MTOpen

Kinetics and Epigenetics of Retroviral Silencing in Mouse Embryonic Stem Cells Defined by Deletion of the D4Z4 Element

Sylvie Rival-Gervier[1,2], Mandy YM Lo[1,3], Shahryar Khattak[1], Peter Pasceri[1], Matthew C Lorincz[4] and James Ellis[1,3]

[1]Developmental and Stem Cell Biology, Hospital for Sick Children, Toronto, Ontario, Canada; [2]INRA, UMR 1198 Biologie du Développement et Reproduction, Jouy en Josas, France; [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; [4]Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Retroviral vectors are silenced in embryonic stem (ES) cells by epigenetic mechanisms whose kinetics are poorly understood. We show here that a 3′D4Z4 insulator directs retroviral expression with persistent but variable expression for up to 5 months. Combining an internal 3′D4Z4 with HS4 insulators in the long terminal repeats (LTRs) shows that these elements cooperate, and defines the first retroviral vector that fully escapes long-term silencing. Using FLP recombinase to induce deletion of 3′D4Z4 from the provirus in ES cell clones, we established retroviral silencing at many but not all integration sites. This finding shows that 3′D4Z4 does not target retrovirus integration into favorable epigenomic domains but rather protects the transgene from silencing. Chromatin analyses demonstrate that 3′D4Z4 blocks the spread of heterochromatin marks including DNA methylation and repressive histone modifications such as H3K9 methylation. In addition, our deletion system reveals three distinct kinetic classes of silencing (rapid, gradual or not silenced), in which multiple epigenetic pathways participate in silencing at different integration sites. We conclude that vectors with both 3′D4Z4 and HS4 insulator elements fully block silencing, and may have unprecedented utility for gene transfer applications that require long-term gene expression in pluripotent stem (PS) cells.

Received 20 December 2012; accepted 21 May 2013; advance online publication 11 June 2013. doi:10.1038/mt.2013.131

INTRODUCTION

Retroviral vectors are transcriptionally silenced in pluripotent stem (PS) cells. This feature facilitated the discovery of induced PS (iPS) cells because delivery of exogenous pluripotency factors in retroviral vectors allowed the transgenes to be silenced as the somatic cells reprogrammed. In virtually all other contexts, silencing of retroviral vectors is considered deleterious for their use in stem cells. To overcome silencing, retroviral vector designs mutate or delete known silencer elements in or adjacent to the long terminal repeats (LTRs).[1,2] However, even self-inactivating (SIN) retroviral vectors with a strong internal promoter are subject to silencing in embryonic stem (ES) cells.[3] Thus, further improvements rely on removing additional unknown silencer elements, or on better defining the mechanisms of silencing and discovering how they can be blocked.

Retrovirus silencing occurs *via* epigenetic mechanisms in ES cells. For example, some retroviral sequences recruit the ZFP809 DNA-binding factor which interacts with repressive complexes including Kap1 (Trim28), the histone methyltransferase ESET (SETDB1), heterochromatin protein 1 (HP1), the nucleosome remodeling and histone deacetylase (NuRD) complex, and the nuclear receptor corepressor complex 1 (N-coR1).[4–8] The binding of this complex results in the deposition of H3K9me3 marks on the sequences nearby and in transcriptional repression. Moreover, DNA methylation is targeted by the *de novo* methyltransferases Dnmt3a and 3b[9] to CpG-rich sequences in LTRs,[10] enhanced green fluorescent protein (EGFP) or other non-mammalian reporter genes.[11] This hypermethylated DNA is bound by MeCP2[12] and recruits histone deacetylases.[13] However, deacetylated histone H3 and H1 can still be associated with silent retrovirus in Dnmt3a and 3b null ES cells,[14] and H3K9me3 marks established by SetDB1 in ES cells are also independent of DNA methylation.[15] The enzymes G9a/GLP write H3K9me2 marks but also promote DNA methylation of LTR elements and other genomic regions independently from their histone methyltransferase activity.[16,17] Inhibiting G9a/GLP activity with a chemical probe (UNC0638) can reactivate the silent internal promoter of a retroviral vector and trigger DNA demethylation.[18] Unfortunately, the kinetics of these epigenetic events is poorly understood, and would be more easily defined by developing a system to synchronize retrovirus silencing in ES cells.

After integration of a SIN retroviral vector, the internal promoter can be subject to residual silencing emanating from cryptic silencer elements in the vector backbone or non-mammalian CpG-rich reporter gene sequences. In addition, despite the propensity of murine leukemia virus-based retroviral vectors

Correspondence: *James Ellis, Developmental and Stem Cell Biology, Hospital for Sick Children, Rm13-310 TMDT, MaRS Centre, 101 College St, Toronto, Ontario M5G 1L7, Canada. E-mail: jellis@sickkids.ca*

to integrate near the promoters of active genes,[19,20] the chromatin environment surrounding the provirus may exert position effects that repress expression.[21,22] Insulator elements can protect transgenes from position effects by: (i) blocking enhancer–promoter communication when positioned between them and (ii) acting as a barrier to prevent transgene silencing by blocking the spread of heterochromatin.[23] The chicken β-globin HS4 insulator is the most characterized insulator in vertebrates.[24] Its 250-bp core has two physically separable and mechanistically distinct insulator properties. Enhancer blocking is mediated by CTCF which is necessary to establish chromosomal loop domains.[25,26] Barrier activity is attributed to VEZF1, which limits DNA methylation and USF1/USF2 which recruits enzymes that write activating modifications on histones.[22,27,28] The HS4 core or larger fragments protect transgenes against silencing,[29,30] but this barrier activity is suboptimal in some contexts including retroviral transduction of ES cells.[31–34]

The D4Z4 element is a 3.3 kb macrosatellite sequence present in 11–150 copies in the subtelomeric region of human chromosome 4. Reduction of D4Z4 copy number is associated with human facioscapulohumeral dystrophy.[35] Expression of an open-reading frame (ORF) called DUX4 from the last D4Z4 repeat in patients is associated with the disease phenotype.[36] A single D4Z4 monomer element can act as an insulator[37] with enhancer blocking activity in its 5′ region that is dependent on CTCF binding. Unusually, D4Z4 can block position effects when inserted on just one side of the transgene, and therefore is not a classical barrier element like HS4 which must flank the transgene.[37] These findings indicate that D4Z4 functions differently from HS4, and thus D4Z4 may protect transgenes from heterochromatin effects that are not shielded by HS4. Here, we investigate the utility of the D4Z4 insulator in preventing retrovirus silencing in ES cells. Our findings demonstrate that 3′D4Z4 cooperates with HS4 barrier activity to drive unprecedented long-term retrovirus expression in ES cells. *In situ* deletion of 3′D4Z4 induces silencing, providing a system to identify kinetic classes of silencing and the epigenetic marks associated with them.

RESULTS

HSC1 retroviral vector is silenced over time even in the presence of the HS4 core

To characterize the kinetics of retrovirus silencing, we analyzed two different silencing events that occur at different timepoints. First, we examined the frequency of silencing that occurs rapidly after proviral integration and without any cell selection. Second, we tracked the kinetics of silencing that extinguishes transgene expression over prolonged periods of cell culture in cells that initially express the virus, isolated either by sorting for EGFP-positive cells or by drug selection with puromycin. For this purpose, we modified the HSC1 SIN retroviral vector to incorporate an internal human *EF1α* long promoter driving the expression of a bicistronic EGFP ires-puromycin (GiP) cassette (**Figure 1a**).

