METHODS: A Companion to Methods in Enzymology
Vol. 2, No. 3, June, pp. 261–271, 1991

LAH 352

Gene-Search Viruses and FACS-Gal Permit the Detection, Isolation, and Characterization of Mammalian Cells with *in Situ* Fusions between Cellular Genes and *Escherichia coli lacZ*

William G. Kerr and Leonard A. Herzenberg

Department of Genetics, Stanford University School of Medicine, Stanford, California 94305

We describe self-inactivating retroviruses that will activate expression of the reporter gene, *Escherichia coli lacZ*, upon integration within a cellular gene (Gensr1) or near a cellular enhancer (Enhsr1), referred to collectively as gene-search viruses. If the Gensr1 virus integrates within an intron of a gene, splicing of a *lacZ* neo-exon to an upstream cellular exon can occur, resulting in transcriptional and translational fusion between *E. coli lacZ* and a cellular gene. The Enhsr1 retrovirus generates a provirus that lacks the LTR enhancer region and thus is dependent upon flanking cellular enhancers to activate expression of *lacZ*. Fluorescence-activated cell sorting permits mammalian cells infected with the gene-search viruses that contain gene fusions between *lacZ* and cellular transcription control elements to be isolated as a population or as clones of single cells. Clones can be analyzed via a rapid, sensitive assay for β-galactosidase activity carried out in 96-well plates, permitting clones with integrations in conditionally expressed genes to be identified. This approach has led to isolation of gene-search virus integrations in developmentally regulated genes and loci. In addition, *lacZ*⁺ clones derived from Gensr1 infections can be screened via a histochemical stain (X-gal) for subcellular targetting of β-gal activity. Molecular characterization of such Gensr1 integrations could allow identification of mammalian proteins with specific subcellular localizations. Finally, we demonstrate the potential of gene-search viruses for obtaining expression of *lacZ* in normal cells both *in vitro* and *in vivo*. © 1991 Academic Press, Inc.

We have previously reported the development of a fluorogenic assay for β-galactosidase (FACS-Gal) that permits the detection and isolation of viable cells expressing *Escherichia coli lacZ* (*lacZ*) (1, 2). FACS-Gal permits the *lacZ* gene to be used as a selectable marker in eukaryotic cells. This new capability should aid the study of gene expression and development in higher eukaryotic organisms for which quantitative assays for gene expression at the single-cell level are lacking. Here we describe a novel methodology for the identification of transcriptionally active sites in mammalian chromatin that utilizes the FACS-Gal technique and *lacZ*-containing retroviruses.

Mobile genetic elements have been used to identify transcriptionally active sites in both eukaryotic (4–10) and prokaryotic (11) genomes. Introduction of transcriptionally disabled reporter gene constructs into mammalian cells has also proven useful for the detection of transcriptionally active sites in mammalian genomes (6, 12, 13). Because *lacZ* expression can be detected in whole embryos (4, 12, 13) or in single, viable cells (1, 2, 6), approaches utilizing *lacZ* as a reporter of expression have been particularly successful in identifying developmentally regulated chromosomal loci (4, 6, 10, 12, 13). Because *lacZ* expression is readily monitored (1–3, 6), both induction (4, 6, 10, 12, 13) and repression (6, 10) of transcriptionally active chromosomal sites are easily detected.

We have previously developed a series of self-inactivating retroviruses containing *lacZ* that activate expression of *lacZ* upon integration near a cellular enhancer (Enhsr1) or within a cellular gene (Gensr1) (6). The gene-search viruses (Enhsr1, Gensr1) have permitted the detection of genetic loci that respond to LPS-stimulated differentiation of a pre-B cell line (6, 10). Others have also developed *lacZ*-containing retroviruses capable of detecting cellular promoters (14) or genes (7). Here we review the use of gene-search retroviruses in concert with FACS-Gal to detect and isolate cells with integrations in transcriptionally active chromatin and genes. In addition, we discuss procedures for identifying integrations in genes with differential regulation patterns or whose gene products exhibit specific subcellular localization.

DESCRIPTION OF THE METHOD

Models for Expression of lacZ via the Gene-Search Retroviruses

In Figs. 1 and 2 we illustrate how the gene-search retroviruses self-inactivate their ability to drive transcription

1046-2023/91 $3.00
Copyright © 1991 by Academic Press, Inc.
All rights of reproduction in any form reserved.

from their 5′ LTR. This approach was originally described by Yu *et al.* (15). When copying viral RNA into DNA, reverse transcriptase uses the 3′ U3 as the template for the U3 region in both the 5′ and the 3′ viral LTR (16). Because a mutation or deletion can be placed in the 3′ U3 of a retroviral construct, infection of target cells will generate proviruses that have a transcriptional phenotype different from that of the parent construct. The gene-search retroviruses exploit this feature of retroviral biology because they generate proviruses that are transcriptionally incompetent relative to the parental construct.

In the case of Enhsrl, the retrovirus construct contains a deletion of the Moloney enhancer region in the 3′ LTR (Fig. 1). Following entry of Enhsrl viral particles into target cells, a provirus that lacks the Moloney enhancer region in either LTR will be generated. The only transcription control elements remaining in the LTR will be the CAAT and TATA motifs of the Moloney promoter. Because the promoter elements are no longer associated with the viral enhancer, the *lacZ* gene must rely on transcription control elements in the flanking cellular chromatin for efficient expression. Thus, *in situ* gene fusions with Enhsrl represent a fusion between cellular transcription control elements and the Moloney promoter driving *lacZ* (6, 10).

In a cell that has β-gal activity following infection with Gensrl, β-gal activity should result from transcriptional and translational fusion of a cellular gene and *lacZ* (Fig. 2). To ensure that β-gal activity is derived from a direct fusion of *lacZ* and cellular coding sequences, we developed the *lacZ* neo-exon, AcLac (6). The AcLac construct was derived by replacing the translation initiation codon (ATG) of *lacZ* with the splice acceptor region of the Moloney *env* gene (6). The splice acceptor region of the *env* gene is predicted to contain three separate splice acceptance points, with each splitting a translational codon at a different position in the triplet (17, 18). The *env* sequence following the three splice acceptors contains no initiation or termination codons and is in frame with *lacZ* (see Fig. 3). This feature enables AcLac to generate transcriptional and translational fusions of *lacZ* with any expressed cellular gene that contains an intron following a coding exon.

