Nucleic Acids Research, 1998, Vol. 26, No. 12 **2849–2858**

Stochastic, stage-specific mechanisms account for the variegation of a human globin transgene

Timothy A. Graubert, Bruce A. Hug, Robin Wesselschmidt+, Chih-Lin Hsieh[1], Thomas M. Ryan[2], Tim M. Townes[2] and Timothy J. Ley*

Division of Bone Marrow Transplantation and Stem Cell Biology, Departments of Medicine and Genetics, Washington University School of Medicine, Campus Box 8007, 660 South Euclid Avenue, St Louis, MO 63110-1093, USA, [1]Department of Urology, Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA, USA and [2]Department of Biochemistry and Molecular Genetics, University of Alabama-Birmingham, Birmingham, AL, USA

Received March 25, 1998; Revised and Accepted May 5, 1998

ABSTRACT

The random insertion of transgenes into the genomic DNA of mice usually leads to widely variable levels of expression in individual founder lines. To study the mechanisms that cause variegation, we designed a transgene that we expected to variegate, which consisted of a β-globin locus control region 5′ HS-2 linked in tandem to a tagged human β-globin gene (into which a Lac-Z cassette had been inserted). All tested founder lines exhibited red blood cell-specific expression, but levels of expression varied >1000-fold from the lowest to the highest expressing line. Most of the variation in levels of expression appeared to reflect differences in the percentage of cells in the peripheral blood that expressed the transgene, which ranged from 0.3% in the lowest expressing line to 88% in the highest; the level of transgene expression per cell varied no more than 10-fold from the lowest to the highest expressing line. These differences in expression levels could not be explained by the location of transgene integration, by an effect of β-galactosidase on red blood cell survival, by the half life of the β-galactosidase enzyme or by the age of the animals. The progeny of all early erythroid progenitors (BFU-E colony-forming cells) exhibited the same propensity to variegate in methylcellulose-based cultures, suggesting that the decision to variegate occurs after the BFU-E stage of erythroid differentiation. Collectively, these data suggest that variegation in levels of transgene expression are due to local, integration site-dependent phenomena that alter the probability that a transgene will be expressed in an appropriate cell; however, these local effects have a minimal impact on the transgene's activity in the cells that initiate transcription.

INTRODUCTION

When fragments of DNA containing one of the human β-like globin genes are randomly integrated into the genome of transgenic mice, the result is generally very low level expression of the transgene in a small percentage of transgenic lines (1–6); however, these transgenes are appropriately expressed at the correct developmental stage and specifically in red blood cells, suggesting that the tissue and development-specific information required for targeted expression lies within or near the genes themselves (1–6). The very low levels of expression seen in a small fraction of the animals prompted a search for DNA elements that were required for more proper, high levels of expression in a higher percentage of the animals. The first clue that these elements might exist came from the work of Tuan *et al.* (7,8) and Groudine and colleagues (9), who showed that erythroid-specific DNase I hypersensitive sites lie upstream from the globin gene cluster, and may contribute to organization of the erythroid domain of the cluster. The finding of an Hispanic patient with γδβ-thalassemia who carried a deletion spanning a region 6–30 kb upstream from the ε-globin gene suggested that this region might be critical for activity of the β-globin locus (10). Grosveld and colleagues (11) indeed showed that linkage of this upstream region to globin genes in transgenic mice greatly increased the percentage of mice that expressed the transgene, and also increased the level of output of the linked globin genes at each integration site, so that transgenes demonstrated 'integration site-independent, copy number-dependent' expression. However, more recent studies have suggested that linkage of single hypersensitive sites to globin genes does not eliminate integration site-specific variation (12–19).

Recent studies by Whitelaw and Martin and colleagues (20–25) have suggested mechanisms by which transgenic variegation may occur. In their studies, globin regulatory sequences were linked with a Lac-Z reporter gene, and transgenic animals were produced. Collectively, these studies have shown that the

*To whom correspondence should be addressed. Tel: +1 314 362 8831; Fax: +1 314 362 9333; Email: timley@im.wustl.edu

+Present address: Genome Systems Inc., 8620 Pennell Drive, St Louis, MO, USA

variation in output seen among different integration sites can largely be explained by different percentages of expressing red blood cells in the transgenic animals (20). Their results also suggested that addition of a locus control region element to a globin transgene seems to increase expression by increasing the probability that a high percentage of red blood cells will express the transgene, not by increasing the level of expression per cell (21,22). These studies complement work that has been performed in tissue culture cells, where we and others have observed that LCR elements seem to act by increasing the probability that an individual integration event will be transcriptionally productive, not by increasing the output of the promoter linked to the LCR element (21,26–28).

In this report, we linked human 5′ HS-2 to a human β-globin transgene that contained a Lac-Z cassette within the β-globin gene. Previous reports suggested that expression of this transgene would be variegated (12,16,19). We created 10 transgenic founder lines with this transgene, and indeed found this to be the case. The basis for this wide variation in transgene expression was found to be primarily due to alterations in the percentages of transgene-expressing cells within each founder line; however, small differences in the output of the transgenes were also noted. We observed that the 'decision' to variegate must be made after the BFU-E stage of erythroid differentiation, suggesting that the regulatory elements in the transgene must interact with stage-specific information in the developing erythroid cell, and that individual integration sites may have differing abilities to respond to that information.

MATERIALS AND METHODS

Production of transgenic mice

The transgene was assembled in pUC19, and contains the 1.9 kb *Kpn*I–*Pvu*II human 5′ HS-2 fragment (previously shown to contain all known HS-2 activity; 16), upstream from the marked β-globin transgene (Fig. 1). The 5′ part of the human β-globin gene (an *Hpa*I–*Nco*I fragment extending from –800 to +46) was fused to a 3.0 kb *Nco*I–*Bgl*II fragment obtained from pLacD. The 3′ part of the β-globin gene consisted of a 2.8 kb *Bam*HI–*Xba*I fragment that contains the 3′ end of exon 2, intron 2, exon 3 and 3′ flanking sequence. 5′ HS-2 was inserted in the genomic (5′→3′) orientation with respect to the transgene.

The transgene was isolated from the plasmid vector backbone by cleavage with *Xho*I and *Sal*I, gel purified, concentrated on a DEAE mini column (Elutip, Schleicher and Schuell), and resuspended in injection buffer (5 mM NaCl, 5 mM Tris pH 7.5, 0.25 mM EDTA). Pronuclear injections of fertilized eggs derived from inbred C57Bl/6 mice, or from B6XC3H mice, were performed. Lines A, B and C were made in the pure Bl/6 background, and lines D–J were made in the hybrid background. Founders were identified by Southern blotting of tail DNA. All 10 founder mice passed the transgene through the germline. Only mice that were F1 or beyond were used for analysis in subsequent studies.