To quantify rapid silencing after infection, we isolated single cell ES clones in the absence of selection. Briefly, we infected ES cells at low multiplicity of infection (**Figure 1b**) and 1 day after infection performed a limit-dilution assay to deposit only one cell per well. These cells were expanded without selection to identify transduced cell clones in which the EGFP transgene is silenced. The remainder of the bulk population was maintained without selection and at 6 days post-infection 2–9% of cells were EGFP-positive after infection with 5 or 25 µl of virus, respectively (**Figure 1c**). To isolate clonal lines, wells with more than one colony were eliminated. About 260–280 clones from two independent transductions were replated on day 6 and assayed for EGFP expression by microscopy on day 8. To gather sufficient genomic DNA (gDNA) for Southern analysis, the clones were cultured until day 17. gDNA used for replicate dot blots was hybridized with an EGFP probe to detect provirus and a mouse endogenous *Thy1* probe as a loading control. These data show that about 10–15% of cells contain the proviral DNA, consistent with single copy integration in the infected clones. We found that 76–84% of provirus positive clones express EGFP at day 8 post-infection (**Figure 1d**). These results show that about 20% of integrated HSC1-GiP proviruses are rapidly silenced in the absence of selection.

To examine long-term silencing of the HSC1 retroviral vector, ES cells were infected with HSC1-GiP at low multiplicity of infection to produce cells with single copy integrations. Two days after infection, 10–20% of cells were EGFP-positive consistent with single copy transduction. These expressing cells were immediately selected with 1 µg/ml of puromycin for 4 days and then cell populations were maintained with or without selection for up to 150 days (**Figure 2a**). All viable cells were EGFP-positive after 4 days of drug selection, and as expected EGFP expression persisted over time when cells were cultured with puromycin (**Figure 2b**). However, upon puromycin withdrawal, frequency of retroviral transgene expression decreased over time. At 60 days post-infection, 80% of cells expressed EGFP and by 150 days post-infection only 40% of the cells are EGFP-positive (**Figure 2b**). This retrovirus silencing is detectable in ES cells, as well as in iPS cells (**Supplementary Figure S1a–c**), but not in NIH3T3 fibroblasts (**Supplementary Figure S1d**). To ensure that the HSC1-GiP population is still undifferentiated, we assessed the pluripotency level of cells by analyzing the presence of SSEA1 protein, an ES cell surface marker, by flow cytometry and the presence of the Nanog ES cell nuclear marker by immunocytochemistry. We show that 99% of the cells are SSEA1-positive (**Supplementary Figure S2a**) and Nanog protein is detectable in most nuclei (**Supplementary Figure S2b**). These results show that silencing and variability of EGFP expression is not a consequence of ES cell differentiation. Overall, these results show that the HSC1-GiP retroviral vector is frequently silenced in undifferentiated PS cells over long-term culture, even when integration sites favorable for transgene expression are preselected shortly after infection.

To attempt to overcome this silencing, we used the well-characterized dimer HS4 core insulator present in the HSC1-GiP LTRs (**Figure 2c**). Under puromycin selection, all cells were EGFP-positive (**Figure 2d**) but without drug selection, at 60 and 150 days post-infection, 90 and 70% of cells expressed EGFP, respectively. Thus, the dimer HS4 core located in the LTRs improves the frequency of transgene expression in ES cells but a significant percentage of cells still show silencing.

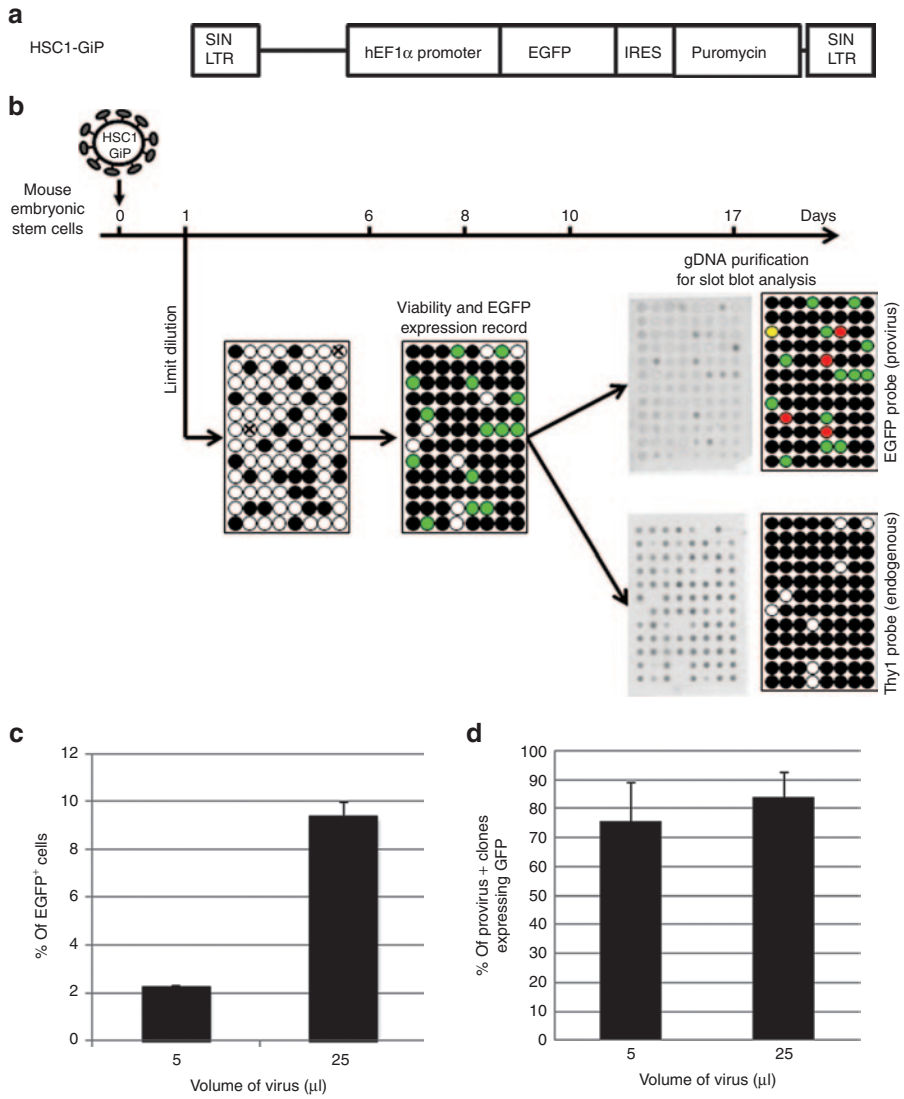

Figure 1 HSC1 retroviral vector is rapidly silenced after infection. (**a**) Structure of HSC1-GiP provirus. The internal human *EF1*α promoter drives expression of the bicistronic gene encoding enhanced green fluorescent protein (EGFP) and puromycin resistance. (**b**) Experimental system. At day 6, wells with more than one colony (circles with a multiplication symbol) were eliminated and wells containing only one colony (closed circles) were passed into duplicate 96-well plates. At day 8, empty wells (open circles), wells with non-expressing cells (closed circles), and wells with expressing cells (green circles) were identified. Genomic DNAs (gDNAs) were harvested at day 17 from both plates and used for slot blot analysis. Membranes were hybridized with an EGFP probe to detect provirus or *mThy1* probe to detect endogenous DNA. Interpretation of expression and genotype data are shown on the right. Green circles represent wells with expressing cells containing EGFP proviral DNA; red circles are non-expressing cells containing the proviral vector; yellow circle is a colony detected as expressing where no provirus was detected. (**c**) EGFP expression in populations of non-selected cells detected by flow cytometry. This graph displays the percentage of EGFP-positive cells at day 6 in the population obtained after infection with 5 or 25 µl of virus (264 and 288 clones picked, respectively; producing 7 and 27 EGFP-expressing clones). (**d**) Percentage of expressing cells among infected cells. The percentage of expressing clones represents the number of EGFP-expressing clones among the infected (proviral DNA detected) and viable clones. This result is the average of two independent experiments with two independent viral productions. Total numbers of dot blot positive clones are 10 and 33, respectively; for the 5 and 25 µl virus infections. Error bars represent the SD. IRES, internal ribosome entry site; LTR, long terminal repeat; SIN, self-inactivating.