In Figs. 4 and 5 we illustrate the ability of the AcLac neo-exon to form gene fusions with cellular genes. Following stable transfection of the AcLac construct into 293 cells, $lacZ^+$ clones were found to contain β-gal fusion proteins (Figs. 4 and 5) and/or discrete subcellular localization of β-gal (Fig. 5). The presence of β-gal fusion proteins (Figs. 4 and 5) in multiple, independent AcLac stable transfectants validates our hypothesis that expression of

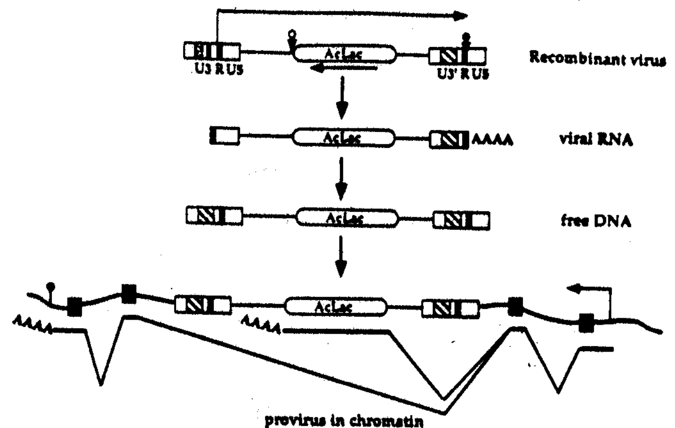

FIG. 2. Model for Gensrl. A flowchart that depicts the generation of Gensrl viral RNA, reverse transcription to a free DNA intermediate, and subsequent proviral integration in an intron of a transcriptionally active gene. A striped rectangle in the viral LTR indicates that the LTR contains a deletion of both the Moloney enhancer and the promoter regions, which results in a provirus that is transcriptionally inert. Horizontal arrows indicate transcriptional orientation; note that transcriptional orientation of AcLac is opposite that of the retroviral construct. We have inserted a unidirectional poly(A) site downstream of AcLac (vertical arrow with open circle) which will function as the poly(A) cleavage and addition site for the *in situ* fusion with a cellular gene. The vertical arrow with a filled circle represents the MLV genomic poly(A) site. The vertical line with filled circle represents the poly(A) site for the putative cellular gene. Construction of Gensrl was achieved as follows: A *Bam*HI-*Dra*I fragment from pAcLac-Δ (6, and Fig. 3) which contains the AcLac neo-exon was isolated and a *Xho*I linker was added to the *Dra*I-digested blunt end, generating a *Bam*HI-*Xho*I fragment. The resulting *Bam*HI-*Xho*I fragment was ligated into *Bam*HI-*Xho*I-digested pJrPro⁻. The resulting plasmid, pArgus⁻, was digested with *Xho*I and the ends were blunted with T4 polymerase. A 400-bp *Sca*I-*Sma*I fragment containing the adeno-associated virus poly(A) site (30, 31) was then ligated with *Xho*I/blunted pArgus⁻. A clone containing the *Sca*I-*Sma*I fragment in the correct orientation was isolated and is referred to as pGensrl. For tranfection of pGensrl into producer cells the plasmid was linearized with *Hin*dIII.

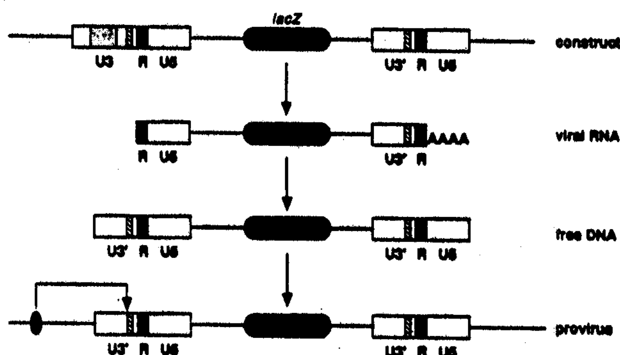

FIG. 1. Schematic depiction of Enhsrl and model for *lacZ* expression. This figure depicts how a *lacZ*-encoding provirus that lacks the Moloney leukemia virus enhancer region (stippled box) would be generated from the Enhsrl construct. The provirus generated from Enhsrl will lack the viral enhancer region (stippled box) but still retain CAAT and TATA box motifs (hatched box). This defective proviral transcriptional unit can be activated by an enhancer element (stippled oval) in the flanking, endogenous chromatin. Thus, with Enhsrl, viral integrations near endogenous enhancer elements will result in *lacZ* expression.

lacZ is derived by transcriptional and translational fusion with a cellular gene via RNA splicing.

Generation of Gene-Search Producer Cell Lines

Both ecotropic and amphotropic producers of the Enhsrl and Gensrl viruses have been generated. Ecotropic producers were generated in the producer cell line, ψ2 (19), whereas amphotropic producers were generated in PA317 cells (20). Since the gene-search viruses lack a drug-selectable marker, clones producing virus were derived by co-transfection of the appropriate gene-search virus construct along with SV2*neo* (21) at a molar ratio of 5:1, respectively (6). Following drug selection for G418-resistant stable transfectants, the G418-resistant cells were single-cell-cloned via FACS. Because G418-resistant cells that produce Enhsrl genomic RNA should also express β-gal, cells producing Enhsrl virions were selected by FACS-Gal. When these FACS-selected clones grew up, the relative titer of their supernatants was assessed by an RNA slot blot screen for *lacZ*-hybridizing viral RNA. The ecotropic (ψ2/E2) and amphotropic (PA/E25) clones that gave the strongest *lacZ* hybridization signal were used in all subsequent Enhsrl infections.

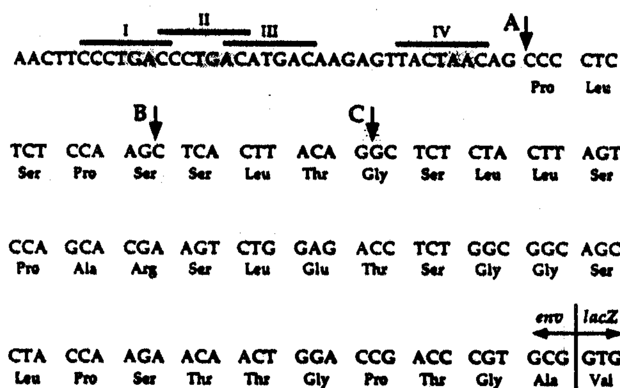

FIG. 3. Sequence of the AcLac neo-exon involved in generation of transcriptional/translational fusion between *lacZ* and cellular genes. The sequence of the *env* splice acceptor–*lacZ* construct showing the four consensus lariat branch points (solid bars, I through IV) and the three potential consensus splice acceptor intron/exon boundaries (arrows, A through C). The potential branch site–acceptor site pairs are I with A, II with A or B, III with A or B, and IV with C (17, 18). Splice donation from an upstream exon to one of the three potential splice acceptor sites will fuse a cellular exon to the AcLac sequence following these sites. The translated peptide sequence of the splice acceptor *lacZ* region is shown below the nucleotide codons. In frame translational stop codons upstream of the splice acceptor sites are indicated by shaded boxes. The junction of *env* and *lacZ* sequences is as indicated. The AcLac neo-exon in Gensrl has been modified from that used in the initial transfection experiments in 293 cells (6) such that an ATG found upstream of the splice sites is now out of frame with *lacZ*. To accomplish this modification, the pAcLac plasmid was digested with *Kpn*I and the staggered ends were removed by treatment with T4 polymerase. The blunt-ended pAcLac was religated and the expected deletion confirmed by sequencing. This plasmid is referred to as pAcLac-Δ and was used for the construction of Gensrl.