Fluorescence *in situ* hybridization (FISH)

The entire plasmid containing the β-globin Lac-Z transgene (pTR159) was nick-translated with Bio-11-dUTP (Sigma) for *in situ* hybridization. Procedures for hybridization, washing, blocking, detection and amplification were described previously (29,30) with minor changes. After denaturation for 15 min at 37°C, the biotinylated probe (at 15 μg/ml) was mixed in 0.1 mg/ml salmon sperm DNA, 50% formamide, 2× SSC and 10% dextran sulfate, and hybridized to chromosome spreads for 16 h at 37°C. After hybridization, the slides were washed three times in 4× SSC, 0.1% Tween 20 for 5 min each time at 45°C, followed by one wash in 0.1× SSC, 0.1% Tween 20 for 5 min at 45°C. Sites of hybridization were detected with avidin conjugated with FITC (Molecular Probes, Eugene, OR) followed by one round of amplification. Metaphase spreads were counterstained with 100 ng/ml propidium iodide, mounted with 90% glycerol and 2.3% DABCO, then scanned, analyzed and photographed on an Olympus BX-60 fluorescent microscope.

Preparation of metaphase spreads

Skin fibroblasts from the transgenic and non-transgenic mice were cultured to obtain metaphase preparations. Fibroblasts were cultured in Dulbecco's Modified Eagle Medium with 10% fetal calf serum and penicillin–streptomycin. Cells were harvested for metaphase preparation by standard methods.

Preparation of peripheral blood or whole animals for determination of β-galactosidase activity

Peripheral blood was obtained from mice between 1 and 24 months of age by retro-orbital bleeding. For X-gal staining, 2 μl of blood was fixed in 1 ml of 2% formaldehyde/0.2% glutaraldehyde in phosphate buffered saline (PBS) at 4°C for 10 min, washed twice with PBS and then stained in 100 μl of a mixture containing 1 mg/ml X-gal, 4 mM potassium ferricyanide, 4 mM potassium ferrocyanide and 2 mM $MgCl_2$ at room temperature for 2–4 h. The reaction was stopped by adding 1 ml of PBS and cooling to 4°C; 100–200 μl of cells were then prepared for examination using a cytospin apparatus.

For quantitative determination of β-galactosidase activity, 3 μl of peripheral blood was lysed in 300 μl of extraction buffer (50 mM Tris pH 8.0, 1 mM DTT and 1% Triton X-100). Extracts were vigorously vortexed, and then 50 μl of lysate was used to normalize for hemoglobin content by measuring the OD_{405}. Thirty μl of normalized lysate was mixed with 270 μl of ONPG cocktail, exactly as described (43). Reactions were incubated at room temperature for 4–24 h. In each assay, samples were performed in duplicate, and 0–200 defined units of *Escherichia coli* β-galactosidase (Boehringer) were assayed to ensure that all samples were measured in the linear range of the assay (OD_{420} measurements of 0.0–1.0). Reactions were stopped by adding 500 μl of dH_20 and immediately measured on a spectrophotometer at OD_{420}.

For determination of β-galactosidase expression in whole organs, mice were anesthetized and the left ventricle was cannulated. The animals were perfused exhaustively (≥5 min, until distinct organ blanching was observed), and then organs were harvested and frozen at –70°C. The organs were thawed in extraction buffer at 4°C, disrupted with a tissue grinder, and then sonicated. Cleared supernatants were evaluated for total protein content using the BioRad assay. Equal amounts of total protein were then evaluated using the ONPG assay; substrate conversion for the highest expressing tissue (peripheral blood) was always performed in the linear range of the assay. All assays were performed in duplicate.

Flow cytometry

Reticulocytes were enumerated on a FACScan (Becton Dickinson) after staining with acridine orange using standard techniques (31). Reticulocytosis was induced with a single subcutaneous injection of phenylhydrazine (Sigma, St Louis, MO) (60 mg/kg), as previously described (32). Reticulocytopenia was induced by a single exposure of the mice to 750 cGy from a ^{137}Cs source.

Flow cytometric examination for β-galactosidase activity was performed using the FACS-FDG staining protocol developed by Fiering *et al.* and Nolan *et al.* (33,34). Basically, 2 µl of peripheral blood was mixed with 1 ml of FACS buffer (PBS containing 10 mM HEPES pH 7.0, 4% fetal calf serum and 300 µM chloroquine). One hundred µl of cells in this mixture were warmed to 37°C and then osmotically loaded by adding 100 µl of fluorescein di-β-D-galactopyranoside (FDG) (Molecular Probes, Eugene, OR) diluted 1:10 in dH_2O, with vigorous mixing. After 30 s of osmotic shock at 37°C, 1.2 ml of FACS buffer at 4°C was added, and the reaction was allowed to proceed for exactly 5 min on ice. The reaction was stopped by adding 30 µl of 1 mM phenylethyl β-D-thiogalactopyranoside to the reaction. Flow cytometric analysis of the gated red cell population was then performed immediately. The duration of substrate loading and enzymatic conversion were carefully controlled to ensure that assays were performed in the linear range. Cell concentration was kept low ($<10^6$ cells/ml) in order to minimize transfer of fluorescein from Lac-Z$^+$ cells to Lac-Z$^-$ cells. One hundred thousand events were collected per sample in order to increase the statistical significance of positive events detected in the low expressing lines.

Methylcellulose culture of bone marrow progenitor cells

Bone marrow cells were flushed from mouse femurs. After washing once with PBS, 5×10^4 cells were plated in M3230 methylcellulose (Stem Cell Technology, Vancouver, Canada) supplemented with erythropoietin (3 U/ml), c-kit ligand (2.5 ng/ml) and IL-3 (2.5 ng/ml). After 3 days (for CFU-E) or 7 days (for BFU-E) of culture at 37°C, colonies were counted and then overlayed with X-gal reagents. Colonies were scored and photographed after an additional 24 h incubation at 37°C.

