

Transcriptionally Defective Retroviruses Containing *lacZ* for the In Situ Detection of Endogenous Genes and Developmentally Regulated Chromatin

W.G. KERR, G.P. NOLAN, A.T. SERAFINI, AND L.A. HERZENBERG
 Department of Genetics, Stanford University, Stanford, California 94305

Previously, we demonstrated that expression of transduced *Escherichia coli lacZ* can be detected in individual mammalian cells with a fluorescence-activated cell sorter (Nolan et al. 1988), a technique referred to as FACS-GAL. This fluorogenic assay permits the quantitative measurement of *E. coli* β -galactosidase (β -Gal) in individual viable cells. This technique also permits rare cells expressing β -Gal to be isolated utilizing the sorting capacity of the FACS. We have taken advantage of this feature to isolate β -Gal-expressing cells where the *lacZ* gene is under the control of endogenous transcription control elements. To permit the isolation of such cells, we have developed systems for delivering a reporter gene, *E. coli lacZ*, into the genome of mammalian cells in such a way that its expression is dependent on endogenous transcription control elements. We have stably introduced *lacZ* into the genome of mammalian cell lines by infection with *lacZ* gene search retroviruses, as well as by transfection with a splice acceptor/*lacZ* construct, AcLac. Each *lacZ* gene search virus we have developed requires that the flanking cellular sequences complement some transcriptional deficiency of the provirus for β -Gal to be expressed by the infected cell. The *lacZ* gene search viruses consist of an enhancer element search virus, Enhsr1, a promoter search virus, Prosr1, and a gene search virus containing a splice acceptor/*lacZ* fusion, Gensr1.

By studying the expression of *lacZ* derived by infection with either Enhsr1 or Prosr1, we demonstrate that individual β -Gal-expressing clones have very different distributions of activity, both qualitatively and quantitatively. We demonstrate that these broad ranges of *lacZ* expression seen in many individual clones represent a controlled variation in expression that is probably cyclical. Since expression in these clones is dependent on endogenous transcription elements, this may also be a common feature of cellular genes. As further proof of the endogenous control, we have been able to identify Enhsr1 clones where *lacZ* expression is regulated in a differentiation-stage-dependent manner. These Enhsr1 clones were derived from infections of a differentiation-inducible pre-B-cell line, 70Z/3, demonstrating the presence of multiple stage-specific enhancers involved in the differentiation of B-lineage cells.

Introduction of a splice acceptor/*lacZ* construct either by transfection of the AcLac construct or by infection with the AcLac virus into mammalian cells results in transcriptional and translational fusion of *lacZ* to endogenous genes and their protein products and thus provides the most pristine in situ measurement of endogenous transcriptional control we have developed. As with Enhsr1 and Prosr1, we find that distributions of β -Gal activity differ both qualitatively and quantitatively between individual clones. Since these proteins retain the enzymatic activity of bacterial β -Gal while also exhibiting the mammalian characteristic of subcellular localization, we propose that this represents a model for molecular evolution of multidomain, multifunctional proteins via exon shuffling. Moreover, the high frequency of success in generating in vivo gene fusions by introduction of the AcLac neo-exon into the genome strongly supports exon shuffling as a mechanism for generating novel protein structures in molecular evolution. Finally, we have used a retrovirus containing the splice acceptor/*lacZ* (Gensr1) to obtain gene fusions that are differentially regulated in B-lineage cell lines.

METHODS

FACS-GAL assay and immunofluorescence. FACS-GAL assay has been described previously (Nolan et al. 1988; Nolan 1989). For two-color analysis by FACS-GAL and immunofluorescence, 70Z/3 cells were first hypotonically loaded with fluorescein-di-galactoside (FDG) (from Molecular Probes, Junction City, Oregon) as described previously, brought back to isotonicity at 4°C, and pelleted. The pellet was resuspended in 100 μ l of cold staining medium containing 10 μ g/ml biotinylated monoclonal anti-Ig κ . After incubation on ice for 15 minutes, the cells were washed three times with 2 ml of ice-cold staining medium and pelleted at 4°C. After the third pelleting, cells were resuspended in 100 μ l of Texas Red (from Molecular Probes) complexed avidin (Vector Laboratories, Burlingame, California) and incubated on ice for 15 minutes. Cells were washed three times as before and resuspended in 100 μ l of staining medium containing 1 μ g/ml propidium iodide (to mark dead cells). In some experiments, where indicated, phenylethylthiogalac-

tosidase (PETG), a strong, hydrophobic competitive inhibitor of β -Gal, was added to a concentration of 1 mM to stop hydrolysis of FDG to fluorescein (Nolan 1989). Single-color FACS analyses are presented as graphs (density 100%). Dual-color FACS analyses are presented as dual-color contour plots, 5% probability.

Infection of cells by *Enhsl1*, *Prosl1*, and *Gensl1* viruses. 70Z/3 cells were infected by cocultivation for 16–18 hours with the Ψ 2 lines producing *Enhsl1* (Ψ 2/E2) or *Prosl1* (Ψ 2/P10) in the presence of 2 μ g/ml polybrene. The *Enhsl1* or *Prosl1* producer cell lines were irradiated with 3000 rads of γ -radiation prior to cocultivation with 70Z/3 to prevent further cell division by the producer cell line. *Gensl1* infections were done by cocultivation of 70Z/3 cells with the ecotropic producer line Ψ 2/A36 or, in the case of NFS-5.3, with the amphotropic producer, PA/A8. Nearly confluent monolayers of *Gensl1* producers were UV-irradiated for 30–40 seconds in a sterile tissue culture cabinet prior to cocultivation with 70Z/3 or NFS-5.3 cells in the presence of 2 μ g/ml polybrene. All LPS treatments were done at a concentration of 10 μ g/ml (Difco, *S. typhosa* 0901 Westphal). IL-4 inductions were done at 20 units/ml (a kind gift from A. O'Garra and M. Howard, DNAX, Palo Alto, California).

Indolyl galactoside (X-Gal) histochemistry. Cells were fixed in 4% paraformaldehyde in phosphate-buffered saline. Cells were stained with the histochemical dye 5-bromo-4-chloro-indolyl-galactoside (Sigma) for a period of either 3 hours or overnight at 37°C, the length of staining prior to photography depending on the activity of the individual 293/AcLac clone. Cells were photographed using a Zeiss D7082 microscope (courtesy of H. Blau, Stanford University).