3′D4Z4 insulator increases the frequency of long-term transgene expression

Given that the HSC1 retroviral vector is strongly silenced in ES cells, we tested the novel strong D4Z4 element, first on its own at an internal location and subsequently combined with the dimer HS4 core in the LTRs. We cloned the full-length D4Z4 monomer upstream of the *EF1*α promoter (**Figure 3a**). Unfortunately, no reporter expression was detected after infection, indicating that the full D4Z4 sequence is incompatible with retroviral vector transduction. As only a monomer of D4Z4 is present and there are no endogenous D4Z4 elements in mouse cells, the inability to detect transduced virus is not due to any repetitive sequence of D4Z4. To identify D4Z4 sequences that can transmit at high titer, we subdivided this element into four DNA fragments derived from 5′D4Z4 or 3′D4Z4 regions and cloned them upstream of the *EF1*α promoter (**Figure 3b**). The 5′ fragment called D4Z4-A contains a putative enhancer that was reported to activate an SV40 viral promoter[38] but not a mammalian promoter,[37] an enhancer blocking

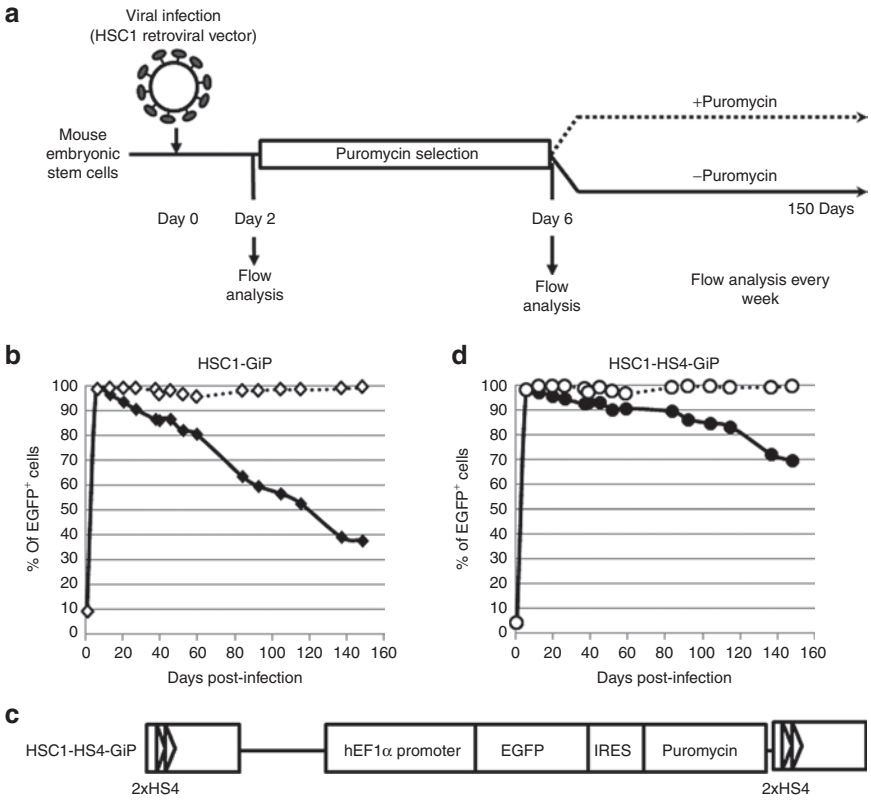

Figure 2 HSC1 retroviral vectors with or without HS4 are silenced over time. (**a**) Experimental system. J1 embryonic stem cells were infected at low multiplicity of infection by HSC1 retroviral vectors. Two days after infection, expressing cells were selected with puromycin for 4 days and maintained as a population in the presence (dashed line) or absence (solid line) of puromycin for up to 5 months. (**b**,**d**) HSC1-GiP and HSC1-HS4-GiP retroviral vector expression over time. Populations were maintained without (black marker, solid line) or with (open marker, dashed line) selection over time. Graphs represent the percent of enhanced green fluorescent protein (EGFP)-positive cells over time in the HSC1-GiP (open and closed diamonds in **b**) and HSC1-HS4-GiP (open and closed circles in **d**) populations. (**c**) Structure of HSC1-HS4-GiP provirus. Two copies of the core HS4 are in the 3′ LTR of the HSC1-GiP retroviral vector. After viral replication, this insulator is present in both LTRs. IRES, internal ribosome entry site; LTR, long terminal repeat.

region with a known CTCF-binding site and a barrier activity.[37,38] The D4Z4-B fragment contains the promoter and 5′ end of the *DUX4* ORF, but because this would introduce a second internal promoter, we excluded this vector from further analysis. To evaluate the 3′ end of D4Z4, we generated two fragments. D4Z4-C contains only 36 bp of the *DUX4* promoter and lacks its TATA box and the 3′ end of the *DUX4* ORF.[39–41] To independently confirm that any activity that blocks retroviral silencing found in 3′D4Z4 does not require *DUX4* expression, we also used D4Z4-D that does not include the promoter nor the 5′ end of the *DUX4* ORF. To investigate the possibility of spacing effects, we inserted a lambda phage spacer at the same position as a control. None of these fragments have enhancer activity on the *EF1α* promoter in transient transfection assays in ES cells (**Supplementary Figure S3**). Therefore, any influence they have on retroviral expression after vector integration likely involves epigenetic effects on chromatin.

Viruses with the spacer fragment and the D4Z4-A and D4Z4-C fragments were successfully produced and ES cells were infected at low multiplicity of infection resulting in 2–8% of cells expressing EGFP 3 days post-infection. In this experiment, instead of using puromycin selection, we enriched for the expressing cell population by sorting for EGFP-positive cells 3 days post-infection. Cells were cultured for 2 months and EGFP expression assayed over time by flow cytometry. As expected, the vector with the spacer element was already subject to silencing at 64 days post-infection, with 70% EGFP-positive cells (**Figure 3c**). This result shows that a spacer element does not prevent silencing. Surprisingly, the D4Z4-A virus infection also displays a prominent peak of non-expressing cells by 64 days post-infection with only 54% of cells still EGFP-positive. We show that transgene DNA in the D4Z4-A population is only detected by Southern blot and quantitative PCR (qPCR) in EGFP-positive and not EGFP-negative sorted cells (**Supplementary Figure S4**). These results show that the D4Z4-A construct is unstable in this retroviral vector. In contrast, 95% of cells are EGFP-positive in the population containing the D4Z4-C fragment (**Figure 3c**) and the transgene is easily detected by Southern blot and qPCR (**Supplementary Figure S4**). Therefore, 3′D4Z4 transmits through retroviral vectors, is stable over extended passage time and ensures a very high frequency of retrovirus expression. These 3′D4Z4 cells display a mean fluorescence intensity (MFI) of 311 (compared with 250 in EGFP-positive cells with the spacer element) but exhibit some cell-to-cell variability of expression (coefficient of variation (CV) = 65) (**Figure 3d**). To design an optimized retroviral vector for PS cells, it may be important to combine 3′D4Z4 with additional insulator activities that are compatible with transduction.