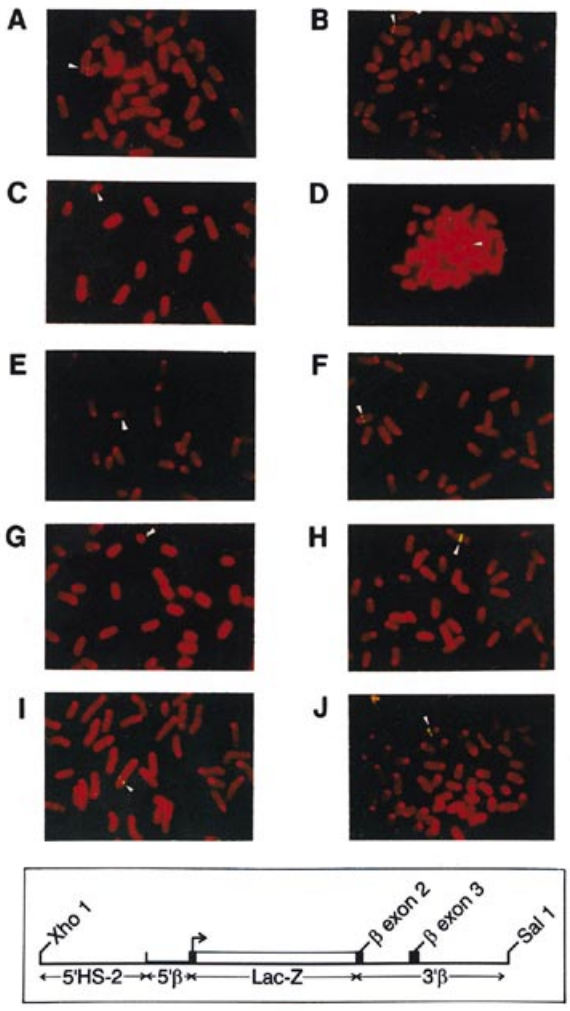

Figure 1. FISH analysis of the integrated transgene. A diagram of the HS-2/β-globin-Lac-Z transgene is shown at the bottom of this figure. The entire DNA fragment was used as a probe for FISH of cells derived from skin fibroblasts of each of the 10 founder lines (A–J). The location of each transgene integration site is shown with an arrowhead. An interpretation of the sites of integration is provided in Table 1.

RESULTS

The β-Lac-Z transgene and production of founder animals

To study the molecular basis of variegation, we wished to create a transgene that would be highly variegated, but expressed in most integration sites, and that would allow us to measure transgene output in individual cells. We therefore assembled the transgene shown in Figure 1. A previously characterized human β-globin LCR 5′ HS-2 fragment was linked upstream in tandem to a human β-globin transgene that contained a Lac-Z cassette. The sequences between the initiation codon and the *Bam*HI site of the β-globin gene were removed, and the Lac-Z cassette was inserted in this site. This transgene contains the known β-globin regulatory sequences in the 5′ flanking region, in IVS-2 and 3′ to the gene. Previous studies suggested that the output of this transgene should vary widely from one integration site to the next (12,16,19). Ten founder lines were created with this transgene. Lines A, B and C were made in inbred C57Bl/6 mice, and lines D–J were made in a hybrid C3H×C57Bl/6 background. Southern blot analysis revealed that the copy numbers of the transgene per haploid genome ranged from 4 to 167; all of the transgenes were arranged as head-to-tail concatamers in the genome, and all had non-rearranged transcription units by Southern analysis (data not shown).

To define the chromosomal integration sites of the transgene in all 10 founder lines, the entire transgene was used as a probe for FISH analysis. The transgene did not hybridize with DNA sequences in non-transgenic mice (data not shown). Hybridization signals were detected in all 10 founder lines, and the intensity of these signals correlated with Southern blot analysis (Fig. 1). Each integration site was unique. Only one of the 10 founder lines exhibited transgene insertion in the centromere; this line demonstrated very low expression of the transgene (Table 1). Five other transgenic lines exhibited low level expression, but none of these had centromeric transgene insertion sites.

Table 1. Characteristics of transgenic lines

Founder line	Transgene copy number (per haploid genome)	Chromosomal integration (site/size of chromosome)	Expression (total adjusted Lac-Z activity)	Adjusted expression/ copy number
A	57	distal/large	1570	27.5
B	25	centromeric/large	5.6	0.22
C	4	proximal/small	97	24.3
D	111	mid/medium	ND	ND
E	4	mid/small	151	37.8
F	35	proximal/medium	ND	ND
G	10	mid/small	1790	179
H	167	mid/large	ND	ND
I	6	mid/large	0.8	0.13
J	28	mid/medium	ND	ND
K	0	–	0	0

ND, none detected over background.

Transgene expression was highly variable and depended on the integration site

Peripheral blood was obtained from young adult animals (ranging from 2 to 4 months of age) from all 10 founder lines and wild type controls, and lysates from red blood cells were prepared and normalized for hemoglobin content. The total β-galactosidase activity in these lysates was then measured using a quantitative ONPG assay. Different quantities of lysates were assayed to be certain that all measurements were performed in the linear range of the assay. Mice derived from all 10 founder lines were assayed simultaneously on at least three separate occasions. The results from one representative analysis are shown in Table 1 (and Fig. 5A, in graphic form). Two of the founder lines expressed very high levels of β-galactosidase activity in the peripheral blood (lines A and G), two lines contained intermediate levels of activity (lines C and E), and six lines contained very low or undetectable levels of β-galactosidase activity (lines B, D, F, H, I and J). The total variation in expression of β-galactosidase was >1000-fold using the quantitative assay. Transgene expression did not correlate positively with transgene copy number (Table 1).

Because the variation in transgene expression was extreme among the 10 founder lines, we wanted to be certain that this β-globin transgene appropriately targeted the adult red blood cell compartment. Therefore, young adult mice from the high expressing and intermediate expressing transgenic lines were examined for β-galactosidase activity in all organs. Peripheral blood lysates were prepared from adult animals ranging from 8 to 12 months of age, and then the animals were extensively perfused with PBS to reduce the amount of blood present in all peripheral organs. Lysates of all organs were made and normalized for total protein content. β-galactosidase activity was then determined in each organ using the ONPG assay. All assays were performed in linear range. For each founder line, the level of β-galactosidase in the peripheral blood lysate was defined as 100%. The normalized values of the perfused organ lysates are shown in Figure 2. In all of these founder lines, the highest level of activity was found in peripheral blood, followed by the spleen and bone marrow (which both contain large numbers of erythroid precursors). Smaller amounts of β-galactosidase activity (<5% of peripheral blood levels, in most cases) were detected in all other organs. The low level of β-galactosidase activity in non-erythroid tissues is most likely accounted for by residual RBC contamination following organ perfusion. Within the bone marrow itself, we detected Lac-Z staining only in cells that had the morphology of erythroid precursors and mature erythroid cells (data not shown). These results show that the transgene was appropriately targeted to the erythroid compartment of adult animals, and that the mutation introduced into the β-globin transgene did not alter its targeting properties.

Variations in transgene expression are predominantly due to variations in the percentages of expressing red blood cells

To determine whether the wide variation in expression of the transgene was due to differences in the percentages of red blood cells that expressed the transgene or due to differences in output of the transgene in all of the red blood cells, we obtained peripheral blood from young adult animals from all 10 founder lines simultaneously. We either fixed these cells and stained them with X-gal reagents, or introduced the biochemical substrate FDG into red cells by osmotic shock and measured β-galactosidase activity using a flow cytometric method. The two independent techniques yielded results that were virtually identical. The results of the X-gal staining of peripheral blood are shown in Figure 3, and the results of flow cytometric analyses in Figures 4 and 5B. The two founder lines with the largest amounts of β-galactosidase activity in peripheral blood lysates (lines A and G), demonstrated very high percentages of red blood cells containing detectable amounts of enzymatic activity (78 and 88%, respectively). The two intermediate expressing lines (C and E) demonstrated intermediate percentages of FDG^+ red blood cells (14 and 12%, respectively). The six lines with very low levels of enzyme activity in peripheral blood demonstrated very small percentages of expressing red blood cells in the periphery (0.3–1.8%; Fig. 4). The plots shown in Figure 4 present β-galactosidase activity on the *x*-axis and an irrelevant parameter on the *y*-axis, to make it possible to see the mean level of fluorescence per cell in the lines that have only a very small percentage of FDG^+ cells. The percentage of positive red cells in each founder line, and the mean fluorescence intensity of the