Immunoprecipitation and SDS-PAGE analysis of 293/AcLac β -Gal fusion proteins. Approximately 10^7 cells of each 293 AcLac clone were preincubated in 1 ml of methionine-deficient RPMI 1640 with 10% fetal calf serum for 1 hour. The cells were then pulsed with 400 μ Ci [35 S]methionine and 80 μ Ci [35 S]cysteine for 4 hours. The cells were then pelleted and lysed in 1 ml of 1% Nonidet P-40, 0.15 M NaCl, 50 mM Tris (pH 7.5), and 1 mM phenylmethylsulfonyl fluoride. The cell lysates were preadsorbed by the addition of 10 μ g of anti-HLA and Pansorbin (Calbiochem, San Diego, California). For the final precipitation, 10 μ g of monoclonal anti- β -Gal (Promega, Madison, Wisconsin) was added to the lysates with precipitation being achieved by addition of Pansorbin. Immunoprecipitated proteins were resolved after reduction by electrophoresis on a 7.5% SDS-polyacrylamide gel and visualized by autoradiography.

AcLac construct and transfection into 293 cells. The 125-bp splice acceptor of the Moloney leukemia virus *env* gene (derived from a *Bam*HI, *Kpn*I digest of pZip-NeoSV[X]) (Cepko et al. 1984) was fused to the 5' terminus of the *lacZ* gene by direct replacement of the

cytomegalovirus promoter and *lacZ* initiator codon (removed by *Kpn*I, *Bam*HI digestion) of pON405 (kindly provided by E. Mocarski, Stanford University). The resulting plasmid is called pAcLac. An approximately 4.2-kb fragment containing the splice acceptor/*lacZ* followed by the SV40 poly(A) site was isolated from a *Bam*HI, *Hind*III digest of pAcLac. This fragment was cotransfected with SV2neo (Southern and Berg 1982) at a fivefold molar excess into 293 cells by a CaPO₄ precipitation procedure (Graham et al. 1980).

Construction of transcriptionally defective retroviruses containing *lacZ*. The *lacZ*-containing plasmid, pON405 (a kind gift from B. Manning and E. Mocarski, Stanford University), was digested with *Dra*I, and a 3.5-kb fragment containing *lacZ* was isolated. *Bam*HI linkers were ligated onto this 3.5-kb *Dra*I fragment, and the ligation was digested with *Bam*HI and *Sal*I. The digest was separated on a 1% low-melting agarose gel, and the 3.2-kb *Bam*HI-*Sal*I fragment was isolated. This *lacZ*-containing *Bam*HI-*Sal*I fragment was ligated to a purified 6.5-kb *Xho*I-*Bam*HI fragment generated by digestion of either plasmid, pJrEnh- or pJrPro- (a kind gift from H. Stuehlmann, Stanford University and Brad Guild, MIT). After transformation into JM109 on X-Gal-containing LB plates, blue colonies were selected for further analysis to confirm the correct orientation of *lacZ*. Construction of the AcLac-containing retrovirus and isolation of *Gensl1* ecotropic and amphotropic producer lines will be described in a subsequent manuscript.

RESULTS

Expression of Transduced *lacZ* Is Controlled by Endogenous Transcription Control Elements

We have demonstrated previously that expression of transduced *lacZ* can be detected in viable mammalian cells with a fluorescence-activated cell sorter (Nolan et al. 1988; Nolan 1989), a technique referred to as FACS-GAL. The retrovirus used in this earlier study had *lacZ* under the control of an internal promoter, the SV40 late promoter; however, expression varied among independent clones containing different integration sites, suggesting that endogenous transcription control elements were influencing the expression of *lacZ* (Nolan 1989). To more accurately assess the control of introduced reporter genes by endogenous transcription control elements, we have constructed self-inactivating (SIN) retroviruses containing *lacZ*. These retroviruses will deliver *lacZ* into the chromatin of a target cell after the removal of the Moloney 72-bp enhancer element (enhancer-search, *Enhsl1*) or after removal of the entire Moloney transcription control region (promoter-search, *Prosl1*) (illustrated for *Enhsl1* in Fig. 1). This can be accomplished because the U3 of the 3' long terminal repeat (LTR) serves as the template for reverse transcription of the U3 region in both 3' and 5' LTRs (Varmus et al. 1982). Thus, if a

Figure 1. Schematic depiction of *Enhsr1* and model for *lacZ* expression. This figure depicts how a *lacZ*-encoding provirus that lacks the Moloney leukemia virus enhancer region (stippled box) but will 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 *Enhsr1*, viral integrations near endogenous enhancer elements will result in *lacZ* expression.

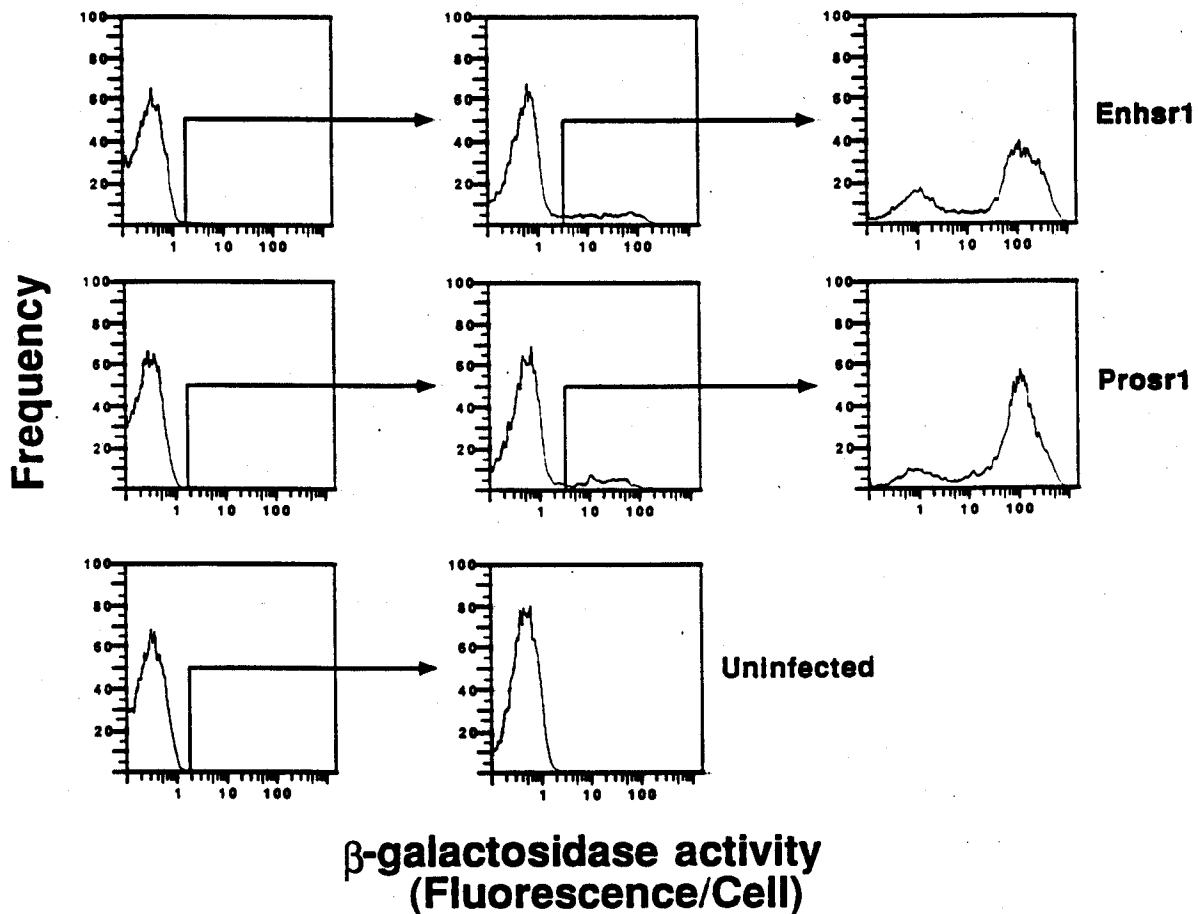
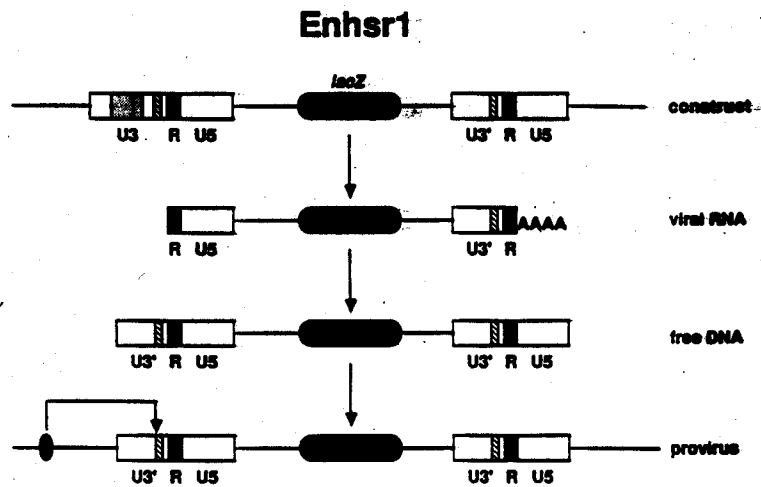