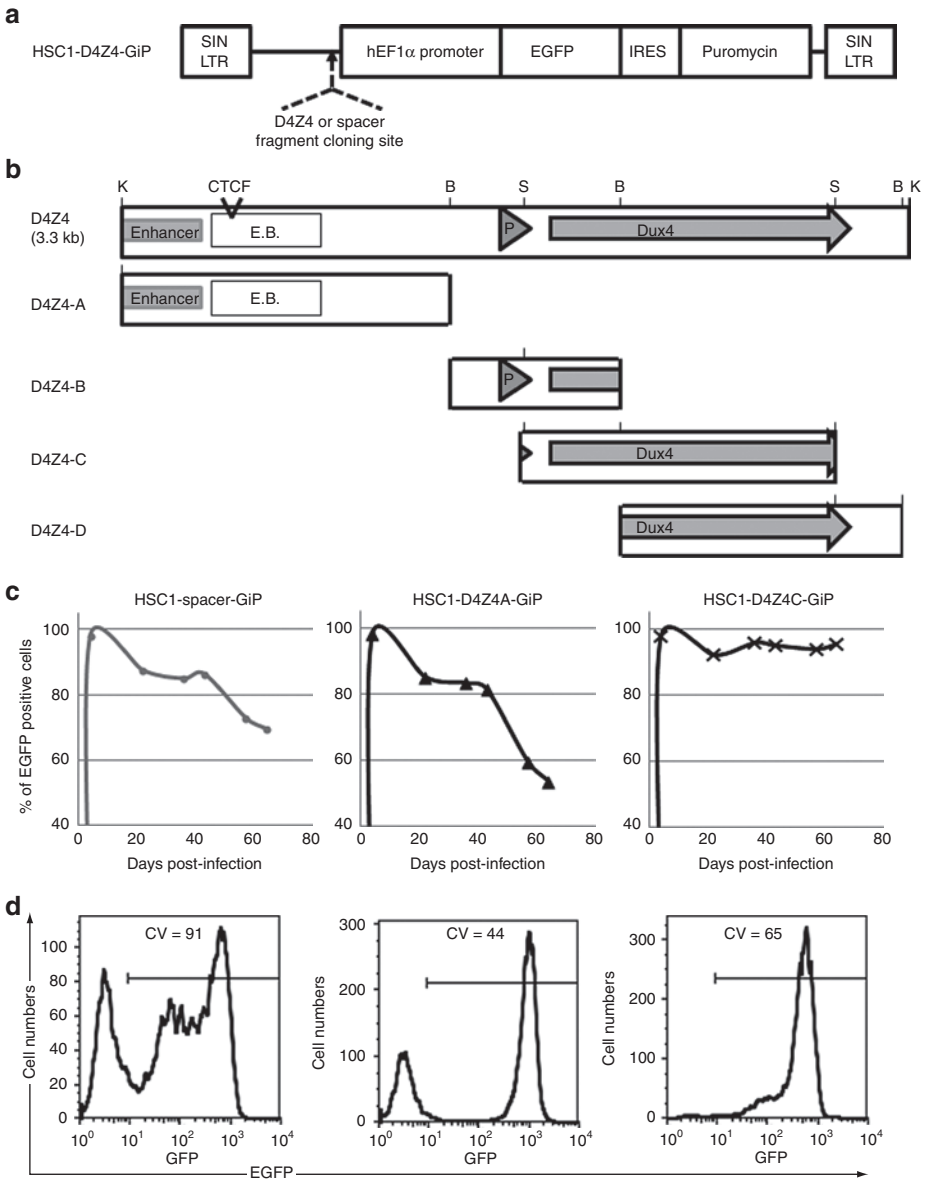

Figure 3 3′D4Z4 protects HSC1 retroviral vector from silencing. (**a**) Structure of HSC1-D4Z4-GiP provirus. D4Z4 elements or a spacer DNA were cloned upstream of the *EF1α* promoter in the HSC1-GiP retroviral vector. (**b**) D4Z4 fragments cloned in HSC1 retroviral vector. The D4Z4 element is 3.3 kb long. It contains an enhancer, an enhancer blocker, and part of the *DUX4* open-reading frame (ORF). D4Z4-A is the *Kpn*I-*Bam*HI 5′D4Z4 fragment (1,380 bp); D4Z4-C is the *Stu*I 3′D4Z4 fragment (1,310 bp); and D4Z4-D corresponds to the *Bam*HI 3′D4Z4 fragment (1,172 bp). K, *Kpn*I; B, *Bam*HI; S, *Stu*I; P, *Dux4* promoter; E.B., enhancer blocker with the CTCF-binding site; *Dux4* gray arrow, Dux4 partial ORF. (**c**) Enhanced green fluorescent protein (EGFP) expression from provirus containing the spacer element, D4Z4-A and D4Z4-C in embryonic stem (ES) cells. Expressing ES cells were sorted and cultured without further selection. (**d**) Flow cytometry plots 64 days post-infection. *X*-axis represents the level of EGFP fluorescence, *Y*-axis represents the cell numbers. CV, coefficient of variation; IRES, internal ribosome entry site; LTR, long terminal repeat; SIN, self-inactivating.

3′D4Z4 and HS4 elements cooperate to optimize long-term expression

While 3′D4Z4 can block silencing when located upstream of the *EF1α* promoter, it may be insufficient to fully protect the adjacent promoter from 3′ position effects that reduce the MFI. Because HS4 appeared to have partial barrier activity when located in the LTRs, we tested whether 3′D4Z4 can cooperate with HS4 in the LTRs to minimize transgene silencing in ES cells. We inserted 3′D4Z4 (fragments C or D) upstream of the *EF1α* promoter in the HSC1-HS4-GiP virus that contains the HS4 core dimer in the LTRs (**Figure 4a**). We infected ES cells and briefly selected for expressing cells with puromycin before culture with or without selection for up to 5 months. After long-term culture, none of these populations shows any evidence of differentiation (**Supplementary Figure S2**). Strikingly, the expression frequency approaches 95–100% EGFP-positive cells without selection over 150 days after infection with D4Z4-C or D4Z4-D vectors (**Figure 4b**), indicating that vectors containing both HS4 and 3′D4Z4 are stable over extended passages. These data confirm that the 3′D4Z4 activity is functional in the presence of HS4 and does