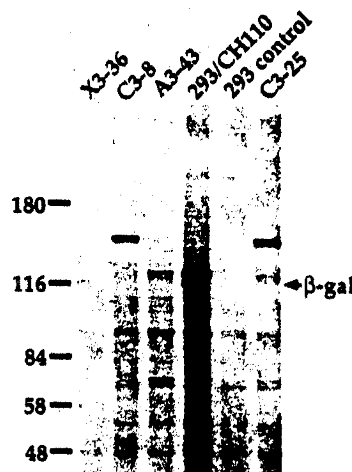

FIG. 4. Clonal variation of β-gal fusion protein size in cells harboring fusions between the AcLac neo-exon and a cellular gene. Immunoprecipitation analysis of β-galactosidase fusion proteins in four representative AcLac/293 clones (X3-36, C3-8, A3-43, C3-25) shows different apparent molecular weights greater than that of *E. coli* β-galactosidase (revealed by Coomassie staining; indicated by an arrow). Purified *E. coli* β-galactosidase (Sigma, St. Louis, MO) (5.5 μg) was added to the labeled cell lysate of 293 cells (293 control) and immunoprecipitated as described above. As a further control, 293 cells stably expressing CH110 (*trpA*-*lacZ* fusion protein) were also analyzed by immunoprecipitation analysis (293/CH110). Migration of molecular weight standards is indicated on the left.

Because producer cells expressing Gensrl genomic RNA will not express *lacZ*, random single-cell clones were made from the pool of G418-resistant stable transfectants. These clones were then screened by the RNA slot blot procedure. Clones that showed a significant *lacZ* hybridization signal on the RNA slot blot were further screened for their ability to infect target cells and transduce β-gal expression. Target cells from these test infections were analyzed by FACS-Gal to determine the frequency of β-gal$^+$ cells. In addition, the β-gal$^+$ cells in the infections were sorted and expanded in culture. The *lacZ* expression in this culture of sorted cells was then analyzed. Because the frequency of successful integrations with Gensrl (i.e., β-gal$^+$ cells) is very low, the reanalysis of the sorted and expanded population proves to be the most reliable assay of a producer clone's ability to transduce the Gensrl genome. Ecotropic (ψ2/A8) and amphotropic (PA/A36) producers of Gensrl were isolated and used in all subsequent Gensrl infections.

Infection of Target Cells with the Gene-Search Viruses

We have routinely infected target cells by co-cultivating target cells with gene-search virus producer cells. We have found that lymphoid cell lines neither adhere to the culture dishes nor do they become irreversibly adhered to the fibroblast-derived producer cells. To establish co-cultures of producer and target cells, virus producer cells are first plated onto 100 × 15-mm LUX dishes at a density

FIG. 5. Subcellular localization of β-galactosidase fusion proteins in cells harboring fusions between the AcLac neo-exon and a cellular gene. (A) Nuclear localized β-galactosidase. (B) Cytoplasmic β-galactosidase. (C) Perinuclear localized β-galactosidase. (D) Diffuse, whole-cell β-galactosidase. X-gal stains were performed as described previously (6).

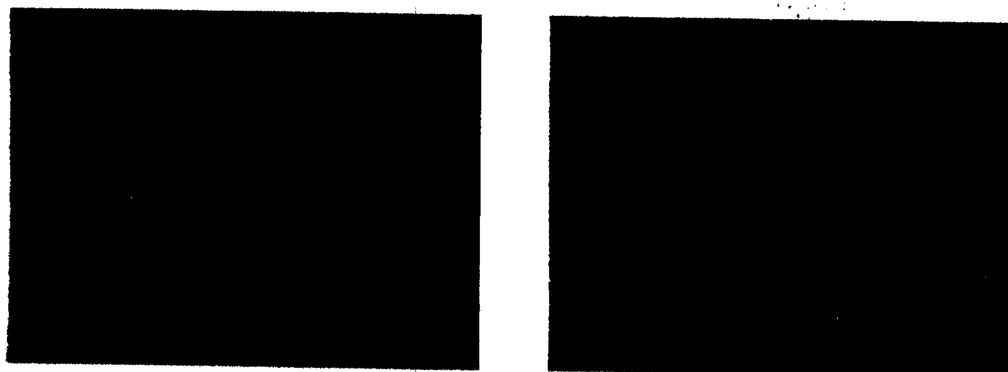

FIG. 10. Subcellular localization of β-galactosidase in Gensr1 integrants, 7a23 and 3a21. X-gal stains were performed as described previously (6).