Figure 2. FACS enrichment of *lacZ*-expressing cells. Following cocultivation of 70Z/3 cells with the $\Psi 2$ Prosr1 and *Enhsr1* producer cell lines (see Methods), the 70Z/3 cells expressing *lacZ* were enriched by sorting based on their fluorescence due to cleavage of FDG to fluorescein. After expansion for 1 week, the cells were again stained by the FACS-GAL technique, reanalyzed, and again sorted for fluorescein-positive cells. This secondary sorted population was reanalyzed by FACS-GAL after 1 week of expansion in culture. Uninfected 70Z/3 cells were also stained and sorted by FACS-GAL. Cleavage was allowed to proceed for 2 hr, and false-positive cells were sorted. After expansion for 1 week in culture, the cells were stained and analyzed by FACS-GAL as before. We saw no increase in false-positive cells.

deletion is introduced into the 3' U3 in the initial retroviral construct, this deletion will be transmitted to both LTRs of the provirus following infection and integration (Yu et al. 1986). Since our constructs contain no internal promoter or other known regulatory elements, *lacZ* expression is dependent on readthrough transcription from an endogenous promoter with Prosr1 or *cis*-activation of the promoter in the viral LTR by a proximal enhancer element with Enhsr1 (Kerr et al. 1989).

Figure 2 shows that rare *lacZ*-expressing cells can be sorted by FACS-GAL following cocultivation of 70Z/3 cells with the Enhsr1 and Prosr1 virus producer lines. The initial percentage of *lacZ*⁺ cells following infection is quite low, as would be expected for these transcriptionally defective retroviruses. In fact, only infection with Enhsr1 gives percentages of positive cells significantly greater than the background frequency of false positives seen in uninfected cells. However, these rare *lacZ*⁺ cells can be sorted under sterile conditions by FACS-GAL and cultured, and a subsequent sort can be done to achieve a high degree of enrichment. Data in Figure 2 show that the initial percentage of *lacZ*⁺ cells following infection (Enhsr1: 0.5–0.6%, Prosr1: 0.1–0.2%) is enriched after two rounds of sorting and culture to 70–90% *lacZ*⁺ cells among the 70Z/3 cells. Thus, infection with either Enhsr1 or Prosr1, followed by sorting for *lacZ*⁺ cells via FACS-GAL, allows one to obtain a population of cells with random integrations of the *lacZ* reporter gene either under the control of nearby endogenous enhancer elements (Enhsr1) or downstream from active, endogenous promoters (Prosr1). In all uninfected mammalian cell lines examined, there is a low frequency of cells that are positive when loaded with FDG and analyzed by FACS-GAL, even though they express no *E. coli lacZ* gene. The positive phenotype of these cells is not heritable, since a sort of these rare positives from uninfected 70Z/3 by FACS-GAL sorting shows no enrichment; therefore, the phenotype of these rare cells is not stable.

The population of *lacZ*⁺ 70Z/3 cells obtained by FACS-GAL sorting contains many cells with independent integrations of *lacZ* at different sites in the genome, since individual clones derived from these pools have different characteristic patterns of *lacZ* expression. Figure 3 shows some representative *lacZ* expression patterns of fluorescence per cell for different clones derived from infections with Enhsr1 and Prosr1. These patterns of expression remain reproducible after repeated reculture and reanalysis, implying that they are stably transmitted to daughter cells following cell division (data not shown). This stability of expression was tested further by sorting low, medium, and high expressing cells from individual Enhsr1 clones and reexamining their pattern of expression after 1 week of culture. An example is shown in Figure 4. Each sort, when regrown, recapitulates the distribution of *lacZ* activity shown by the parent clone. The data in Figure 4 suggest that the broad *lacZ* expression patterns we see by FACS-GAL represent a controlled variation in transcription which, we suggest, is cyclical. Expression of many eukaryotic genes may also be under such cyclical control.