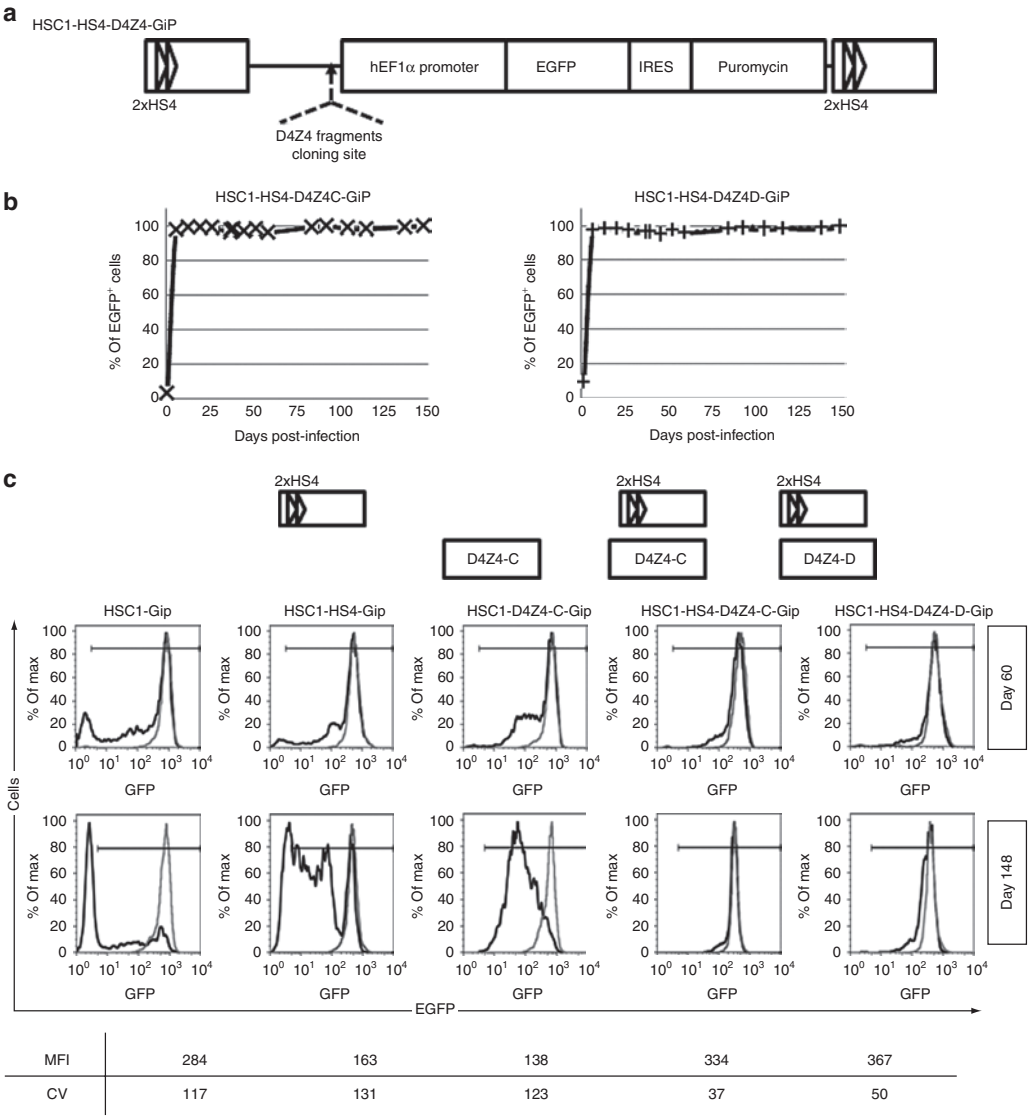

Figure 4 3′D4Z4 and HS4 insulators cooperate to improve frequency, level, and reproducibility of enhanced green fluorescent protein (EGFP) expression. (**a**) Structure of HSC1-HS4-D4Z4-GiP proviruses. 3′D4Z4 fragments (D4Z4-C or D4Z4-D) were cloned upstream of the *EF1α* promoter in the HSC1-HS4-GiP retroviral vector. (**b**) HSC1-HS4-D4Z4-C-GiP and HSC1-HS4-D4Z4-D-GiP retroviral vector expression over time. J1 embryonic stem (ES) cells were infected and selected with puromycin. Cell populations were maintained without (black marker, solid line) or with (open marker, dashed line) further selection. (**c**) EGFP expression at 60 (top panel graphs) and 148 days (bottom panel graphs) post-infection in all populations. Flow plots in populations cultured in presence (gray line) or absence (black line) of selection. *X*-axis shows the level of EGFP fluorescence, *Y*-axis shows the percentage of maximum cell numbers. Levels (the mean fluorescence intensity (MFI)) and variability (the coefficient of variation (CV)) of EGFP expression at 148 days are displayed below. IRES, internal ribosome entry site.

not require expression of the *DUX4* ORF. The D4Z4-C vector also maintains the frequency of EGFP expression after differentiation into embryoid bodies *in vitro* (**Supplementary Figure S5**).

To determine whether the HS4 element improves the level (MFI) and consistency (CV) of expression in cooperation with 3′D4Z4, we analyzed the flow cytometry data at day 60 (for comparison to the results shown after EGFP sorting in **Figure 3**) and day 148 post-infection (**Figure 4c**). In cells grown without selection, the vector without any insulators has a low frequency of expression with increasing numbers of EGFP-negative cells over time, and the cells that do express have a high CV at the day 148 timepoint (**Figure 4c** and **Supplementary Figure S6**). Addition of HS4 improves the frequency of expression, but the MFI is reduced and the CV is almost unchanged, suggesting that the barrier activity of HS4 may protect the transgene from surrounding genomic sequences but may not block silencing directly recruited by the provirus itself. When D4Z4-C is present without HS4, the frequency of expression is very high at both timepoints, but the MFI is reduced and the CV remains at the same level consistent with a residual position effect. The

shoulder of low-expressing D4Z4-C cells at day 60 is very similar to the behavior of these cells after EGFP-positive cell sorting and growth for 60 days as shown in **Figure 3d**, indicating that the two selection protocols generate similar results. Only in the presence of D4Z4-C and HS4 does the frequency stay high and the CV is very low producing a very sharp peak with an MFI of around 300 at both timepoints. This finding is corroborated by virtually identical results from combining D4Z4-D with HS4. Overall, these data demonstrate that 3′D4Z4 cooperates with HS4 barrier activity to promote persistent transgene expression with unprecedented consistency in ES cells.

3′D4Z4 and HS4 cooperate to open chromatin structure on the provirus

To determine whether the 3′D4Z4 insulator protects retroviral transgene expression from silencing through formation of an open chromatin structure, we analyzed the DNA methylation and histone modification profiles on the provirus. We harvested gDNA from transduced ES cells cultured for more than 5 months without selection for bisulfite sequencing of the *GAG*, *EF1α* promoter, and *EGFP* regions upstream and downstream of 3′D4Z4. In all but one population, the three regions have a similar level of CpG methylation (**Figure 5a**). In the population without