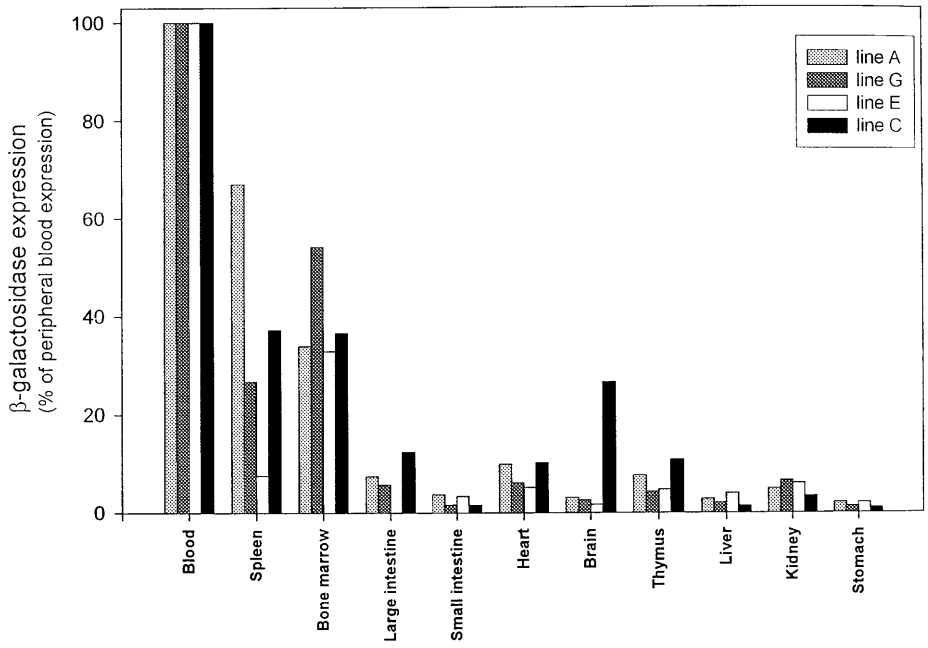

Figure 2. β-galactosidase expression in organ lysates from founder lines. Individual mice from founder lines A, C, E and G were obtained, and the anesthetized mice were exhaustively perfused with PBS using standard techniques. The organs were then harvested and lysates were made and normalized for protein content. Equal amounts of protein were evaluated for total β-galactosidase content with the ONPG assay. The amount of β-galactosidase activity was defined as 100% in the peripheral blood lysate. The levels of detectable β-galactosidase in the other organs is represented as a percentage of that value. Note that the highest level of expression in every animal is in the peripheral blood, followed by the bone marrow and spleen. Non-hematopoietic organs invariably contain small or nearly undetectable amounts of β-galactosidase activity, indicating that the transgene is correctly targeted to hematopoietic tissues in all founder lines tested.

expressing cells, are summarized in Figure 5B. From these data, it is evident that the mean fluorescence intensity per positive cell (a function of the number of β-galactosidase molecules per cell) (34) varies only 10-fold between the lowest expressing line and the highest line. The percentage of positive cells varies by >250-fold, suggesting that most of the variegation can be best explained by differences in the percentage of cells that express the transgene in the peripheral blood. We also detected a correlation between the percentage of positive cells and the mean fluorescence intensity of the positive cells (R = 0.8).

The β-galactosidase enzyme has a long half life in adult red blood cells, and has no adverse effects on red cell survival

To be certain that our measured transgene activity was not affected by alterations in red cell sub-populations, or by altered red blood cell half life due to transgene expression, we performed complete blood counts on at least two independent animals from each founder line, and determined the reticulocyte counts from at least two independent animals from each founder line. The hemoglobin content in the peripheral blood from all examined animals was within normal limits, and the reticulocyte counts, as determined both by manual counting, and by flow cytometry, were <3% for all examined animals (data not shown).

To determine whether the detected β-galactosidase was present at higher levels in very young red cells (reticulocytes), we induced hemolytic anemia in several low to intermediate expressing animals by treating the animals with phenylhydrazine, which induced reticulocyte counts of 50–80% 3 days after administration of the drug (data not shown). The percentage of β-galactosidase cells in the peripheral blood, and the Lac-Z content per red cell were unaltered by phenylhydrazine-induced hemolysis (data not shown). In addition, we irradiated (750 cGy) several high expressing and intermediate expressing animals, and then measured the reticulocyte counts and percentages of positive cells 1 week later. In all cases, no reticulocytes were detected in the irradiated animals. However, the percentage of positive cells was unchanged from the pre-irradiation values (data not shown). In sum, these data suggest that the Lac-Z protein has a very long half life in red blood cells (perhaps as long as the red cell itself), and that high level expression of β-galactosidase in red blood cells does not alter their half life *in vivo*.

Minimal changes in transgene expression as a function of time

Whitelaw and colleagues have previously reported that globin transgenes can exhibit age-dependent silencing (35). We selected three lines that permitted accurate measurements of β-galactosidase activity in peripheral blood, and followed mice for up to 2 years. Levels of transgene expression and the frequency of positive cells were similar from animal to animal, and revealed little or no reduction in levels of expression as a function of time (Fig. 6). Many of the other animals and additional founder lines were followed for >1 year, and similar levels of transgene expression were noted in all animals at all time points measured (data not shown). In particular, 11 animals were studied for up to 1 year for line B, which demonstrated centromeric insertion of the transgene. All of these animals had similar patterns of expression at all time points tested (data not shown). Although these data differ from the findings of Whitelaw and colleagues (35), there are a number of technical differences that make a meaningful comparison