The ability of *lacZ* to be regulated in a differentiation-stage-dependent manner in individual Enhsr1-infected clones of 70Z/3 further demonstrates endogenous control of *lacZ* expression in integrations derived from Enhsr1. The pre-B-cell line, 70Z/3, differentiates *in vitro* to an IgM κ -expressing B cell after 24 hours of culture in lipopolysaccharide (LPS)-containing medium (Paige et al. 1978). One of the Enhsr1 clones, 7e17-17, shows nearly complete repression of *lacZ* expression after 24 hours of culture in medium with LPS (Fig. 5a), whereas other clones examined were either unresponsive to LPS treatment or showed only minor variations in expression (Kerr et al. 1989). To demonstrate that 7e17-17 cells had differentiated from pre-B to the B-cell stage while repressing *lacZ*, the cells were analyzed simultaneously for β -Gal activity as well as for IgM κ expression on the cell surface. In response to LPS, 7e17-17 cells acquire ex-

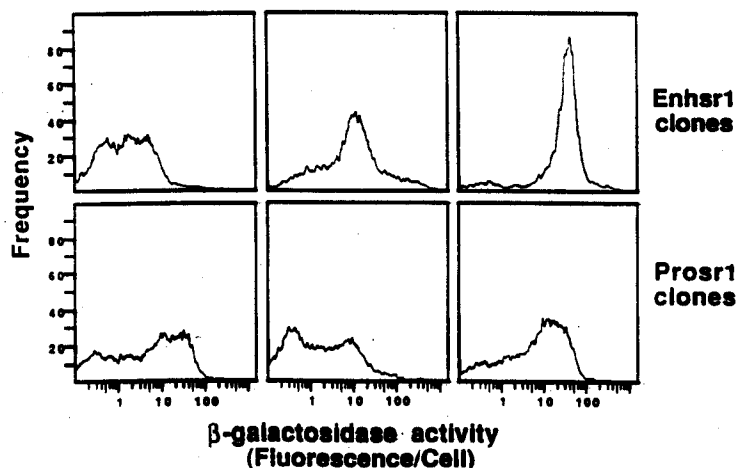


Figure 3. *lacZ*-expressing clones exhibit different patterns of activity. Cells of the individual Enhsr1 and Prosr1 clones were loaded hypotonically with FDG and analyzed by FACS after 2 hr of enzymatic cleavage at 4°C.

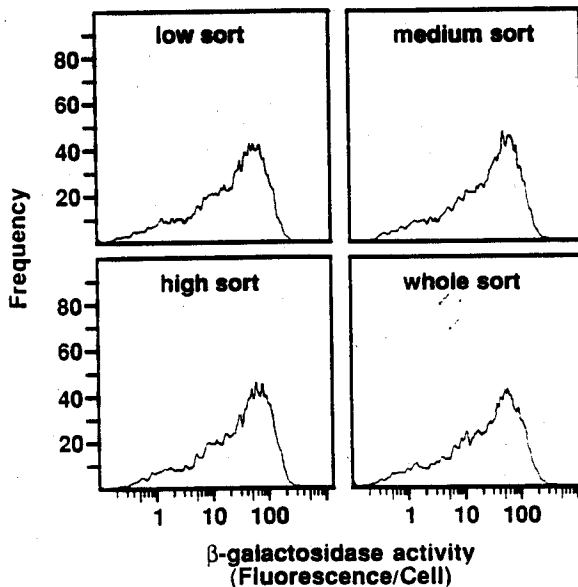


Figure 4. Stability of *lacZ* expression pattern in Enhsl1 clones. The Enhsl1 clone, 7e15, was stained and analyzed by FACS-GAL with enzymatic cleavage being halted at 1 hour by bringing the cells to 1 mM PETG. 50,000 cells were sorted from the 5–10, 45–50, and 90–95 percentiles of the 7e15 distribution (low, medium, and high, respectively). 50,000 cells from the whole, ungated distribution were also sorted. The sorted cells were grown for 1 week and reanalyzed by FACS-GAL as before.

pression of surface IgM κ expression while β -Gal is repressed (Fig. 5b). Therefore, LPS either simultaneously transmits positive and negative regulatory signals to a differentiating pre-B cell or initiates a cascade of events that result in both positive and negative regulation of genes in the nucleus of the pre-B cell. Induction of κ light-chain expression via NF- κ B is one well-documented effect of LPS on 70Z/3 cells (Sen and Baltimore 1986a,b). However, the repression of *lacZ* expression in 7e17-17 cells suggests that LPS can act to repress the transcription of a locus or loci expressed at the pre-B-cell stage of B-lineage differentiation. We believe this to be the first indication that LPS can play a negative regulatory role in the gene expression of B-lineage cells.

If Enhsl1 is integrated in a region of chromatin near a B-cell stage-specific enhancer in 70Z/3, *lacZ* expression would be induced by LPS. To increase our chances of finding integrations of Enhsl1 near enhancers active in the B-cell stage and not in the pre-B-cell stage, we infected cells in the presence of LPS and kept them in LPS up until cloning of *lacZ*⁺ cells by FACS-GAL sorting. Analysis of 230 clones derived from two different infections revealed several 70Z/3 clones with Enhsl1 integrations in chromatin regions where transcription is induced by LPS. Two examples, clones 7e129-3 and 7e131-3, show induction of *lacZ* expression following culture of the cells in LPS for a period of 24 hours (Fig. 6). Interestingly, only a portion of the cells within the clone appear to undergo an induction of *lacZ*

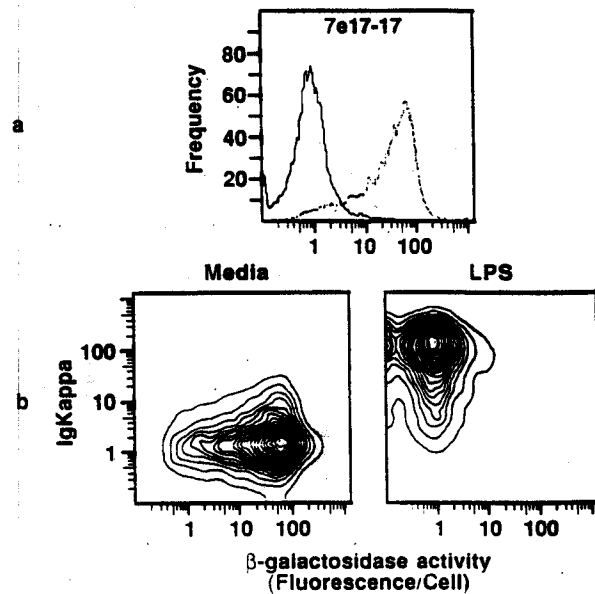


Figure 5. Regulation of *lacZ* and κ light-chain expression following LPS induction of the Enhsl1 clone, 7e17-17 cells. Equal numbers of 7e17-17 cells were cultured either in normal media or in media containing 10 μ g/ml of LPS for 24 hr. The cells were then stained as described above and analyzed on the FACS. (a) Graphs representing β -Gal activity of 7e17-17 cells after 24 hr in LPS-containing media (solid line) or in normal media (broken line). (b) Dual parameter FACS analysis of κ light-chain expression (Texas Red) vs. β -Gal (fluorescein) in 7e17-17 cells cultured in normal media or in LPS-containing media.

expression by LPS, despite the fact that all the cells have expressed surface IgM κ (demonstrated for 7e129-3 and 7e131-3, Fig. 6). DNA cell-cycle analysis of these Enhsl1 clones indicates that no stage of the cell cycle is specifically correlated with *lacZ* expression in these clones (W. Kerr et al., in prep.). Thus, there may be B-cell stage-specific enhancers that do not *cis*-activate constitutively during this stage of differentiation but rather can be considered "fluctuating" enhancers. The gene products such enhancers regulate may require only a transitory period of transcription to achieve the level of expression necessary for proper B-cell function. These fluctuating enhancers may be a necessary requirement for the rapid changes in phenotype that occur during the antigen-dependent phase of B-cell differentiation.