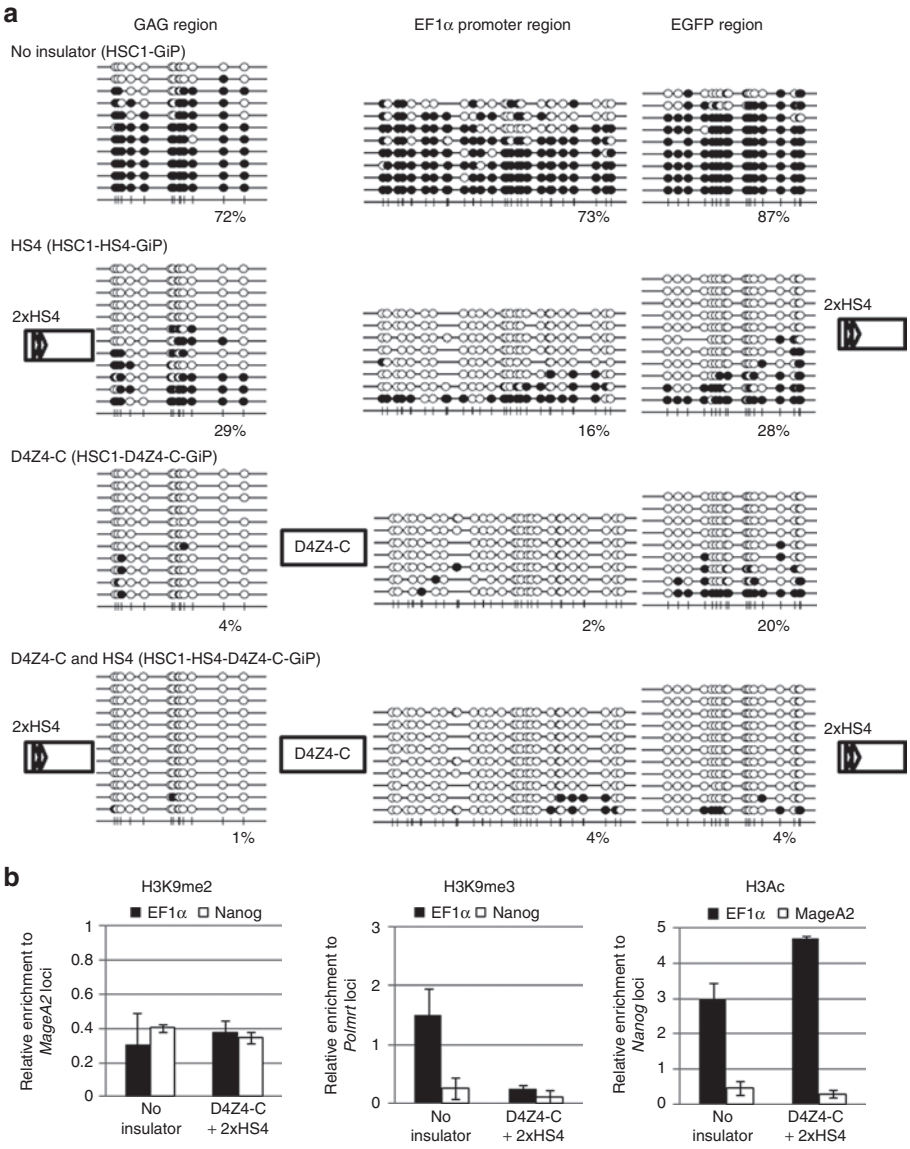

Figure 5 Epigenetic profiles of silent and expressing proviruses. (**a**) Bisulfite sequencing of proviral DNA at *GAG*, *EF1α* promoter, and *EGFP* regions from HSC1-GiP, HSC1-HS4-GiP, HSC1-D4Z4-C-GiP, and HSC1-HS4-D4Z4-C-GiP proviruses 148 days post-infection. Open and closed circles indicate unmethylated and methylated CpG sites, respectively. Percentages below each methylation profile represent the percentage of total CpG methylation. (**b**) Analysis of repressive and active chromatin marks on the *EF1α* promoter by ChIP. ChIP experiments were performed with anti-H3K9me2, H3K9me3, and H3Ac antibodies. DNA isolated from immunoprecipitated chromatin was subjected to quantitative PCR to amplify the *EF1α* promoter from the retroviral vector and endogenous positive controls (*MageA2* for H3K9me2, *Polmrt* for H3K9me3, and *Nanog* for H3Ac). Relative enrichments correspond to the enrichment of the studied chromatin mark on the *EF1α* promoter normalized to endogenous positive control (closed bars). Endogenous negative controls (open bars, *Nanog* for H3K9me2 and H3K9me3, and *MageA2* for H3ac) relative to the enrichment on the corresponding endogenous positive control are shown below. EGFP, enhanced green fluorescent protein.

insulators that have only 35% EGFP-positive cells, all regions are hypermethylated (72–87% CpG methylation). In contrast, the proviruses containing HS4 in both LTRs are partially methylated (16–29% CpG methylation), whereas the proviruses containing D4Z4-C insulator display 4 and 2% CpG methylation in *GAG* and *EF1α* promoter regions, respectively. Partial methylation of 20% in *EGFP* may predispose the cells to the variation in expression observed over time. Finally, in the presence of 3′D4Z4 and HS4, the proviral DNA is hypomethylated (1–4% CpG methylation). Thus, the methylation pattern of proviral DNA is in complete concordance with the expression profile of each population. These findings suggest that 3′D4Z4 and HS4 alone are partially effective at blocking DNA methylation, but when present together cooperate to maintain hypomethylated retroviral transgenes that are persistently expressed.

To analyze whether the presence of both insulators alters the histone modification pattern on the provirus compared with the same retroviral vector without insulators, we performed ChIP-qPCR on the *EF1α* promoter after 150 days of culture without puromycin. H3K9 dimethylation (H3K9me2) and trimethylation (H3K9me3) modifications were examined as repressive chromatin marks, and H3 pan-acetylation (H3Ac) as an active chromatin mark. It has been shown that H3K9me2 and H3K9me3 correlate with silencing of some retroviral vectors.[7,15] With the HSC1-GiP retroviral vector, we found comparable enrichment of H3K9me2 on the *EF1α* promoter and the expressed endogenous *Nanog* region with no difference between the two populations (**Figure 5b**). This result reveals that H3K9me2 does not correlate with silencing of a SIN retroviral vector. However, we cannot exclude that it promotes DNA methylation in some silent genomic contexts but not in more active contexts, or that it plays a role in a subset of viral integrations that are not detected in the bulk ChIP data on cell populations. On the other hand, we found a sixfold enrichment in H3K9me3 marks on provirus without insulators compared with provirus containing both insulators. The presence of the H3K9me3 mark may reflect spreading of repressive chromatin over the provirus, which is blocked by cooperating 3′D4Z4 and HS4 elements. Finally, we detected a marginally greater enrichment in H3Ac histone marks on provirus containing both insulators. Overall, these data suggest that 3′D4Z4 and HS4 insulators together reduce H3K9me3 repressive marks to the same level as the expressed endogenous *Nanog* locus and support active chromatin marks on the provirus to promote high frequency long-term transgene expression.

Deletion of 3′D4Z4 induces retroviral vector silencing

3′D4Z4 may promote stable transgene expression by directing retroviral vector integration away from hostile genomic environments that would otherwise impose a silent chromatin state on neighboring proviral integrants. If redirection to more favorable sites were the case, then deletion of 3′D4Z4 after integration would presumably have no effect on EGFP expression. Furthermore, if 3′D4Z4 acts as an enhancer, its deletion in the context of the HS4 construct would be expected to reduce the expression level at all integration sites but to not cause transgene silencing. In contrast, if 3′D4Z4 activity is responsible for preventing heterochromatin spread at hostile integration sites, then deletion of 3′D4Z4 after integration would induce silencing.