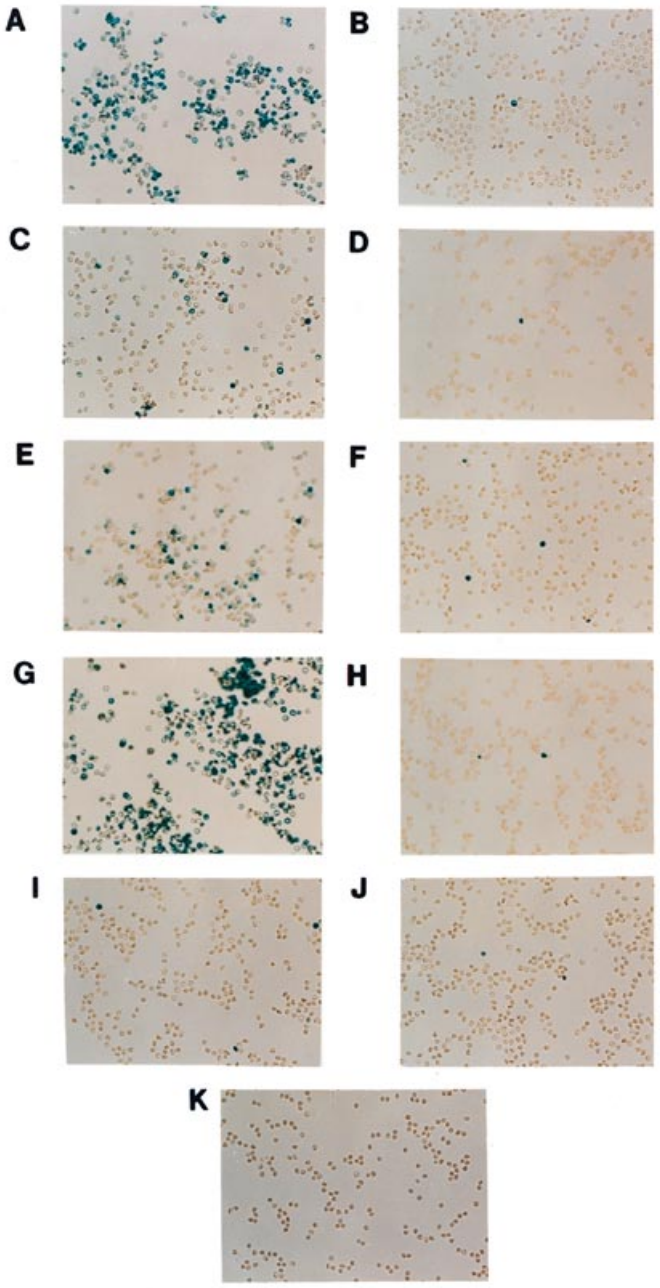

Figure 3. X-gal staining of peripheral blood from adult animals in all 10 founder lines. Peripheral blood was obtained from mice between 2 and 5 months of age for all 10 founder lines (and a non-transgenic control, K). Peripheral blood cells were stained with the X-gal reagent for 2 h at room temperature, and then cytospin preparations were made and photographed. Note that the intensity of X-gal staining of red blood cells is approximately the same for X-gal$^+$ cells in all 10 transgenic founder lines, but that the percentage of positive cells varies widely from one founder line to the next. Data is shown from one representative experiment. These results were repeated at least twice with independent groups of mice, and were found to be virtually identical in every case.

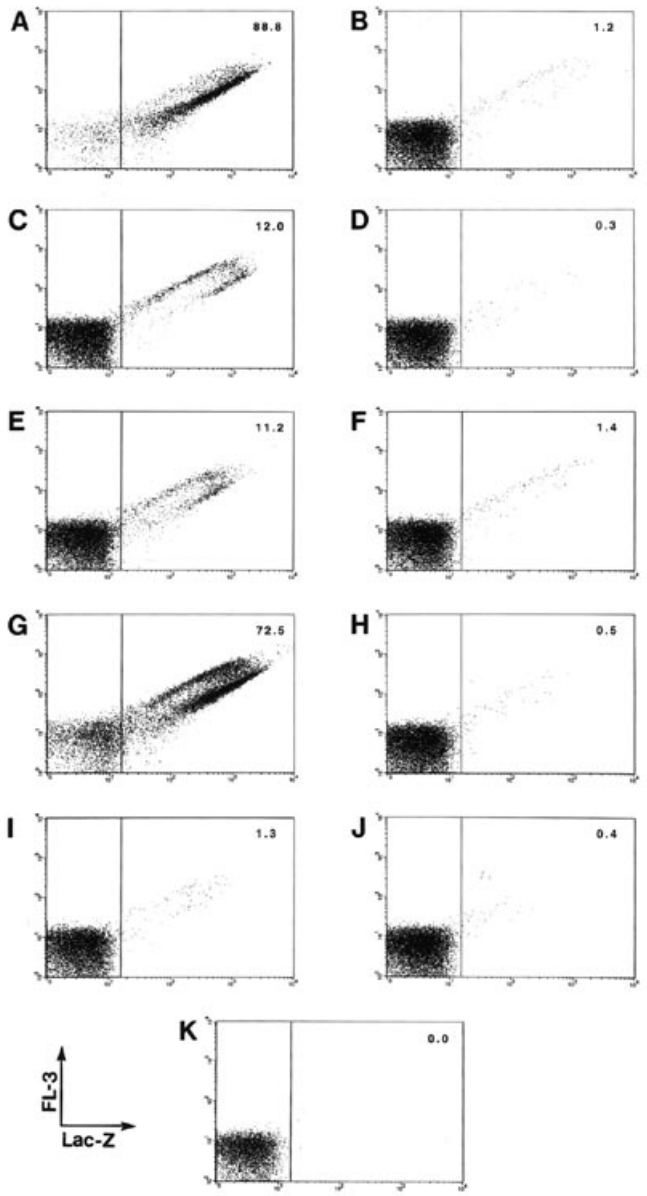

Figure 4. Flow cytometric determination of β-galactosidase expression in the red blood cells of all 10 founder lines. Peripheral blood was obtained from all 10 founder lines and a non-transgenic control (K) on the same day. The FDG reagent was introduced into red cells using osmotic shock for 30 s, and then conversion of the fluorescent substrate was allowed to proceed for 5 min at 4°C. Flow cytometric analysis was then immediately performed. An irrelevant parameter is shown on the *x*-axis, and green fluorescence (a reflection of β-galactosidase activity) on the *y*-axis. The data is plotted in two dimensions so that the Lac-Z activity in founder lines with a very small percentage of positive events can be clearly seen off the baseline. Note that the percentage of FDG$^+$ cells correlates precisely with the percentage of X-gal positive cells shown in Figure 3, and that the mean fluorescence intensity of the positive cells varies little from one founder line to another. Data is shown from one representative experiment. These results were repeated at least twice with independent groups of mice, and were found to be virtually identical in every case.

between the two studies difficult. In the prior study, age-dependent silencing was most striking in the lines utilizing an embryonic (ξ) promoter. Thus, silencing could be interpreted as an appropriate development switch. However, five of five founder lines containing a transgene with a mini LCR and a human β-globin promoter driving a Lac-Z cassette also exhibited significant transgene silencing by 8 weeks of age. Since we did not analyze mice younger than 4 weeks of age, it is possible that we failed to