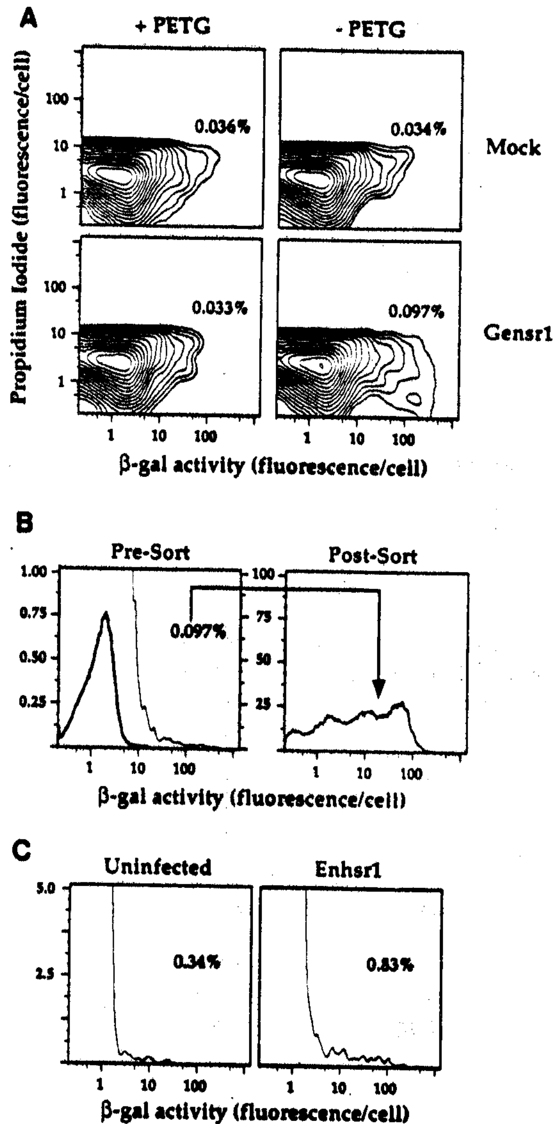

FIG. 6. Detection, estimation, and isolation of cells with *lacZ* gene fusions from Gensrl infections. (A) Detection and estimation of 70Z/3 cells harboring *lacZ* gene fusions following infection with Gensrl. Dual-color log 50% contour plots of PI vs fluorescein for mock-infected and Gensrl-infected 70Z/3 cells loaded and incubated with FDG in the absence and presence of the competitive inhibitor, PETG. The shaded areas of the contour plots indicate software gates used to estimate the percentage of fluorescein-positive cells in each sample (indicated above the shaded area). Mock infections were carried out by co-cultivating 70Z/3 cells with ψ2 cells plated at a density equal to that of the ψ2 producer of Gensrl, ψ2/A8. Both infections were done at 4 μg/ml polybrene. Cells were removed from the co-cultivation after 18 h as described in the text, replated in normal RPMI 1640 medium, and expanded for another 24 h. The cells were recovered from culture dishes by vigorous pipetting, pelleted, and resuspended in 100 μl of staining medium. The cells were then hypotonically loaded with FDG (1–3) and incubated for ~3 h on ice. For samples containing PETG, PETG was present at 2 mM when cell pellets were resuspended in 100 μl of staining medium. The cold, isotonic staining medium used to halt loading contained 1 mM PETG. Prior to FACS analysis, cells were pelleted and resuspended in 100 μl of staining medium containing 1 μg/ml propidium iodide and 1 mM PETG. FACS analysis was as described (1–3). (B) FACS-Gal selection and enrichment of *lacZ*-expressing cells following Gensrl infection. Presort density histograms (100%—bold line, 1%—thin line) of Gensrl-infected cells incubated in the absence of PETG (same data sample as that in A). Following sorting of the cells in the shaded area, they were expanded in culture for 1 week and reanalyzed by FACS-Gal (postsort 100% density histogram below arrow). Note that for the 100% density histograms (bold line) the relative cell number is indicated in the center of the two plot boxes. The relative cell number scale for the 1% density histogram (thin line) is on the left side of the presort plot box. (C) Estimation of 70Z/3 cells harboring *lacZ* gene fusions following Enhsrl infection. Density histograms (5%) of uninfected and Enhsrl-infected cells. Shaded areas represent software gates used to estimate the percentage of fluorescein-positive cells in each sample. This percentage is indicated in the shaded area.

of 1.5 × 10^4/cm^2 in a total volume of 10 ml of DMEM. The producer cells are then grown to near confluency (~48 h). To start co-cultivation, half of the producer cell medium is removed from the dish (5 ml) and 2 × 10^6 target cells in fresh medium (5 ml) containing 8 μg/ml polybrene are added to the plate of producer cells. In experiments where we wished to isolate integrations in which *lacZ* expression is induced by LPS, co-cultivation of target cells with producer cells was done in the presence of LPS (10 μg/ml).

Analysis and Isolation of lacZ⁺ Target Cells with Productive Integrations of a Gene-Search Retrovirus

Recovery of target cells from adherent producer cell monolayers is achieved by first washing the culture medium over the surface of the dish two to three times. Following these washes, remove the culture medium, which now contains most of the target cells, to a sterile centrifuge tube. Following the first wash, some target cells will remain adhered; these can be dislodged by hitting the bottom of the culture dish against a flat hard surface two or three times. Recover the residual target cells by rinsing the plate with fresh medium. Combine the two washes and pellet the target cells to concentrate them for loading of fluorescein di-β-galactopyranoside (FDG). For hypotonic loading, the cells are resuspended in a small volume (50–200 μl) of sterile staining medium (3) and warmed to 37°C. The cells are hypotonically loaded with FDG and brought back to isotonicity as described previously (1–3).

For analysis and sorting of *lacZ*$^+$ gene-search integrations, samples should be incubated on ice for at least 3 h. This ensures that even cells with low levels of *lacZ* expression will have sufficient time to cleave sufficient FDG to fluorescein, enabling the FACS to distinguish them from negative cells. Prior to FACS analysis all samples are brought to 1 μg/ml propidium iodide to allow exclusion of dead cells and to enhance the exclusion of rare false positives.

For one to accurately assess whether successful infection of target cells has occurred, it is best to hypotonically load FDG into an aliquot of the infected cells in the presence of 1 mM PETG. Because PETG is an inhibitor of β-

gal, it will prevent cells that are β-gal⁺ in the sample from generating fluorescein and being detected; however, the autofluorescent false positives will still be detected. Thus, comparison of the infected target cells plus or minus PETG permits an accurate estimate of the β-gal⁺ cells in the infection as well as allows the selection of sorting gates that avoid as many of the false positive cells as possible, although these cells cannot be completely excluded. This control is valid only if PETG is present prior to loading, during loading, and at all times following loading until the cells are analyzed.

To demonstrate that β-gal⁺ cells can be detected in Gensr1-infected target cells, we show FACS-Gal analysis of mock- and Gensr1-infected 70Z/3 cells following incubation in the presence and absence of PETG (Fig. 6A). Although the frequency of β-gal⁺ cells following infection with Gensr1 is low (0.05–0.06%), we can readily and reproducibly distinguish true positives in Gensr1 infections by the +/− PETG comparison. These analyses are always performed with hardware gating to exclude cells that are above the normal distribution of yellow fluorescence (propidium iodide channel) for the target cells. In addition, events that are below the normal distribution of forward light scatter for the target cells should also be excluded, since they are primarily dead cells or cell debris. Analysis of the infected cells +/− PETG with the above hardware gates allows one to determine whether productive integrations have occurred. For comparison we show that the Enhsr1 virus is able to transduce *lacZ* expression into approximately 0.5% of 70Z/3 cells after co-cultivation (Fig. 6C). Enhsr1's transduction efficiency is approximately an order of magnitude higher than that for Gensr1, presumably because their target sizes differ significantly.

Selection of β-gal⁺ cells from the gene-search infection is then brought about by sorting cells as a population or by sorting single cells into the individual wells of a 96-well plate containing growth medium. We use three hardware gates for this purpose: (i) propidium iodide gate—exclude all cells above the normal distribution of yellow fluorescence for the target cells; (ii) forward light scatter—exclude all cells smaller than the normal distribution of light scatter for the target cell population; (iii) fluorescein gate—select all cells that are at least one log more fluorescent than the mean fluorescence of the target cells. If the frequency of fluorescein-positive cells is significantly greater in the gene-search infected cells not stopped with PETG than in those stopped with PETG, one should see a significant enrichment for *lacZ*-expressing cells upon FACS-Gal analysis of the sorted cells after growth in culture. For Gensr1 infections, we typically find that the sorted and expanded cells are 60–70% β-gal⁺ upon analysis by FACS-Gal (see Fig. 6B).