To distinguish these models of 3′D4Z4 function, we cloned the D4Z4-C fragment flanked by two FRT sites upstream of the *EF1α* promoter in vectors with or without HS4 in the LTR (**Figure 6a**). First for a proof-of-concept, we infected ES cells and sorted EGFP-positive cell populations (**Figure 6b**), before moving to individual cell clones in subsequent experiments described below. After expansion, FLP recombinase-mCherry was transiently expressed and mCherry-positive cells were sorted 2 days later to enrich the population of cells which expressed FLP recombinase and flipped out D4Z4-C (**Figure 6b**). We maintained transfected and untransfected cells in parallel. These populations have proviruses at the same integration sites that differ solely by the presence or absence of the D4Z4 element. Southern blot and qPCR analyses show that 3′D4Z4 was deleted in ~80% of cells (**Figure 6c**,**d**). In both populations with and without the HS4 core insulator in the LTRs, EGFP expression is silenced when D4Z4-C is deleted as shown in **Figure 6e**. As expected, expression frequency is enhanced when HS4 remains in the LTRs (83 versus 67% of EGFP-positive cells). It is not possible to determine the role of 3′D4Z4 at each integration site using these cell populations. However, these results agree with the earlier finding that 3′D4Z4 is not an enhancer of the *EF1α* promoter in the vector. These data also demonstrate that 3′D4Z4 does not direct integration into favorable sites, but are consistent with a model in which the transgene is silenced at a subset of integration sites when 3′D4Z4 is deleted.

Deletion of 3′D4Z4 in cell clones reveals rapid and gradual kinetics of silencing

To identify the kinetics of retrovirus silencing, we generated single cell clones where proviruses with and without 3′D4Z4 are integrated at the same integration site. As outlined in **Figure 6b**, we infected ES cells and 24 hours after infection sorted single EGFP-positive cells into 96-well plates. This single cell sorting method was chosen as it is more efficient at generating unique cell clones than puromycin selection. Eleven clones with D4Z4-C and 10 clones with D4Z4-C and HS4 in the LTRs were expanded. All clones expressed the transgene consistently when grown without puromycin. To delete 3′D4Z4 in each clone, we transfected FLP recombinase-mCherry plasmid and mCherry-positive cells were sorted to increase the probability of D4Z4-C removal. The 3′D4Z4 deletion frequency ranged from 70 to 98% (**Supplementary Figure S7**). We verified by Southern blot that 19 out of 21 clones have a single copy provirus (**Supplementary Figure S8**) further supporting that our infection conditions primarily generate cells infected with a single provirus. Each clone with and without 3′D4Z4 was cultured in parallel, and EGFP expression was monitored for more than 60 days (**Supplementary Figure S9** for HSC1-FRT-D4Z4-GiP clones and **Supplementary Figure S10** for HSC1-HS4-FRT-D4Z4-GiP clones). These data show that clones can be grouped into three distinct classes based on the pattern of transgene expression: rapid, gradual or no silencing. Representative single copy clones were chosen for further investigations.

Rapid silencing. Two of 11 HSC1-FRT-D4Z4-GiP clones (clones 1B10 and clone 1C1) were silenced by day 7, the first analysis after

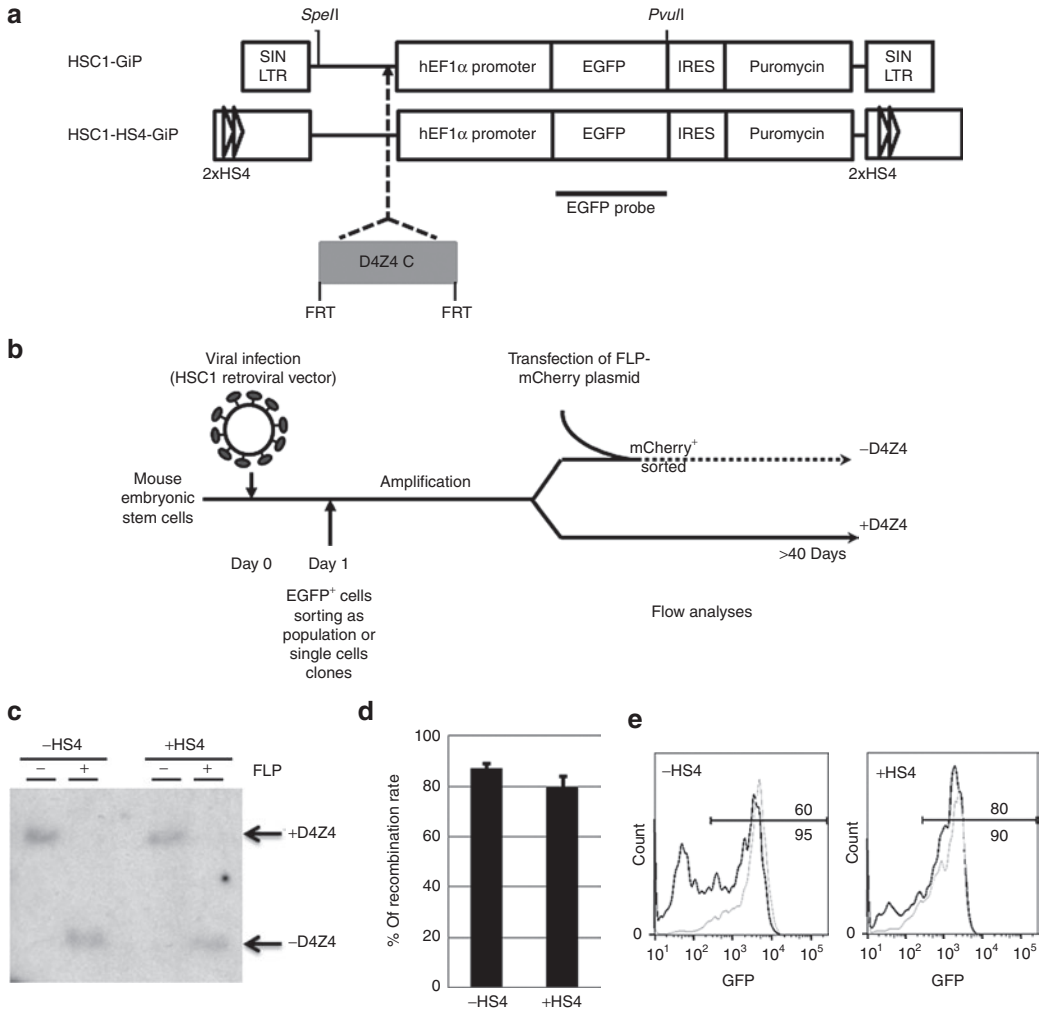

Figure 6 Induction of provirus silencing by FLP recombination. (**a**) Structure of HSC1-FRT-D4Z4-C-GiP and HSC1-HS4-FRT-D4Z4-C-GiP provirus. D4Z4-C fragment flanked by two FRT sites was inserted upstream of the *EF1α* promoter, in HSC1-GiP and HSC1-HS4-GiP retroviral vector. The black line indicates the enhanced green fluorescent protein (EGFP) probe used in the Southern blot analysis. *Spe*I and *Pvu*I restriction sites are shown. (**b**) Experimental system. R1 embryonic stem cells were infected with HSC1-FRT-D4Z4-C-GiP and HSC1-HS4-FRT-D4Z4-C-GiP viral vectors. One day after infection, EGFP-positive cells were sorted as cell populations or as single cell clones and expanded. A portion was transiently transfected with FLP recombinase-mCherry plasmid and 2 days later mCherry-positive cells were sorted and maintained in culture. (**c**) Deletion of 3′D4Z4 by FLP recombinase was verified by Southern blot. gDNA was digested by *Spe*I and *Pvu*I and provirus was detected by ^{32}P-dCTP–labeled EGFP probe. In the presence of 3′D4Z4, the *Spe*I-*Pvu*I fragment is about 4,100 bp and about 2,790 bp when 3′D4Z4 is flipped out. (**d**) Estimation of the deletion frequency by quantitative PCR. D4Z4-C and *EF1α* promoter regions were amplified and the deletion frequency calculated as $1–2^{-(\Delta\Delta Ct)}$ between the flipped out D4Z4-C fragment and the remaining *EF1α* promoter relative to non-flipped cell populations. Each bar represents the mean of three PCRs each performed in triplicate. Error bars represent the SD. (**e**) EGFP expression. Fluorescence was analyzed by flow cytometry 20 days after 3′D4Z4 removal in the HSC1-FRT-D4Z4-GiP population (top chart) and HSC1-HS4-FRT-D4Z4-GiP population (bottom chart) where D4Z4 was flipped (black line) or not flipped (gray line). Top figures are the percentage of EGFP-positive cells in flipped population and bottom figures in non-flipped population. gDNA, genomic DNA; IRES, internal ribosome entry site; LTR, long terminal repeat; SIN, self-inactivating.