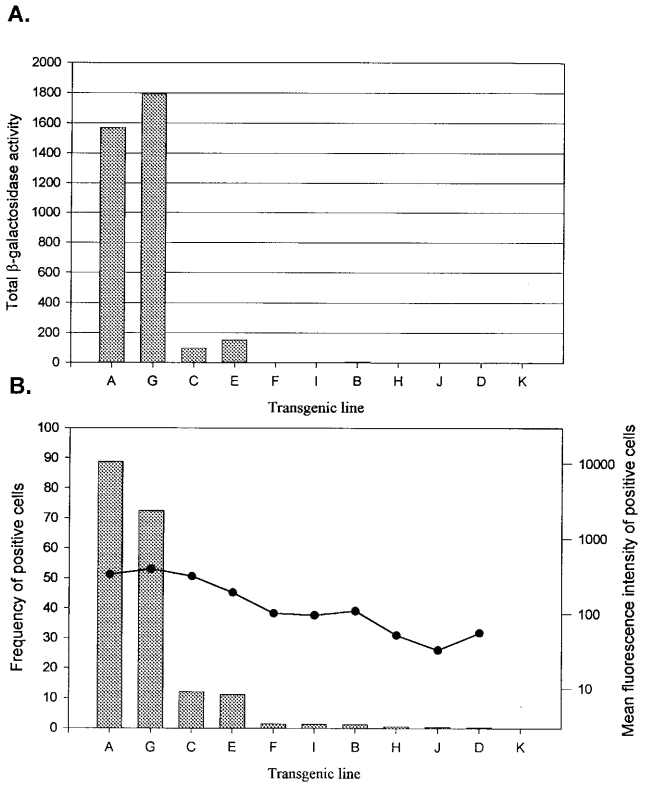

Figure 5. Graphical representation of β-galactosidase expression in the 10 founder lines. (**A**) The adjusted ONPG values in the peripheral blood for all 10 founder lines is shown. This data was extrapolated from the data presented in Table 1. (**B**) The data from the flow cytometric analysis is presented in graphical form. The data is presented from the highest to the lowest expressing line. The bar graphs represent the percentage of FDG^+ cells in each founder line, and the superimposed line represents the absolute mean fluorescence of the FDG^+ cells present within the peripheral blood of each line. Note that the total ONPG activity in the peripheral blood is highly correlated with the percentage of positive cells determined by FACS. Also, note that the average level of β-galactosidase per cell (reflected by mean fluorescence) varies only 10-fold from the lowest to the highest expressing founder line.

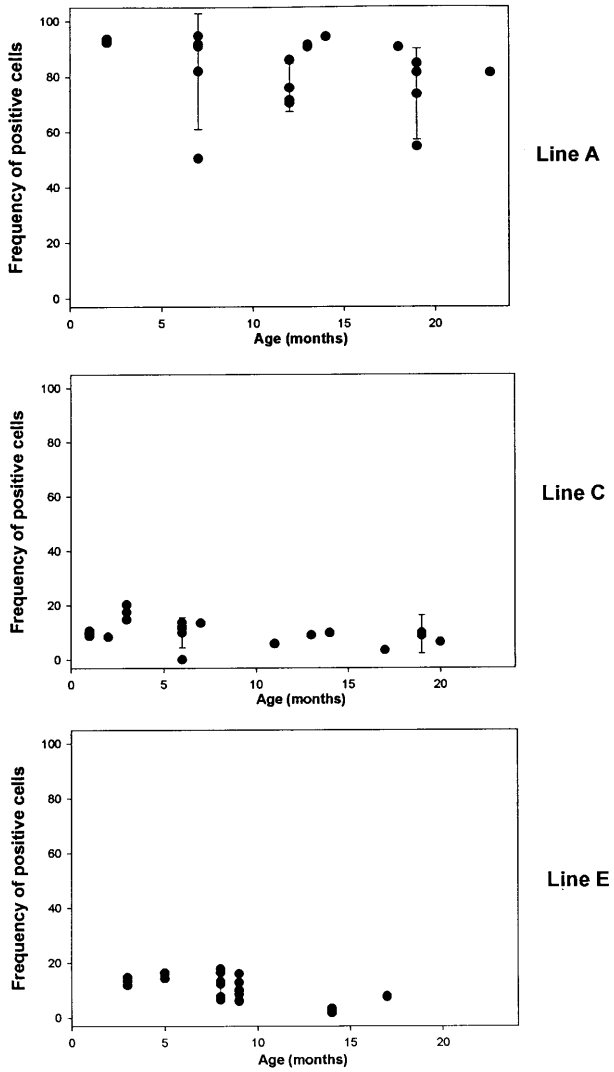

Figure 6. Minimal changes in the levels of transgene expression as a function of time. Several different F1 and F2 progeny from the transgenic founder lines had peripheral blood examined at various ages by flow cytometry and ONPG determinations on peripheral blood lysates. The percentage of FDG^+ cells in the peripheral blood of each animal is plotted as a function of time. Notice that the mice exhibit little or no change in the frequency of positive cells as a function of time.

detect early silencing in some lines. However, since lines A and G in our study express the transgene in >80% of red blood cells from the earliest time point tested, early silencing is not a universal feature of β-globin transgenes. In any case, silencing during adult life seems to us to be a more appropriate endpoint for analysis of lines expressing adult globin transgenes.

The 'decision' to variegate occurs after the BFU-E stage of erythroid differentiation

We next wished to determine whether the heterocellular pattern of transgene expression was due to the fact that different erythroid progenitors had different propensities to express the transgene, or whether each early erythroid progenitor would give rise to progeny with or without the potential to express the transgene. If the decision to variegate occurred before the BFU-E stage, then all of the cells within a given BFU-E colony would be expected to either express the transgene or not express it; the percentage of Lac-Z^+ colonies should reflect the percentage of positive red blood cells in the periphery. If the decision was made after the BFU-E stage, then all colonies should have the same pattern of transgene expression; the percentage of positive cells within an individual colony should reflect the percentage of positive cells in the peripheral blood (Fig. 7).

We obtained bone marrow from founder lines that demonstrated high, intermediate or low levels of transgene expression, and obtained peripheral blood for Lac-Z staining at the same time. The bone marrow was plated in methylcellulose cultures under conditions that favored the development of erythroid colonies. After 7 days of culture, the methylcellulose plates were treated with X-gal stain, allowed to develop overnight at 37°C, and then photographed and scored. Representative results are shown in Figure 8. No X-gal^+ cells were found in the peripheral blood or the erythroid colonies of non-transgenic mice. The frequency of BFU-E, and the appearance of BFU-E colonies, was the same for all founder lines (not shown). The percentage of X-gal positive cells within individual BFU-E reflected the percentage of positive

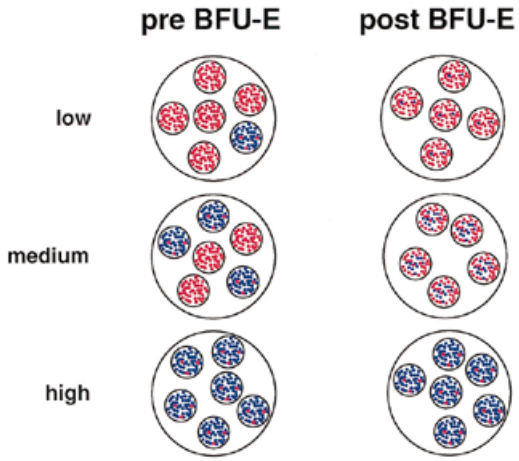

Figure 7. Potential models for transgene variegation in BFU-E colonies. Patterns of Lac-Z$^+$ cells in BFU-E colonies are presented for two different models of variegation. If the decision for the transgene to be expressed or not expressed occurs prior to the BFU-E stage of erythroid differentiation, some colonies will be negative (depicted here as red cells) and others should exhibit Lac-Z staining (depicted as blue cells) in all of the BFU-E progeny. If the decision to variegate occurs after the BFU-E stage of differentiation, then all BFU-E colonies should exhibit both positively and negatively staining red blood cells. The percentage of positive cells within the BFU-E should reflect the percentage of positive cells in the peripheral blood.

cells within the peripheral blood. Similar data were obtained for CFU-E (not shown). These data strongly suggest that integration site-dependent variegation occurs after the BFU-E stage of differentiation (Fig. 7), since virtually all BFU-E exhibited the same potential to either express or not express the transgene in their progeny.