In Vivo Mammalian Gene Fusions with a Splice Acceptor *lacZ* Neo-exon

To improve the accuracy of measuring transcriptional control of endogenous genes in situ, we have developed a system in which transcriptional and translational fusion with the endogenous gene and its protein product are both required for *lacZ* expression. To accomplish this, we constructed AcLac, a *lacZ* artificial exon (neo-exon), by replacing the 5' terminus and

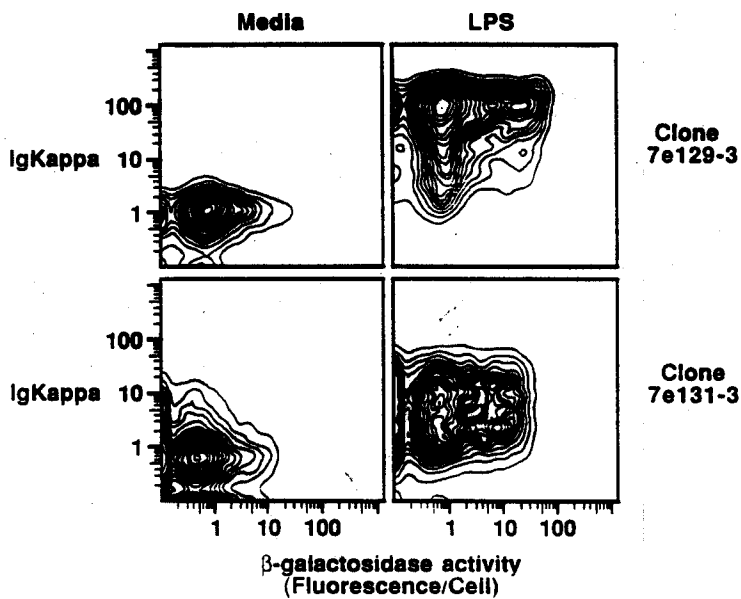


Figure 6. B-cell stage-specific expression of *lacZ* in the Enhsr1 clones, 7e129-3 and 7e131-3. Cells were cultured, stained, and analyzed similarly to 7e17-17 cells in Fig. 5.

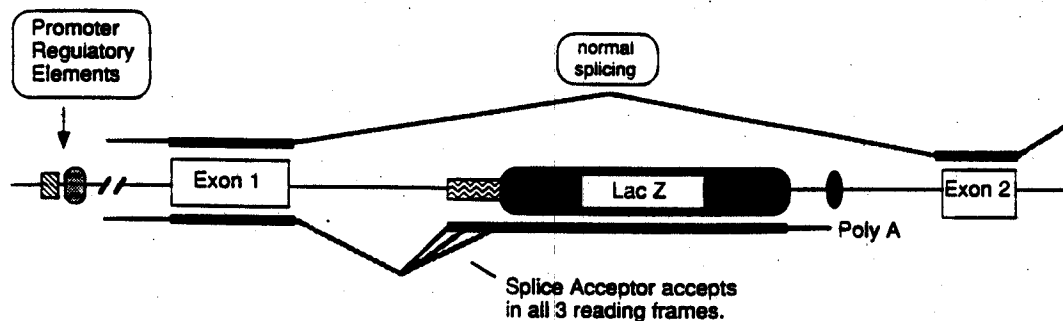
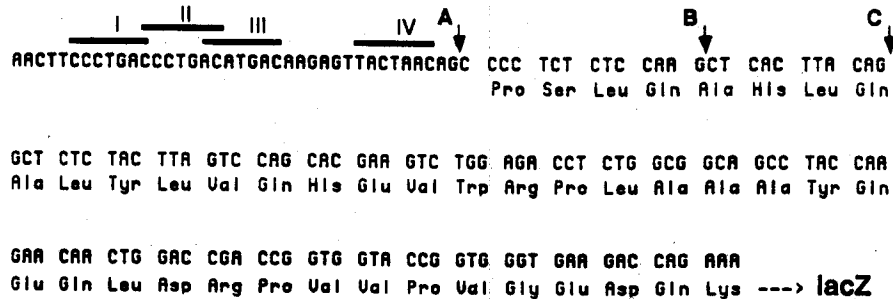


Figure 7. Splice acceptor/*lacZ*, AcLac, construct for in situ transcriptional/translational fusion to endogenous genes. The upper half of the figure is the sequence of the *env* splice acceptor/*lacZ* construct prior to transfection showing the four consensus lariet branch points (solid bars, I-IV) and the three potential consensus splice acceptor intron/exon boundaries (arrows, A-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 (Mann and Baltimore 1985; Lazo et al. 1987). Splice donation from an upstream exon to one of the three potential splice acceptor sites will fuse to the acceptor-*lacZ* sequence shown. The translated peptide sequence of the splice acceptor *lacZ* region is shown. The lower half of the figure is a schematic representation of a model fusion of *lacZ* to an upstream exon mediated by the MLV *env* splice acceptor. The endogenous promoter/regulatory elements drive transcription. The normal splicing pattern of the gene is depicted. Introduction of the splice acceptor/*lacZ* construct into an intron of an endogenous gene leads to interruption of normal splicing and direct transcriptional/translational fusion to the *lacZ* gene (only one of the three potential transcriptional fusions can lead to translation of a *lacZ* fusion protein). The polyadenylation signal is provided by the introduced construct.

initiator codon of *lacZ* with the Moloney *env* gene splice acceptor (Mann and Baltimore 1985; Lazo et al. 1987) (see Fig. 7). Following introduction of AcLac into cells, generation of enzymatically active β -Gal requires the following: (1) integration of the *lacZ* neo-exon into an intron of a gene, (2) interruption of the normal splicing of this gene by splicing of the *lacZ* neo-exon to an upstream exon resulting in transcriptional fusion of *lacZ* to the endogenous transcript, (3) initiation of translation at the AUG of the endogenous transcript with translational elongation into the coding frame of *lacZ*, and (4) folding of the β -Gal "domain" during assembly of this new multidomain protein in a manner that permits association into enzymatically active tetramers. Since AcLac is completely dependent on cellular sequences for expression of β -Gal, successful integrations should more accurately reflect the transcriptional control of the endogenous gene relative to *Enhsr1* and *Prosr1*.