Isolation of Regulated Gene-Search Integrations

In our hands the primary role of the gene-search viruses has been to tag differentially regulated cellular genes or loci with a transcriptionally incompetent reporter gene, thus permitting the isolation and characterization of genes responsive to a given stimulus. Because the overwhelming majority of *lacZ*⁺ gene-search integrations will be constitutive with respect to a given stimulus, one must be able to identify cells with the rare integration in a differentially regulated locus or gene.

We have been successful in finding integrations of the gene-search viruses in which *lacZ* expression is responsive to LPS-stimulated differentiation of a pre-B lymphoma, 70Z/3. When stimulated with LPS, 70Z/3 cells differentiate from the pre-B to the B-cell stage (22). The cells return to the pre-B cell stage when LPS is removed. With Enhsr1 we have found such LPS-regulated integrations at a frequency of ~1/50 (8/411 clones). With Gensr1 the frequency appears to be lower; 2 clones with significant responses to LPS stimulation have been found among approximately 400 total clones. All regulated integrations were identified by cloning individual β-gal⁺ cells into 96-well plates and expanding the clones until there was a sufficient number of cells to remove an aliquot for characterization by either FACS-Gal or the 4-methylumbelliferyl-β-D-galactoside (MUG) assay (2, 3).

Because regulated integrations are rare, a large number of clones must be screened to provide a reasonable chance of identifying a clone with such an integration. This is best achieved by the 96-well-based fluorogenic assay for β-gal, the MUG assay (2, 3). This procedure permits the response of *lacZ* expression in several hundred clones with or without stimulus to be determined within 1–2 days of work by a small modification of the previously published procedure for the MUG assay (2, 3). To screen a large number of clones +/− stimulus, remove two equal aliquots (50 μl each) of cells from the well of the original 96-well plate into which *lacZ*⁺ cells were sorted and expanded. One aliquot is added to 50 μl of normal growth medium in a 96-well plate while the other aliquot is placed in a parallel well of another 96-well dish containing 50 μl of 2× stimulus medium. These 96-well dishes then represent replicate plates plus or minus the stimulus. The cells are then cultured for an amount of time sufficient for the cells to respond to the stimulus. The relative β-gal activity of each clone with respect to the stimulus is then determined by adding 50 μl of 1.8 mM MUG/0.3% Triton X-100 in Z-buffer to the 100-μl 96-well culture. Addition of this mixture to the culture with sufficient pipetting produces a cell lysate of the clones in the presence of a fluorogenic substrate for β-gal, MUG. After the 96-well plates are incubated at 37°C for 1–2 h, relative β-gal activities are quantified directly in the 96-well dish by a Titertek Fluoroskan II (Flow Labs, MacLean, VA) 96-well plate fluorometer.

Comparison of the lacZ Expression in Gene-Search Integrations with Other Cellular Parameters

The isolation and study of differential gene expression via the gene-search viruses offer the ability to study

expression of a gene with respect to a stimulus while simultaneously measuring other cellular parameters on a single-cell level. *lacZ* expression in gene-search clones can be compared with other cellular parameters by multiparameter FACS analysis. Cellular parameters with which *lacZ* expression can be compared include: (i) position in the cell cycle (Hoechst dye vs fluorescein) (23); (ii) expression of cell surface proteins (fluorochrome-conjugated antibodies vs fluorescein) (6, 10); and (iii) cell size (light scatter vs fluorescein). Multiparameter FACS-Gal analysis of gene-search clones offers a distinct advantage over cDNA library procedures designed to enrich for differentially expressed mRNAs because $lacZ^+$ integrations with very precise phenotypes can be screened for or selected. Thus, $lacZ^+$ integrations that have the most desirable phenotype can be chosen for further characterization by molecular cloning procedures.

Multiparameter analysis of LPS-responsive gene-search integrations illustrates this point (Fig. 7). In our case, we are most interested in characterizing gene-search integrations in which *lacZ* expression responds to the differentiation of pre-B cells rather than to the mitogenic effects of LPS. When pre-B cells differentiate to the B-cell stage, they express a high density of IgMκ on their cell surface, whereas at the pre-B cell stage they lack surface IgMκ but can express a low density of μ chains in association with a light-chain-like complex of proteins (24, 25). These cell surface parameters can be simultaneously compared to *lacZ* expression via FACS-Gal combined with cell surface staining (6). In Fig. 4 we demonstrate that there is heterogeneity among the LPS-responsive integrations with respect to LPS-stimulated differentiation. In the clones 7e17-17 and 7a291-1, *lacZ* expression is found at the pre-B cell stage and thus is repressed in cells when they differentiate to the B-cell stage as determined by Igκ expression. In the clones 7e129-3 and 7a309-1, *lacZ* expression is induced by LPS and there is a clear bias for β-gal activity found in the Igκ^+ cells (B-cell stage). Expression of *lacZ* is also induced by LPS in 7e131-3; however, cells with β-gal activity show no clear bias for either Igκ^+ or Igκ^- cells. We hypothesize that *lacZ* induction in 7e131-3 cells may be due to mitogenic effects of LPS on B-lineage cells and that this integration is not responsive to differentiation-specific signals transmitted by LPS. Thus, multiparameter FACS-Gal analysis permits the expression of a gene or locus to be compared with a wide variety of cellular parameters prior to actual initiation of molecular cloning procedures. This capability allows integrations with very specific expression patterns to be identified and represents one advantage of this technology over subtractive hybridization in the study of differentially regulated genes.

The Potential of Gene-Search Viruses for Studying Gene Expression in Vivo

In addition to following developmentally regulated loci in transformed cell lines, the gene-search retroviruses may allow gene expression to be followed *in vivo*. We have focused our attention on cells of the hematopoietic system and thus we first wanted to determine whether *lacZ* expression could be obtained in primitive hematopoietic cells. To determine whether this was feasible, cells from the fetal liver of a Day 14 embryo were mock-infected or cultured with Gensrl producer cells or with Gensrl producer cell supernatant. Following infection of the cells for 24 h, the frequency of cells expressing *lacZ* was esti-