DISCUSSION

In the present study, we have examined the mechanism of variegated transgene expression at cellular and molecular levels. Using a construct designed to direct expression of the Lac-Z reporter to the erythroid compartment of transgenic mice, we found that the different levels of transgene expression among lines was mostly related to the percentage of Lac-Z$^+$ red cells in the peripheral blood, while differences in the level of Lac-Z expression per cell accounted for a much smaller proportion of the variegation. In other words, our data suggest that a stochastic mechanism is responsible for much of this transgene's variegation, and it supports recent studies that have addressed the mechanisms involved in the production of variegation (36–38).

Several studies have shown that enhancers act to increase the probability of expression of a linked promoter, not the transcriptional activity of that promoter (21,26–28). Collectively, these data and ours argue that the probability of transgene expression largely depends upon integration site. A recent alternative model (which is at odds with our data), suggests that β-globin enhancers act to increase the level of transgene expression per cell in a graded (rather than a binary) fashion (39). These authors based their conclusions on data derived from fetal liver cells, rather than mature adult red blood cells, making comparisons of the data difficult, because of the hemoglobin switch that occurs in that compartment.

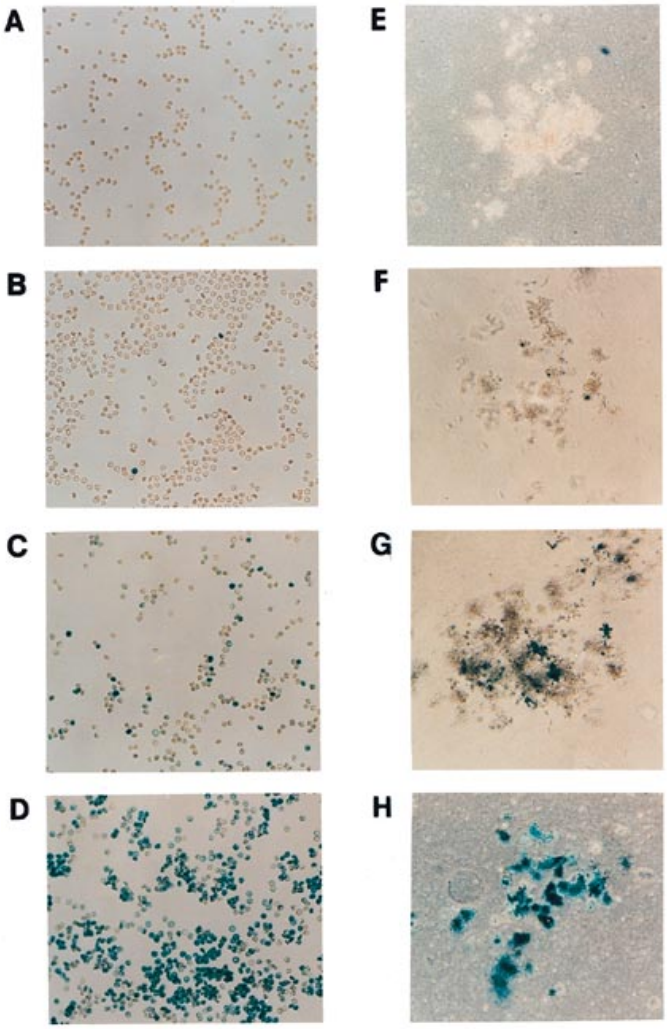

Figure 8. Lac-Z staining of BFU-E derived from a low, medium and high expressing transgenic mouse. Bone marrow cells from a mouse from a low expressing line (line B), an intermediate expressing line (line C) and a high expressing line (line A) were obtained and plated in methylcellulose under conditions that favored erythroid colony formation (see Materials and Methods). All of the colonies on the plate were overlayed with X-gal stain on day 10 after plating, and examined for Lac-Z$^+$ cells. Peripheral blood obtained from the mouse at the time of harvest was obtained, and X-gal staining was performed on it as well. The peripheral blood stains are shown in (**A**)–(**D**), and a representative BFU-E colony from each bone marrow plating experiment are shown in (**E**)–(**H**). Virtually all of the BFU-E colonies on an individual plate had the same pattern of Lac-Z expression. (A) and (E) are from a wild-type mouse; (B) and (F) are from line B (low); (C) and (G) are from line C (intermediate); and (D) and (H) are from line A (high expressing). Note that the percentage of Lac-Z$^+$ cells in the peripheral blood is closely mimicked by the percentage of Lac-Z$^+$ cells within each BFU-E, suggesting that the decision to express the transgene occurs after the BFU-E stage of differentiation.

In our model, we did not observe significant changes in the pattern of transgene expression over the life span of mice derived from several different founder lines. Within a single line, the results of ONPG, X-gal, and FACS-gal assays performed on cohorts of mice (or serially in individual mice) aged from 2 to 24 months did not differ significantly in most cases. The interpretation of this experiment might have changed if the expression of X-gal protein affected the survival of red cells. We therefore showed

that reticulocyte counts and hemoglobin levels were normal in all animals. In addition, we noted that the heterocellular pattern of Lac-Z expression remained consistent in animals after induction of hemolysis (causing an increase in reticulocytes) or irradiation (causing a decrease in reticulocytes). The β-galactosidase protein, therefore, must be long-lived within red cells. We therefore suggest that β-galactosidase expression does not perturb red cell kinetics, and that determination of Lac-Z activity in peripheral blood accurately reflects the stochastic nature of transgene expression in circulating red cells of all ages.

The site of integration of the transgene into the genome differed among all 10 lines we analyzed. In contrast to other studies (40,41), low level transgene expression was not specifically associated with centromeric chromosomal integration in our study. Centromeric integration may therefore be sufficient to cause transgene silencing, but it is not necessary, and it is not the only mechanism that can cause it.

In vitro analyses of erythroid development in our mice revealed that the decision to activate the transgene occurs after the progenitor stage (i.e. post-BFU-E, post-CFU-E). For example, within a single BFU-E colony from an intermediate expressing line, some of the cells stain positively for Lac-Z activity and some do not; all colonies have the same appearance. Thus, the heterocellular pattern of Lac-Z activity evident in peripheral blood is recapitulated in individual erythroid colonies derived from BFU-E and CFU-E of transgenic mice.