We believe that generation of β -Gal expression via introduction of AcLac into the genome mimics exon shuffling that is proposed to take place in the molecular evolution of multidomain proteins (Blake 1978; Gilbert 1978). The basic tenet of these theories is that eukaryotic genes are collections of exons that were brought together by recombination within intron sequences. There are numerous examples of genes that have arisen via duplication and diversification of a single exon (e.g., Ig superfamily) (Gilbert 1985; Gilbert et

al. 1986). The evolution of genes derived from several different ancestral exons is less common, but the LDL receptor is one notable example (Sudhof et al. 1985). A corollary of this hypothesis is that the individual domains encoded by exons represent integrally folded protein units (Blake 1978). This would permit evolution of multidomain proteins with multiple functions to occur at a higher probability.

We introduced the AcLac neo-exon into the chromatin of a human embryo kidney cell line, 293, by transfection. After selecting for stable transfectants with G418, we sorted a population of *lacZ*⁺ cells from the pool of G418-resistant 293 transfectants. Following expansion of this *lacZ*⁺ pool of 293 cells, individual *lacZ*⁺ cells were cloned by FACS-GAL. Analysis of these individual clones by FACS-GAL showed patterns of expression that were unique to individual clones (Fig. 8), indicating that *lacZ* expression is under the control of different endogenous transcription elements. We find that expression can vary quite significantly within an individual clone, indicating that our original observations with *Enhsr1* and *Prosr1* were not an artifact introduced by the presence of Moloney viral sequences. This controlled variation in expression may

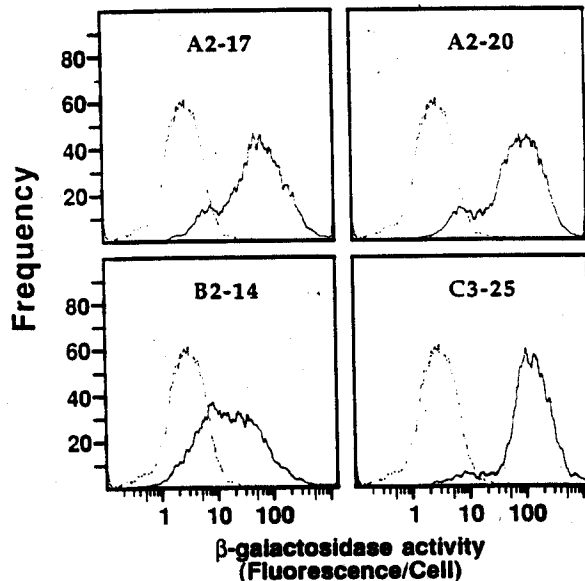


Figure 8. Splice acceptor/*lacZ* 293 clones display different patterns of β -Gal expression when analyzed by FACS-GAL. A suspension of each 293/AcLac clone was loaded hypotonically with FDG at 37°C. At 1 min, 2 ml of ice-cold isotonic medium was added to stop hypotonic loading. At 5 min, the cell suspension was brought to 1 mM PETG. The cells were kept on ice until analysis by FACS. The solid lines represent fluorescence per cell for the indicated AcLac/293 clone, and the dotted lines represent the parent line, 293, treated in a similar fashion.

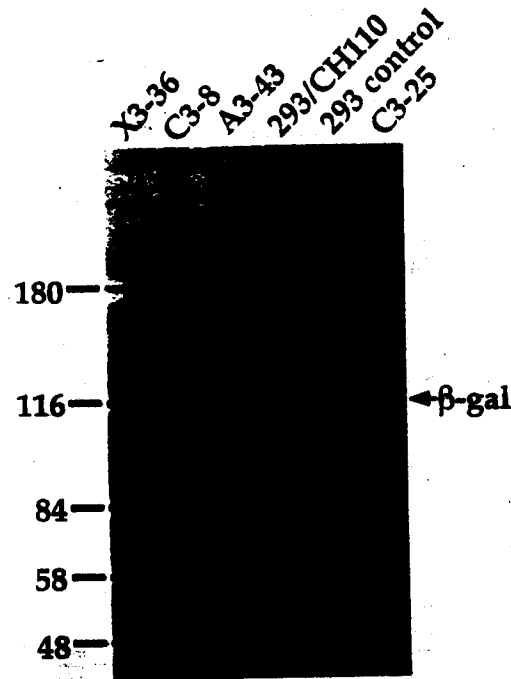


Figure 9. Clonal variation in molecular weight of splice acceptor/*lacZ* in vivo gene fusions. Immunoprecipitation analysis of β -Gal fusion proteins in four representative AcLac/293 clones (X3-36, C3-8, A3-43, C3-25) show different apparent molecular weights greater than that of *E. coli* β -Gal (revealed by Coomassie staining; indicated by an arrow). Purified *E. coli* β -Gal (Sigma) (5.5 μ g) 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 (*irpA-lacZ* fusion protein) (Hall et al. 1983) were also analyzed by immunoprecipitation analysis (293/CH110). Migration of molecular-weight standards is indicated at the left.

thus be reflective of the transcriptional control of many cellular genes.

To confirm that β -Gal activity in the 293 AcLac clones was derived by transcriptional and translational fusion with an endogenous gene, we biosynthetically labeled 293/AcLac clones, immunoprecipitated with a monoclonal anti- β -Gal, and analyzed these immunoprecipitates by SDS-PAGE. In Figure 9, we show the results of this analysis for four representative clones, which indicate that the β -Gal in these clones is of a higher apparent molecular weight than *E. coli* β -Gal or a *trpA-lacZ* fusion protein (CH110). The increased apparent molecular weight of β -Gal expressed by these cells indicates that it results from transcriptional and translational fusions with endogenous upstream exons in transcriptionally active genes.

Histological analysis of the X-Gal precipitates resulting from β -Gal activity in these 293/AcLac clones revealed that β -Gal activity is confined to discrete cellular and subcellular locations in each clone (see Fig. 10). Thus, in addition to the AcLac neo-exon forming novel multidomain proteins that have retained the enzymatic activity of β -Gal, these proteins retain functional targeting domains of the endogenous protein.

These clones have been readily isolated by FACS-GAL from three independent transfections of AcLac into 293 cells where the initial percentage of positive cells was about 1% in all three transfections (data not shown). Thus, the AcLac neo-exon can be seeded into the genome of a mammalian cell and generate enzymatically active fusion proteins with endogenous mammalian proteins that retain the subcellular localization determined by upstream encoded, endogenous

protein domains. Although the mechanism for molecular evolution of multidomain proteins cannot be directly studied, the introduction of the AcLac neo-exon into the genome of a cell provides a model system for molecular evolution of novel protein structures. Our ability to readily isolate novel multidomain proteins with AcLac provides examples of exon shuffling similar to what occurs in molecular evolution of multifunctional, multidomain proteins composed of different ancestral exons.