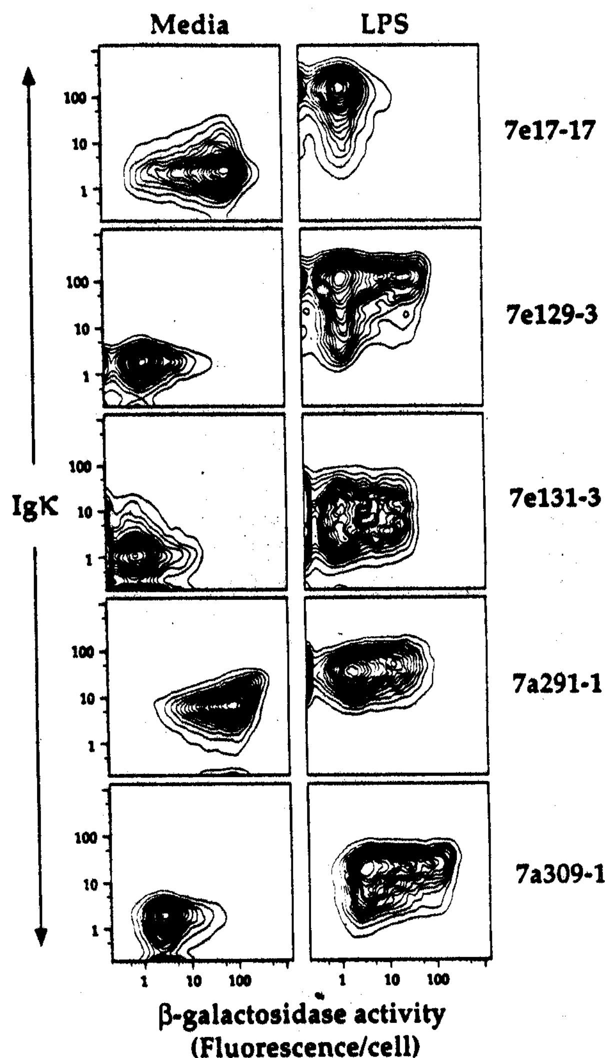

FIG. 7. Regulation of *lacZ* and κ-light-chain expression in gene-search integrants following LPS-stimulated differentiation. Dual-parameter FACS analysis of κ-light-chain expression (Texas Red) vs β-galactosidase activity (fluorescein) in normal medium or in LPS-containing medium. With the exception of 7a291-1 cells, equal numbers of individual gene-search clones (Enhsrl clones 7e17-17, 129-3, 131-3; Gensrl clones 7a291-1, 309-1) were cultured either in normal medium or in medium containing 10 μg/ml of LPS for 24 h. For 7a291-1, cells were cultured at equal density in either normal medium or medium containing 10 μg/ml LPS and 100 U/ml IFN-γ. The cells were then stained and analyzed on the FACS as described previously (6, 10).

mated by FACS-Gal analysis (Fig. 8). FACS-Gal analysis demonstrates that Gensr1 can transduce *lacZ* expression into normal hematopoietic cells *in vitro* at a low but detectable frequency (~0.1%).

To test the feasibility of using gene-search viruses to follow gene expression *in vivo*, we undertook reconstitution experiments with gene-search virus-infected fetal liver cells. These reconstitution experiments have two separate but distinct goals that depend upon the expression pattern of *lacZ*. One goal is to obtain gene-search integrations in a constitutively expressed locus (e.g., cellular metabolism gene) of a hematopoietic stem cell. Such an integration provides a marker for the reconstitution potential of an individual stem cell clone through multiparameter FACS-Gal analysis of reconstituted hosts. The second goal is to obtain hematopoietic stem cells with integrations in developmentally regulated loci. The expression pattern of such gene-search integrations in the hematopoietic system can be determined by a combination of Southern blot analysis (to determine which cells of the hematopoietic system harbor the integration) and FACS-Gal analysis (to determine which cells in the hematopoietic system express *lacZ*). If the expression pattern appears novel or interesting, molecular cloning of the cellular gene is possible because *lacZ* sequences are present at the cellular locus.

In reconstitution experiments we have used either total gene-search infected fetal liver cells or β-gal$^+$ gene-search infected fetal liver cells as the source of donor cells. No β-gal$^+$ cells have been detected in spleen (Spl), bone marrow (BM), or peritoneal cavity (PerC) cells in 8 mice reconstituted with total gene-search infected fetal liver cells. This result is not surprising given the low frequency of β-gal$^+$ cells present following infection with Enhsr1 or Gensr1 (Figs. 6 and 8). In experiments in which sorted β-gal$^+$ cells were used as the sole reconstitution source, only 1 mouse of a total of 11 survived lethal irradiation. This mouse had been reconstituted with ~5000 β-gal$^+$ cells sorted from an Enhsr1 infection of Day 14 fetal liver cells from an IgH-6a donor. Expression of *lacZ* was detected in 0.6% of the PerC cells of the IgH-6b host (Fig. 9) while Spl and BM were negative for *lacZ* expression (data not shown). Multiparameter FACS-Gal analysis indicates that all β-gal$^+$ cells express the Ig allotype of the donor and Mac1 (Fig. 9) and are very large based on light scatter measurements (data not shown). The β-gal$^+$ cells are either of the Ly-1 B lineage (26) or of the myeloid lineage due to their light scatter properties and Mac1 staining. The origin of the β-gal$^+$ cells depends on whether the cells synthesize donor IgM (IgH-6a) or have acquired it via binding of secreted IgM to Fc receptors.

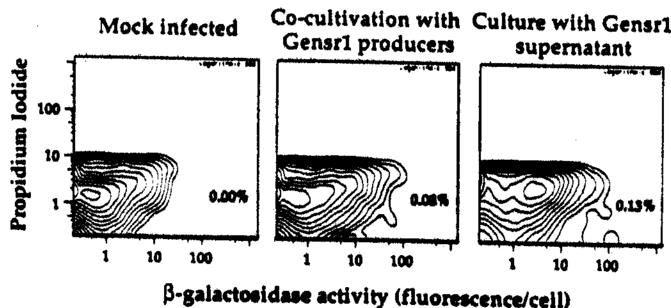

FIG. 8. Transduction of *lacZ* expression into hematopoietic progenitor cells via Gensr1. Day 14 fetal liver cells were plated at 5×10^6 cells/ml either on an irradiated monolayer of Gensr1 producers (ψ2/A8) or in Gensr1 supernatant with 4 μg/ml polybrene for 24 h. The mock-infected cells were cultured at an equivalent cell density in polybrene for an equal amount of time. The cells were then analyzed by FACS-Gal. The propidium iodide-positive cells (dead cells) were excluded from the data collection. Data on ~200,000 cells were collected. The percentage of β-gal$^+$ cells is indicated within the gray rectangles of each contour plot (contours are drawn at consecutive twofold levels).

The above experiments demonstrate that gene-search viruses can be used to transduce expression of *lacZ* into normal hematopoietic cells *in vitro* and *in vivo*. However, the efficiency of this approach is quite low and may require improvements such as the introduction of a drug-selectable marker into the gene-search viruses. Their low efficiency may also be due to difficulties in expressing *lacZ* in all hematopoietic lineages (3). We are currently developing an alternative FACS-based reporter gene that may not experience the same difficulties with expression *in vivo*.