These results are reminiscent of the pattern of HbF expression by normal adult erythroid bursts. Previous studies have shown that individual BFU-E progenitors are capable of giving rise to subclones containing only HbA, or a mixture of HbA and HbF (42). These data are best explained by a stochastic model, in which a probability function governs whether the progeny of adult BFU-E express HbF. The similarity between these observations and our own suggests that similar mechanisms may be responsible for the activation of the HbF program as well as our transgene. There are, however, two important differences between the systems. First, transgene-expressing red blood cells were evenly distributed throughout the entire BFU-E colony in our experiments. In contrast, HbF expression appears to be 'all or none' within subclones of a BFU-E. Thus, HbF variegation is much 'coarser' (occurring in large clusters of cells) than the variegation we observed. The second phenotypic difference is that all of the colonies derived from an individual animal in our studies had the same pattern of Lac-Z staining, while human BFU-E derived from a single patient were quite heterogeneous with respect to HbF staining. These findings suggest that activation of this transgene appears to be a stochastic event late in erythroid development, well beyond the BFU-E and CFU-E stages. The decision to express HbF, on the other hand, probably occurs just a few cell divisions beyond the BFU-E stage and within a narrow temporal 'window'.

These findings have significance for transgenic experiments in general. Integration site-dependent variegation of transgene expression complicates the interpretation of experiments where the direct transcriptional output of different constructs is being compared. However, if the phenotype under analysis is a complex event downstream of transgene expression (e.g. development of tumors in mice with targeted expression of oncogenes), then line to line variegation may not be a critical problem. Our data suggests that low level transgene expression in an organ may represent high level expression in a small number of cells. Therefore, in the example given, all cells that express the transgene should share the same risk for development of a phenotype (i.e. neoplastic transformation), but transgenic lines would differ in the number of cells per tissue at risk for development of the phenotype.

In summary, our analysis of transgenic lines expressing an HS-2/β-globin/Lac-Z transgene supports a model in which differences in transgene expression are largely due to differences in the frequency of expressing cells, rather than differences in the level of expression per cell. Our data do not suggest that low level expression is frequently associated with centromeric chromosomal integration, or that age-dependent silencing is a universal property of globin transgenes. We have provided evidence that activation of this transgene is probably a stochastic event that occurs late in erythroid development, much like the decision to express HbF in adult red cells. The mechanisms that underlie these stochastic decisions are currently unknown, but of importance for the correct interpretation of many experiments involving transgenic mice.

ACKNOWLEDGEMENTS

This work was supported by NIH grants DK38682, DK 49786 and CA 49712 (T.J.L.). The authors thank Pam Goda for excellent care of our animals and Nancy Reidelberger for preparation of the manuscript.

REFERENCES

1 Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Costantini, F. (1985) *Nature*, **314**, 377–383.
2 Magram, J., Chada, K. and Costantini, F. (1985) *Nature*, **315**, 338–340.
3 Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1985) *EMBO J.*, **4**, 1715–1723.
4 Magram, J., Niederreither, K. and Costantini, F. (1989) *Mol. Cell. Biol.*, **9**, 4581–4584.
5 Behringer, R.R., Hammer, R.E., Brinster, R.L., Palmiter, R.D. and Townes, T.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7056–7060.
6 Kollias, G., Hurst, J., deBoer, E. and Grosveld, F. (1987) *Nucleic Acids Res.*, **15**, 5739–5747.
7 Tuan, D. and London, I.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2718–2722.
8 Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.
9 Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359–1363.
10 Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7470–7474.
11 Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) *Cell*, **51**, 975–985.
12 Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) *Genes Dev.*, **3**, 314–323.
13 Curtin, P.T., Liu, D., Liu, W., Chang, J.C. and Kan, Y.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7082–7086.
14 Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucleic Acids Res.*, **18**, 3503–3508.
15 Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2169–2178.
16 Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1626–1630.
17 Liu, D., Chang, J.C., Moi, P., Liu, W., Kan, Y.W. and Curtin, P.T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3899–3903.
18 Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. (1993) *EMBO J.*, **12**, 127–134.
19 Caterina, J.J., Ciavatta, D.J., Donze, D., Behringer, R.R. and Townes, T.M. (1994) *Nucleic Acids Res.*, **22**, 1006–1011.

20 Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5371–5375.
21 Walters, M.C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. and Martin, D.I.K. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7125–7129.
22 Walters, M.C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. and Martin, D.I.K. (1996) *Genes Dev.*, **10**, 185–195.
23 Magis, W., Fiering, S., Groudine, M. and Martin, D.I.K. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 13914–13918.
24 Garrick, D., Sutherland, H., Robertson, G. and Whitelaw, E. (1996) *Nucleic Acids Res*., **24**, 4902–4909.
25 Sutherland, H.G.E., Martin, D.I.K. and Whitelaw, E. (1997) *Mol. Cell. Biol*., **17**, 1607–1614.
26 Moon, A. and Ley, T.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 7693–7697.
27 Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) *Nucleic Acids Res*., **9**, 6047–6068.
28 Weintraub, H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5819–5823.
29 Lichter, P., Cremer, T., Borden, J., Manuelidis, L. and Ward, D.C. (1988) *Hum. Genet*., **80**, 224–234.
30 Pinkel, D., Straume, A.T. and Gray, J.W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2934–2938.
31 Schmitz, F.J. and Werner, E. (1986) *Cytometry*, **7**, 439–444.
32 Cynshi, O., Shimonaka, Y., Higuchi, M., Imai, N., Suzuki, H., Togashi, M., Okamoto, M.T. and Hirashima, K. (1990) *Br. J. Haematol*., **76**, 414–419.
33 Nolan, G.P., Fiering, S., Nocolas, J.-F. and Herzenberg, L.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2603–2607.
34 Fiering, S.N., Roederer, M., Nolan, G.P., Micklem, D.R., Parks, D.R. and Herzenberg, L.A. (1991) *Cytometry*, **12**, 291–301.
35 Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. and Whitelaw E. (1996) *Nucleic Acids Res*., **24**, 1465–1471.
36 Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser P. (1996) *Cell*, **87**, 105–114.
37 Boyes, J. and Felsenfeld, G. (1996) *EMBO J*., **15**, 2496–2507.
38 McDevitt, M.A., Fujiwara, Y., Shivdasani R.A. and Orkin S.H. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 7976–7981.
39 Guy, L.-G., Kothary, R. and Wall, L. (1997) *Nucleic Acids Res*., **25**, 4400–4407.
40 Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. and Kioussis, D. (1996) *Science*, **271**, 1123–1125.
41 Dobie, K.W., Lee, M., Fantes, J.A., Graham, E., Clark, A.J., Springbett, A., Lathe, R. and McClenaghan, M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6659–6664.
42 Stamatoyannopoulos, G., Kurnit, D.M. and Papayannopoulou, T. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7005–7009.
43 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 16.66–16.67.