusions in mammalian cells. Differential sorting of Neo^r clones for β-galactosidase expression may be used to isolate regulated promoters. For example, in the present study, promoters repressed by growth-promoting stimuli could be identified by sequential sorting of quiescent and serum-stimulated cells for changes in β-galactosidase expression. Cell lines containing reporter genes linked to regulated promoters may also be used to identify *trans*-acting factors which modulate expression and to assay for potential drugs which influence signal transduction pathways. Finally, it should be possible to study temporal and spatial patterns of gene expression in vivo by constructing transgenic mice from embryonal stem cells infected with the U3LacZ virus.

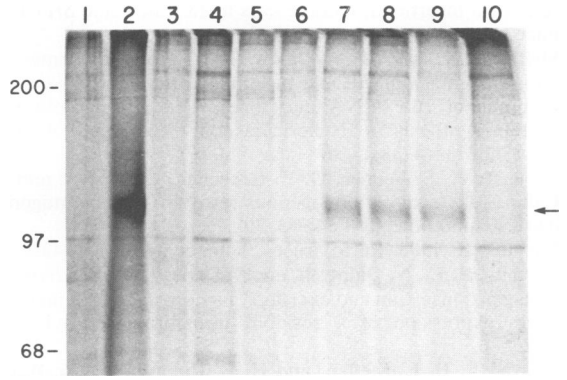

FIG. 7. β-Galactosidase expression in LacZ$^+$ and LacZ$^-$ clones. Cells were labeled with [^{35}S]methionine and immunoprecipitated with anti-β-galactosidase antibody as follows: NIH 3T3 cells (lane 1); B6, a control cell line expressing *E. coli lacZ* (12) (lane 2); LacZ$^-$ clones W1C, W4A, W5B, and W8A (lanes 3 to 6); and LacZ$^+$ clones CD15, ED2, GD5, and KD4 (lanes 7 to 10).

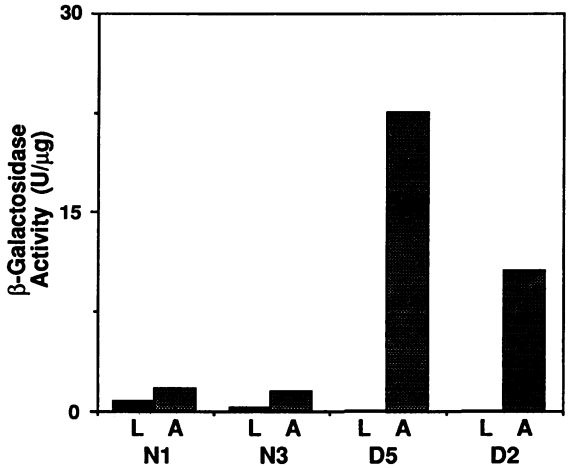

FIG. 8. Growth-dependent regulation of *lacZ* expression in BALB 3T3 cells. β-Galactosidase enzyme activities were measured in log-phase (L) or growth-arrested (A) cells. N1 and N3 are randomly selected Neo^r clones isolated following infection with the U3LacZ virus; D5 and D2 are clones isolated after sorting for cell clones expressing a quiescent:LacZ$^+$/log:LacZ$^-$ phenotype. One unit (U) of β-galactosidase is equivalent to 1 nM ONPG hydrolyzed per min at 37°C.

ACKNOWLEDGMENTS

We thank Stewart Connor for assistance with the FACS; K. Hunter and N. Speck for providing helper Moloney murine leukemia

virus; P. Robbins, S. Darling, and J. Rossant for the pSDKLacZ plasmid; and M. Marvin and R. McKay for the B6 cell line.

This work was supported by Public Health Service grant PO1CA42063 from the National Cancer Institute to H.E.R. and partially supported by Cancer Center core grant CA14051 to P. A. Sharp. H.M. was supported by the Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg, Federal Republic of Germany. H.E.R. is a Latham Family Professor.

REFERENCES

1. **Allen, N. D., D. G. Cran, S. C. Barton, S. Hettle, W. Reik, and M. A. Surani.** 1988. Transgenes as probes for active chromosomal domains in mouse development. Nature (London) **333:** 852–855.
2. **Barklis, E., R. C. Mulligan, and R. Jaenisch.** 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. Cell **47:**391–399.
3. **Bellofatto, V., L. Shapiro, and D. A. Hodgson.** 1984. Generation of a Tn5 promoter probe and its use in the study of gene expression in *Caulobacter crescentus*. Proc. Natl. Acad. Sci. USA **81:**1035–1039.
4. **Berg, C. M., D. E. Berg, and E. A. Groisman.** 1989. Transposable elements and the genetic engineering of bacteria, p. 879–925. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
5. **Bier, E., H. Vaessin, S. Shepard, K. Lee, K. McCall, B. Barbel, L. Ackerman, R. Carretto, T. Uemura, E. Grell, L. Y. Jan, and Y. N. Jan.** 1989. Searching for pattern and mutation in the Drosophila genome with a p-lacZ vector. Genes Dev. **3:**1273–1287.
6. **Brenner, D. G., S. Lin-Chao, and S. N. Cohen.** 1989. Analysis of mammalian cell genetic regulation *in situ* by using retrovirus-derived "portable exons" carrying the *Escherichia coli lacZ* gene. Proc. Natl. Acad. Sci. USA **86:**5517–5521.
7. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143:**971–980.
8. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1499. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath, and B. Roizman (ed.), Virology, 2nd ed. Raven Press, New York.