Subcellular Localization of β-gal Fusion Proteins with Gensr1

Because *lacZ*$^+$ integrations of Gensr1 represent translational fusions between β-gal and cellular proteins, we observe that certain Gensr1 clones exhibit subcellular localization of β-gal activity as detected by indolyl galactoside (X-gal) histochemistry. This was observed originally in clones transfected with the AcLac construct (6). In this assay the 4-Cl-5-Br-3-indolyl moiety is cleaved from the galactoside residue by β-gal in the presence of an appropriate histochemical solution causing it to precipitate immediately, thus demarcating the location of β-gal within the cell. In Fig. 10 we show phase contrast micrographs of two Gensr1 clones where β-gal is localized to specific subcellular locations. In 7a23, β-gal is nuclear-localized, while in 3a21 it has a perinuclear localization. Because an X-gal precipitate from a relatively small number of β-gal molecules is difficult to detect by phase contrast microscopy (requires several hundred molecules of β-gal/cell) (3), other more sensitive means could be used to detect precise subcellular localizations of proteins expressed at low levels in the cell. Because the X-gal precipitate contains electron-dense Fe atoms, electron microscopy is capable of detecting these precipitates (27). Therefore, electron microscopy of X-gal-stained Gensr1 clones could increase our capability of detecting mammalian proteins with precise subcellular localizations and

possibly reveal cellular proteins fused to β-gal that have novel localization patterns.

DISCUSSION

Gene-search viruses that allow the detection of cellular genes (Gensrl) or their transcription control elements (Enhsrl) have been developed by the creation of *in situ* gene fusions with *E. coli lacZ*. Because retroviral infection is the most efficient procedure for stable introduction of a reporter gene into mammalian cells, these viruses permit a large pool of *lacZ*-expressing integrations to be isolated via FACS-Gal. Because *lacZ* expression is easily monitored via the MUG assay (2, 3) integrations with differential regulation patterns can readily be detected (6, 10). Differential sorting based on the *lacZ* phenotype of a cell with respect to other cellular parameters can also lead to the identification of cells with integrations in differentially regulated genes (14). Because of Gensrl's requirement that *lacZ* expression result from translational fusion with cellular proteins, this virus also permits the identification of mammalian proteins with specific subcellular localizations (6) and may also permit the identification of proteins with novel subcellular localizations.

Infection of cells with the gene-search viruses results in *lacZ* being dispersed throughout the genome within a transcriptionally defective provirus. Expression of *lacZ* is then dependent upon the transcriptional state of the surrounding chromatin. Thus, if a cellular gene is conditionally expressed, *lacZ* expression will also be conditional. Because expression of *lacZ* is easily monitored, integrations near differentially regulated genes can be identified by screening the response of *lacZ* in individual clones or via differential cell sorting based on *lacZ* expression (6, 14). We have succeeded in detecting and isolating cells with integrations in differentially regulated loci with both the Enhsrl and the Gensrl viruses. A total of 8 Enhsrl integrations have been identified out of approximately 400 clones (~1/50) that have significant responses to LPS-stimulated differentiation of a pre-B lymphoma (6, 10). Two Gensrl integrations out of approximately 400 (~1/200) that show a significant response to LPS stimulation have been identified.

Although Enhsrl integrations in developmentally regulated domains of chromatin could lead to the identification of the cellular gene(s) contained within such a domain, this task could prove cumbersome from a technical perspective. Because β-gal activity in Gensrl-infected cells results from RNA splicing of a cellular exon(s) to the AcLac neo-exon, cDNA cloning procedures should permit identification of the involved cellular gene. The recent development of combined cDNA/PCR procedures (28, 29) may greatly facilitate isolation and identification of the cellular sequences fused to AcLac. In addition, we have found that the intron length in gene fusions with AcLac can be estimated by Northern blot analysis of total poly(A)$^+$ RNA (W. Kerr, unpublished). This estimation permits cloning of genomic DNA fragments containing Gensrl with reasonable confidence that genomic clones will contain exonic sequences. The cellular portion of the genomic clone containing Gensrl can then be used to probe standard cDNA libraries to obtain full-length cDNA clones of the involved cellular gene.

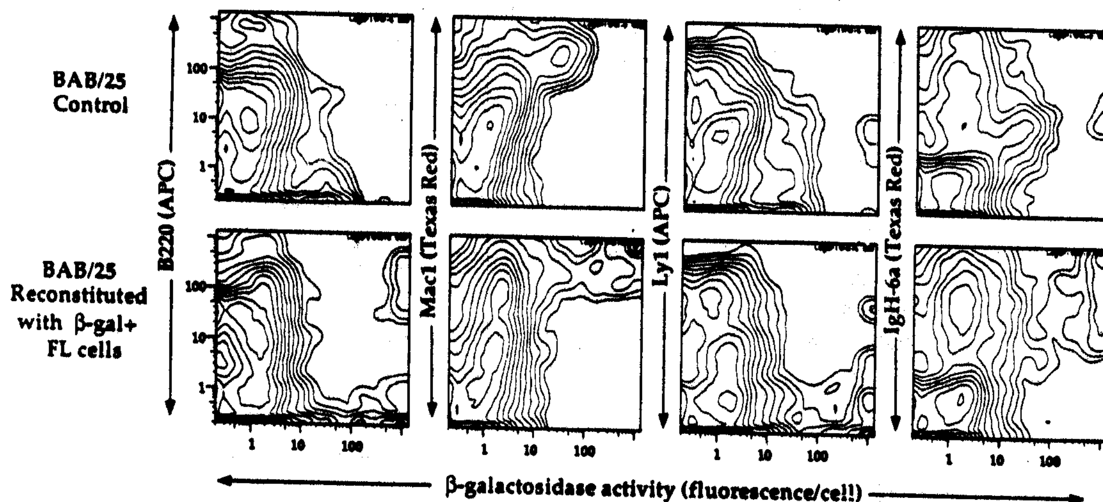

FIG. 9. *In vivo* expression of *lacZ* following reconstitution with β-gal$^+$ fetal liver cells obtained by infection with the Enhsrl retrovirus. Multiparameter FACS analysis of hematopoietic lineage markers (*y* axis) versus β-gal activity (*x* axis) of PerC cells from a BAB/25 host (IgH-6b) reconstituted with 5000 β-gal$^+$ BALB/C (IgH6a) fetal liver cells infected with Enhsrl. The percentage of fluorescent cells in an unreconstituted BAB/25 PerC was estimated at ~0.05%, while in the BAB/25 host reconstituted with Enhsrl-infected *lacZ*$^+$ fetal liver cells there were 0.63% β-gal$^+$ cells. The stains were B220 (APC) vs Mac1 (TexRed) vs β-gal (fluorescein) and Ly-1 (APC) vs IgH6a (TexRed) vs β-gal (fluorescein).