Transduced AcLac Forms Gene Fusions in B-lineage Cell Lines

We have inserted the AcLac construct into the same transcriptionally defective retrovirus construct that was used to construct Prosr1 (manuscript in prep.). If integration of the provirus containing AcLac is in an intron of an endogenous gene in the proper transcriptional orientation, β -Gal should be expressed from the integrated provirus in the same manner as transfection of AcLac was in 293 cells. Integration of the provirus derived from this retroviral construct will place *lacZ* in the chromatin of a cell without the known Moloney transcription control elements being present, and, thus, expression of *lacZ* should mimic the transcriptional regulation at this locus.

In Figure 11, we demonstrate the endogenous control of *lacZ* in clones derived from infection of the pre-B cell lines, 70Z/3 and NFS5.3. We demonstrate endogenous control of expression in these clones because they are differentially regulated. In the case of the 70Z/3 clones, 7a65 and 7a135, the expression of

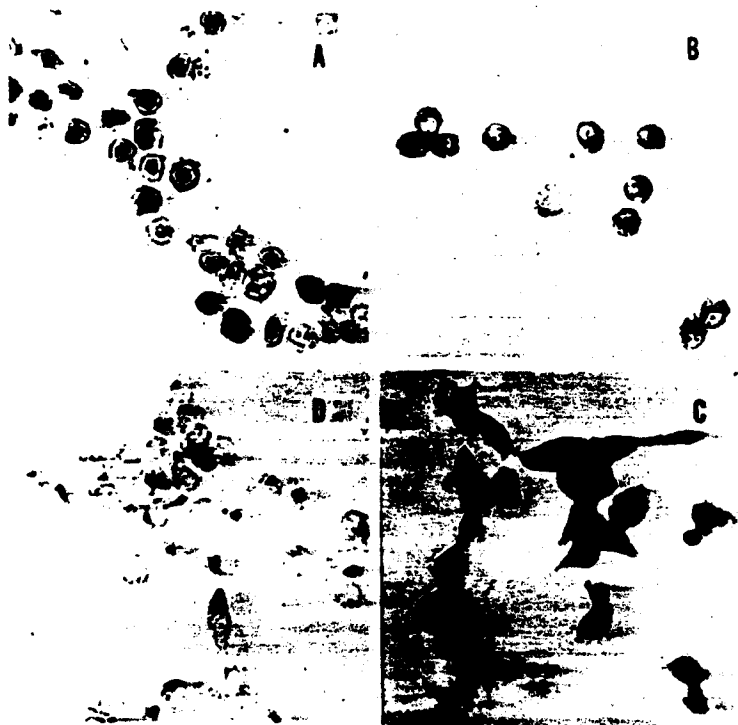


Figure 10. Subcellular localization of β -Gal with splice acceptor/*lacZ* in vivo gene fusions. (A) Nuclear localized β -Gal; (B) cytoplasmic β -Gal; (C) diffuse, whole cell β -Gal; (D) perinuclear localized β -Gal.

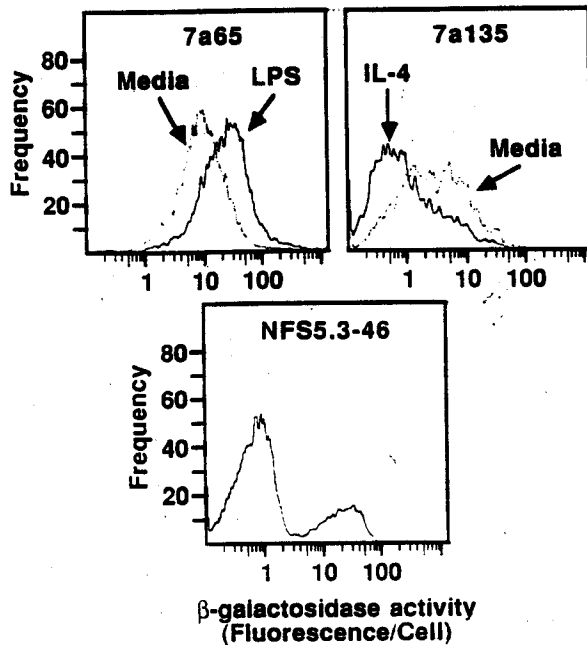


Figure 11. Differential expression of *lacZ* in pre-B-cell lines infected with Gensr1. LPS (7a65) and IL-4 (7a135) treatment of cells is described above. All clones were hypotonically loaded with FDG with enzymatic cleavage halted at 2 hr by bringing the cells to 1 mM PETG. The cells were then analyzed by FACS.

lacZ is modified by treatment of the cells with immunomodulators (IL-4, LPS). *lacZ* expression undergoes a threefold induction in 7a65 with LPS and is partially repressed in 7a135 by IL-4. In NFS5.3-46, the expression of *lacZ* fluctuates within the clone. That variation may be linked to cell-cycle events is presently under investigation.

DISCUSSION

We have demonstrated that the genome of a mammalian cell can be seeded with transcriptionally defective, *lacZ*-encoding retroviruses to isolate cells expressing *lacZ* under the control of endogenous transcription control elements. This goal was made feasible by the recently developed FACS-GAL technique, which permits rare cells expressing *lacZ* to be sorted from a population that is primarily negative (Nolan et al. 1988; Kerr et al. 1989). With the combination of the *lacZ* gene search retroviruses and FACS-GAL, we can study, in individual viable cells, the transcriptional control of a gene in its native context, thus permitting transcriptional control to be appreciated as a cell distribution rather than as a mean of this distribution (e.g., CAT assay, RNA assay).

Utilizing different *lacZ* gene search retroviruses (Enhsr1, Prosr1, and Gensr1) that have different requirements for *lacZ* expression, we have consistently found that the distribution of β -Gal activity differs both quantitatively and qualitatively among either Enhsr1, Prosr1, or Gensr1 clones, attesting to their transcrip-

tional control by cellular sequences flanking the provirus containing *lacZ*. The unique distributions we see are inherited by daughter cells, since the patterns are stable with continuous culture as well as in subclones. The wide variation in β -Gal activity we see in cells of a single clone appears to be controlled, since cells of either low, medium, or high β -Gal activity within the distribution of the clone recapitulate the distribution of the parental cells. The most likely explanation for the variation of β -Gal activity in these clones is fluctuations in the transcription rate of the reporter gene, *lacZ*. If this variation is indeed determined at the transcriptional level, our results would suggest that this is a common feature in the transcriptional control of cellular genes.