8a. **Darling, S., and J. Rossant.** Unpublished data.

9. **Dougherty, J. P., and H. M. Temin.** 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3′ RNA processing. Proc. Natl. Acad. Sci. USA **84:**1197–1201.
10. **Draper, N. R., and W. E. Lawrence.** 1970. Probability: an introductory course. Markham Publishing Co., Chicago.
11. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:**266–267.
12. **Frederiksen, K., P. S. Jat, N. Valtz, D. Levy, and R. McKay.** 1988. Immortalization of precursor cells from the mammalian CNS. Neuron **1:**439–448.
13. **Gossler, A., A. L. Joyner, J. Rossant, and W. C. Skarnes.** 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science **244:**463–465.
14. **Hantzopoulos, P. A., B. A. Sullenger, G. Ungers, and E. Gilboa.** 1989. Improved gene expression upon transfer of the adenosine deaminase minigene outside the transcriptional unit of a retroviral vector. Proc. Natl. Acad. Sci. USA **86:**3519–3523.
15. **Kerr, W. G., G. P. Nolan, A. T. Serafini, and L. A. Herzenberg.** 1989. Transcriptionally defective retroviruses containing *lacZ* for the *in situ* detection of endogenous genes and developmentally regulated chromatin. Cold Spring Harbor Symp. Quant. Biol. **54:**767–776.
16. **King, W., M. D. Patel, L. I. Lober, S. P. Goff, and M. C. Nguyen-Huu.** 1985. Insertion mutagenesis of embryonal carcinoma cells by retroviruses. Science **228:**554–558.
17. **Koncz, C., N. Martini, R. Mayerhofer, K.-K. ZsuZsanna, H. Korber, G. P. Redei, and J. Schell.** 1989. High-frequency T-DNA-mediated gene tagging in plants. Proc. Natl. Acad. Sci. USA **86:**8467–8471.
18. **Kozak, M.** 1984. Selection of initiation sites by eukaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucleic Acids Res. **12:**3873–3893.
19. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:**283–292.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
21. **Le Mouellic, H., Y. Lallemand, and P. Brulet.** 1990. Targeted replacement of the homeobox gene *Hox-3.1* by the *Escherichia coli lacZ* in mouse chimeric embryos. Proc. Natl. Acad. Sci. USA **87:**4712–4716.
22. **Lewin, B.** 1980. Genes, 1st ed., p. 292–299. John Wiley & Sons, Inc., New York.
23. **Mann, R., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell **33:**153–159.
24. **Miller, A. D., and C. Buttimore.** 1986. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell. Biol. **6:**2895–2902.
25. **Nolan, G. P., S. Fiering, J.-F. Nicolas, and L. A. Herzenberg.** 1988. Fluorescence-activated cell analysis and sorting of viable mammalian cells based on β-D-galactosidase activity after transduction of *Escherichia coli lacZ*. Proc. Natl. Acad. Sci. USA **85:**2603–2607.
26. **Norton, P. A., and J. M. Coffin.** 1985. Bacterial β-galactosidase as a marker of Rous sarcoma virus gene expression and replication. Mol. Cell. Biol. **5:**281–290.
27. **Reddy, S., K. Ozgur, M. Lu, W. Chang, S. R. Mohan, C. C. Kumar, and H. E. Ruley.** 1990. Structure of the human smooth muscle α-actin gene: analysis of a cDNA and 5′ upstream region. J. Biol. Chem. **265:**1683–1687.
28. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61:**336–343.
29. **Sanes, J. R., J. L. R. Rubenstein, and J.-F. Nicolas.** 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO J. **5:**3133–3142.
30. **Shih, C.-C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
31. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
32. **Spence, S. E., D. J. Gilbert, D. A. Swing, N. G. Copeland, and N. A. Jenkins.** 1989. Spontaneous germ line virus infection and retroviral insertional mutagenesis in 18 transgenic *Srev* lines of mice. Mol. Cell. Biol. **9:**177–184.
33. **Steitz, J. A., and K. A. Jakes.** 1975. How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. Proc. Natl. Acad. Sci. USA **72:**4734–4738.
34. **Stoye, J. P., S. Fenner, G. E. Greenoak, C. Moran, and J. M. Coffin.** 1988. Role of endogenous retroviruses as mutagens: the hairless mutation. Cell **54:**383–391.
35. **Varmus, H.** 1988. Retroviruses. Science **240:**1427–1435.
36. **Varmus, H. E., N. Quintrell, and S. Ortiy.** 1981. Retroviruses as mutagens: insertion and excision of a non-transforming provirus alters expression of a resident transforming provirus. Cell **25:**23–36.
37. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683–692.
38. **von Melchner, H., S. Reddy, and H. E. Ruley.** 1990. Isolation of cellular promoters using a retrovirus promoter trap. Proc. Natl. Acad. Sci. USA **87:**3733–3737.
39. **von Melchner, H., and H. E. Ruley.** 1989. Identification of

cellular promoters by using a retrovirus promoter trap. J. Virol. **63**:3227–3233.
40. **Wagner, M. J., J. A. Sharp, and W. C. Summers.** 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type I. Proc. Natl. Acad. Sci. USA **78**:1441–1445.
41. **Wallenfels, K., and R. Weil.** 1971. β-Galactosidase, p. 617–663. *In* P. Boyer (ed.), The enzymes. 3rd ed. Academic Press, Inc., New York.
42. **Weiher, H., T. Noda, D. A. Gray, A. H. Sharpe, and R. Jaenisch.** 1990. Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. Cell **62**:425–434.