One advantage of the Gensrl virus with respect to Enhsrl and other search vectors (7, 13, 14) is that the *lacZ* gene fusion is more likely to be the result of fusion with a cellular gene due to the stringent criteria required for generation of β-gal activity from AcLac. The AcLac neo-exon lacks a translational initiator codon and contains several translation stop codons 5′ to the splice sites that are in frame with *lacZ* (see Fig. 3). Thus, it is improbable that β-gal activity could be generated without splicing of AcLac to an upstream coding exon such that β-gal is co-translated with the upstream cellular sequence. Despite these stringent requirements for expression, we can readily detect *lacZ*-expressing cells following Gensrl infection at a frequency of 0.05–0.07%. We are also able to significantly enrich (~1000-fold) for these cells after a single round of sorting. With other search viruses enrichments of ~250-fold (7) and ~1000-fold (14) have been found after two rounds of sorting. This may imply that expression of *lacZ* with these approaches is not stable in the majority of integrations, although stably expressing clones can be obtained following two rounds of sorting (7, 14). A possible explanation for this disparity among Gensrl and the other viruses (7, 14) is that expression of *lacZ* is stable for most Gensrl integrations because AcLac requires a fusion with a cellular gene at both the transcriptional and the translational level. The *lacZ* genes in other vectors (7, 14) contain a translational initiator codon and thus expression of *lacZ* may require only transcriptional readthrough from outside the provirus. Such readthrough transcription may occur transiently in a significant fraction of integrations, resulting in unstable expression and thus poor enrichment by cell sorting for *lacZ* expression.

The use of Gensrl to identify differentially expressed genes is not limited to those genes whose differential expression is controlled at the transcriptional level. Because *lacZ* expression from a Gensrl integration requires participation in RNA splicing, it is feasible that *lacZ* expression could be activated or repressed by alternative RNA splicing events operating on the cellular gene. In addition, because a fusion protein is formed with β-gal, Gensrl also has the potential to detect genes whose expression is regulated at the translational or post-translational level. Translational control of β-gal activity could occur if mammalian sequences in the fusion mRNA or in the NH_2 terminal portion of the fusion protein altered the translational capacity during certain cell states. Post-translational control of β-gal activity could occur if the mammalian NH_2 terminal portion of the fusion protein contained protein determinant(s) that specified preferential degradation of the protein during certain cell states. The ability of Gensrl to detect differentially regulated genes whose expression is controlled beyond the level of transcription, even at the post-translational level, is a distinct advantage of Gensrl over subtractive hybridization techniques. Although Gensrl is not meant as a replacement for subtractive hybridization cloning, it will certainly complement this technology in the elucidation of genes whose expression is necessary for mammalian development, differentiation, and cellular responses to stimuli.

ACKNOWLEDGMENTS

We acknowledge the considerable contributions of Dr. Garry Nolan to the development of the gene-search viruses. We also thank the members of the Herzenberg laboratory for critical comments and suggestions throughout the development and utilization of the gene-search viruses. This work was supported in part by National Institutes of Health Grant CA42509 to L.A.H. and NIH Training Grant AI 07290 (W.G.K.). W.G.K. is a Fellow of the Irvington Institute for Medical Research.

REFERENCES

1. Nolan, G. P., Fiering, S., Nicolas, J.-F., and Herzenberg, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **85**, 2603–2607.
2. Fiering, S. N., Roederer, M., Nolan, G. P., Micklem, D. R., Parks, D. R., and Herzenberg, L. A. (1991) *Cytometry* **12**, 291–301.
3. Roederer, M., Fiering, S. N., and Herzenberg, L. A. (1991) *Methods, Companion Methods Enzymol.* **1**, 248–260.
4. O'Kane, C. J., and Gehring, W. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9123.
5. Barklis, E., Mulligan, R. C., and Jaenisch, R. (1986) *Cell* **47**, 391–399.
6. Kerr, W. G., Nolan, G. P., Serafini, A. T., and Herzenberg, L. A. (1989) *Cold Spring Harbor Symp. Quant. Biol.* **54**, 767–776.
7. Brenner, D. G., Lin-Chao, S., and Cohen, S. N. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5517–5521.
8. Konzc, C., Martini, N., Mayerhofer, R., ZsuZsanna, K.-K., Korber, H., Redei, G. P., and Schell, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8467–8471.
9. von Melcher, H., Reddy, S., and Ruley, H. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3733–3737.
10. Kerr, W. G., Nolan, G. P., Johnsen, J. B., and Herzenberg, L. A. (1991) Mechanisms of Lymphocyte Activation and Immune Regulation. III. Developmental Biology of Lymphocytes, in press.
11. Casadaban, M. J., Chou, J., and Cohen, S. N. (1990) *J. Bacteriol.* **143**, 971–980.
12. Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S., Reik, W., and Surami, M. A. (1988) *Nature* **333**, 852–855.
13. Gossler, A., Joyner, A. L., Rossant, J., and Skarnes, W. C. (1989) *Science* **244**, 463–465.
14. Reddy, S., DeGregori, J. V., von Melchner, H., and Ruley, E. (1991) *J. Virol.* **65**, 1507–1515.
15. Yu, S. F., von Ruden, T., Kanoff, P. W., Garber, C., Seiberg, M., Ruther, V., Anderson, W. F., Wagner, E. F., and Gilboa, E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3194–3198.
16. Varmus, H. *Science* **216**, 812–820.
17. Mann, R., and Baltimore, D. (1985) *J. Virol.* **54**, 401–407.
18. Lazo, P. A., Prasad, V., and Tsichlis, P. N. (1987) *J. Virol.* **61**, 2038–2041.
19. Mann, R., Mulligan, R. C., and Baltimore, D. (1983) *Cell* **33**, 153–159.

20. Miller, A. D., and Buttimore, C. (1986) *Mol. Cell. Biol.* **6**, 2895–2902.
21. Southern, P., and Berg, P. (1982) *J. Mol. Appl. Genet.* **1**, 327–341.
22. Paige, C. J., Kincade, P. W., and Ralph, P. (1978) *J. Immunol.* **121**, 641–647.
23. Fiering, S. N., Northrop, J. P., Nolan, G. P., Mattila, P. S., Crabtree, G. R., and Herzenberg, L. A. (1990) *Genes Dev.* **4**, 1823–1834.
24. Pillai, S., and Baltimore, D. (1987) *Nature* **329**, 172–174.
25. Kerr, W. G., Cooper, M. D., Feng, L., Burrows, P. D., and Hendershot, L. M. (1989) *Int. Immunol.* **1**, 354–361.
26. Hayakawa, K., Hardy, R. R., Herzenberg, L. A., and Herzenberg, L. A. (1985) *J. Exp. Med.* **161**, 1554–1568.
27. Bunge, M. B., Wood, P. M., Tynan, L. B., Bates, M. L., and Sanes, J. R. (1989) *Science* **243**, 229–231.
28. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
29. Loh, E., Elliott, J. F., Cwirla, S., Lanier, L. L., and Davis, M. M. (1989) *Science* **243**, 217–219.
30. Srivastava, A., Lusby, E. W., and Berns, K. I. (1983) *J. Virol.* **45**, 555–564.
31. Laughlin, C. A., Tratschin, J.-D., Coon, H., and Carter, B. J. (1983) *Gene* **23**, 651–673.