Convincing evidence for the endogenous control of *lacZ* is the identification of Enhsr1 clones where *lacZ* expression is dependent on the differentiation stage of 70Z/3 cells. We have identified an Enhsr1 clone (7e17-17) where *lacZ* is repressed when 70Z/3 cells differentiate to the B-cell stage under the influence of LPS. In addition, we have also found Enhsr1 clones (7e129-3, 7e131-3) where LPS-induced differentiation causes induction of *lacZ* expression; however, only a portion of the cells within these clones express *lacZ* while in LPS. We suggest that this B-cell stage enhancer has an on/off or fluctuating expression pattern, even though the 70Z/3 cells are all at the B-cell stage of differentiation. One could envision such fluctuating enhancers as being responsible for the regulation of genes that are expressed at the B-cell stage, which are later down-regulated if the cell encounters its antigen during the antigen-dependent phase of B-cell differentiation (e.g., Ia expression).

We were also able to obtain endogenously controlled expression of *lacZ* through transfection with a splice acceptor/*lacZ* neo-exon, AcLac, or by infection with a retrovirus containing AcLac, Gensr1. Since expression of *lacZ* in these clones is derived by both transcriptional and translational fusion with an endogenous gene, we believe that these in situ gene fusions with the *lacZ* reporter gene offer a more accurate measurement of endogenous transcriptional control. As with Enhsr1 and Prosr1 clones, we find that AcLac and Gensr1 clones have unique and characteristic distributions of β -Gal activity, reflecting the transcriptional regulation of the loci where they have integrated. We have also found that certain Gensr1 clones differentially regulate expression of *lacZ*, suggesting that Gensr1 will allow us to tag genes expressed at different stages in lymphoid differentiation.

We believe that, in addition to acting as an in situ measure of transcription, the expression of β -Gal following the introduction of the AcLac neo-exon into the genome mimics events that take place in the exon-shuffling model of molecular evolution. Because these in vivo gene fusions with *lacZ* retain features of both the introduced AcLac neo-exon (enzymatic activity) and the endogenous exon (subcellular targeting), we believe these experiments demonstrate that exon shuffling is a valid mechanism for evolution of novel

multidomain protein structures. The high frequency with which we were able to obtain β -Gal expression in three independent transfections with the AcLac neo-exon suggests that these transfections may provide a model system for examining the requirements for successful and aberrant exon-shuffling events.

The development of the *lacZ* gene search retroviruses and the FACS-GAL assay come at a time when the *E. coli* gene, *lacZ*, is being used in a variety of systems to study developmentally regulated gene expression in higher eukaryotes (O'Kane and Gehring 1987; Allen et al. 1988; Gossler et al. 1989). The *lacZ* gene search viruses offer an efficient means to introduce *lacZ* into the genome of mammalian cells, and FACS-GAL permits the rare cells expressing *lacZ* to be isolated following infection. The combination of these two techniques provides a novel means to identify and study transcriptionally active regions of the mammalian genome and, thus, will facilitate the study of mammalian gene expression.

ACKNOWLEDGMENTS

We thank Dr. Leonore Herzenberg (Stanford University) for helpful discussion. We also thank Bruce Blakeley, Dr. Simon Hughes, and Dr. Helen Blau for assistance with microscopy and photography. In addition, we thank C.M. Tarlinton, G. Jager, C. Rattazzi, and J. Johnsen for expert technical assistance. W.G.K. was supported by the National Institutes of Health training program in immunology, AI-07290. This work was supported by NIH grant CA-42509.

REFERENCES

- Allen, N.D., D.G. Can, S.C. Barton, S. Hettle, and M.A. Surami. 1988. Transgenes as probes for active chromosomal domains in mouse development. *Nature* 333: 852.
- Blake, C.C.F. 1978. Do genes-in-pieces imply proteins-in-pieces? *Nature* 273: 267.
- Cepko, C.A., B.E. Roberts, and R.C. Mulligan. 1984. Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell* 37: 1053.
- Gilbert, W. 1978. Why genes in pieces? *Nature* 271: 501.
- . 1985. Genes-in-pieces revisited. *Science* 228: 823.
- Gilbert, W., M. Marchionni, and G. McKnight. 1986. On the antiquity of introns. *Cell* 46: 151.
- 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.
- Graham, F.L., J. Smiley, W.C. Russell, and R. Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* 36: 59.
- Graham, F.L., S. Bacchetti, and R. McKinnon. 1980. Transformation of mammalian cells with DNA using the calcium technique. In *Introduction of macromolecules into viable mammalian cells* (ed. R. Baserga et al.), p. 3. Alan R. Liss, New York.
- Hall, C.V., P.E. Jacob, G.M. Ringold, and F.J. Lee. 1983. Expression and regulation of *Escherichia coli lacZ* gene fusions in mammalian cells. *Mol. Appl. Genet.* 2: 101.
- Kerr, W.G., G.P. Nolan, and L.A. Herzenberg. 1989. *In situ* detection of transcriptionally active chromatin and genetic regulatory elements in individual viable mammalian cells. *Immunity* (in press).
- Lazo, P.A., V. Prasad, and P.N. Tsichlis. 1987. Splice acceptor site for the *env* message of Moloney murine leukemia virus. *J. Virol.* 61: 2038.
- Mann, R. and D. Baltimore. 1985. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. *J. Virol.* 54: 401.
- Nolan, G.P.J. 1989. "Individual cell gene regulation studies and *in situ* detection of transcriptionally active chromatin using fluorescence-activated cell sorting with a viable cell fluorographic assay." Ph.D. thesis. Stanford University.
- 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.* 85: 2603.
- O'Kane, C.J. and W.H. Gehring. 1987. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci.* 84: 9123.
- Paige, C.J., P.W. Kincade, and P. Ralph. 1978. Murine B cell leukemia with inducible surface immunoglobulin expression. *Science* 121: 641.
- Sen, R. and D. Baltimore. 1986a. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 46: 705.
- . 1986b. Inducibility of κ immunoglobulin enhancer-binding protein NF- κ B by a post-translational mechanism. *Cell* 47: 921.
- Southern, P. and P. Berg. 1982. A new dominant hybrid selective marker for higher eukaryotic cells. *J. Mol. Appl. Genet.* 1: 327.
- Sudhof, T.C., J.L. Goldstein, M.S. Brown, and D.W. Russell. 1985. The LDL receptor gene: A mosaic of exons shared with different proteins. *Science* 228: 815.
- Varmus, H.E. 1982. Form and function of retroviral proviruses. *Science* 216: 812.
- Yu, S.F., T. von Ruden, P.W. Kantoff, C. Garber, M. Sieberg, U. Ruther, W.F. Anderson, E.F. Wagner, and E. Gilboa. 1986. Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. *Proc. Natl. Acad. Sci.* 83: 3194.