deletion of 3′D4Z4 (**Supplementary Figure S9**). Clone 1B10 (**Figure 7a**) is representative of this group, demonstrating that 3′D4Z4 is required for high frequency persistent expression at some integration sites. This observation provides strong evidence that 3′D4Z4 does not function by redirecting retrovirus integration to favouable locations but rather acts by preventing silencing spread over the provirus. It is possible that these rapidly silenced clones have never been studied in cells transduced with non-insulated vectors. They may correspond to the 20% of silent clones shown in **Figure 1** after infection with HSC1-GiP. However, any selection with puromycin would eliminate such rapidly silenced clones and only select for integration sites that exhibit gradual or no silencing of the transgene. These clones are the first definitive evidence for very rapid silencing of a retroviral vector at a specific integration site. Interestingly, none of the 10 HSC1-HS4-FRT-D4Z4-GiP clones displays rapid silencing. One potential explanation for this is that continued presence of HS4 in the LTRs may prevent rapid spread of epigenetic silencing into the provirus, allowing only gradual silencing to proceed.

Gradual silencing. Three of 11 HSC1-FRT-D4Z4-GiP and 3 of 10 HSC1-HS4-FRT-D4Z4-GiP cell clones are gradually silenced

over time (**Supplementary Figures S9** and **S10**). Clone 1A2, which is progressively silenced over the course of 50 days, is representative of this group (**Figure 7a**). These six clones demonstrate that 3′D4Z4 is required to prevent silencing at a substantial proportion of integration sites, even when HS4 is present. This finding implies that 3′D4Z4 blocks gradual silencing that occurs even in the presence of HS4.

No silencing. Six of 11 HSC1-FRT-D4Z4-GiP and 7 of 10 HSC1-HS4-FRT-D4Z4-GiP clones do not silence after 3′D4Z4 deletion.

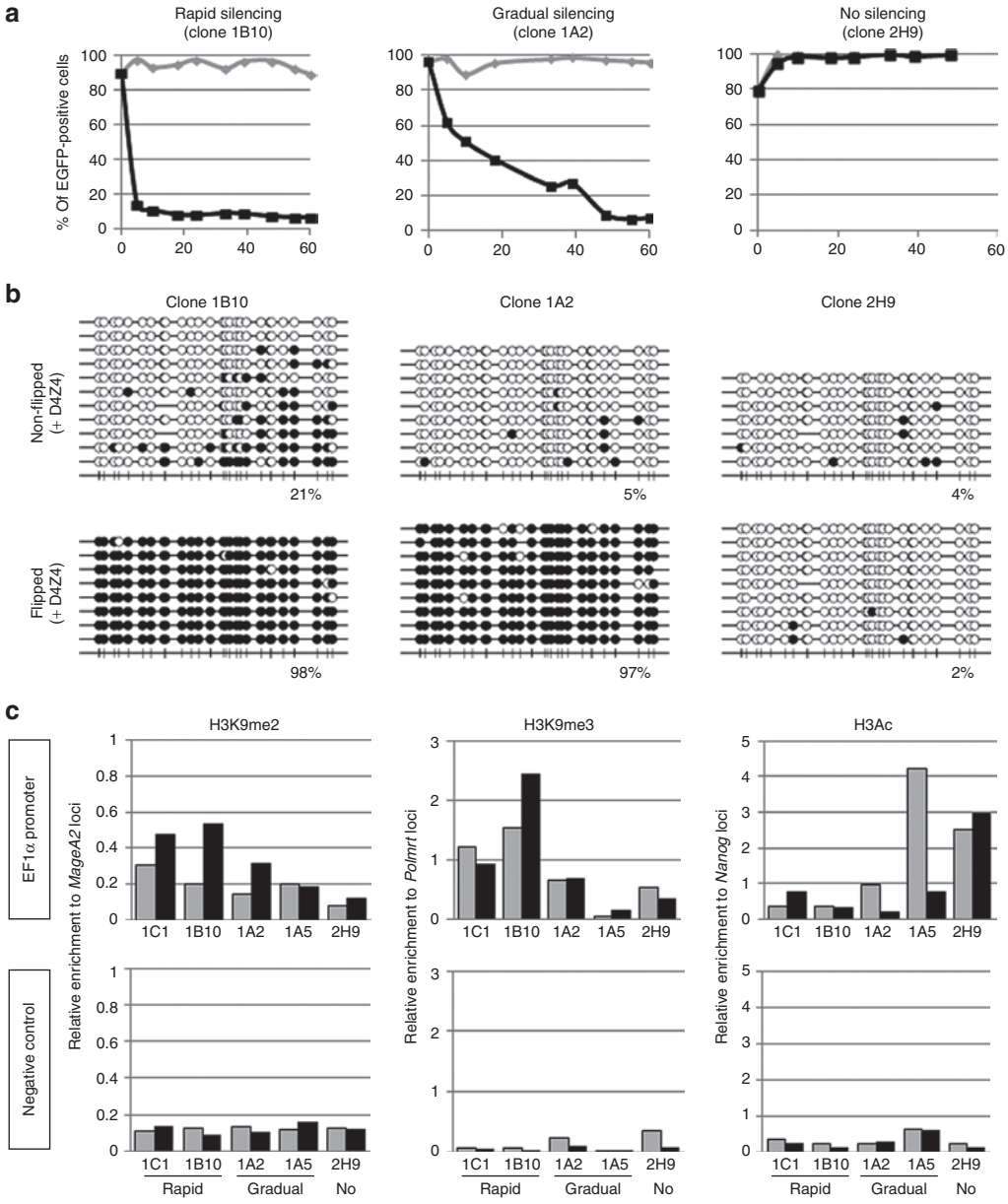

Figure 7 3′D4Z4 protects retrovirus expression from silencing independent of position effects. (**a**) Enhanced green fluorescent protein (EGFP) time course expression in three representative single cell clones with HSC1-FRT-D4Z4-C-GiP provirus. Cells were transfected with FLP recombinase-mCherry plasmid and those containing the provirus with (gray line) and without (black line) 3′D4Z4 were maintained in culture in parallel. (**b**) DNA methylation status of 27 CpG sites in the human *EF1α* promoter obtained from bisulfite sequencing of cells containing provirus with (non-flipped) or without (flipped) 3′D4Z4 36 or 37 passages after FLP recombinase-mCherry plasmid transfection and mCherry sorting. Closed circles represent methylated cytosines and open circles represent unmethylated cytosines. Percentages below each methylation profile represent the percentage of total CpG methylation. (**c**) Analysis of H3K9me2, H3K9me3, and H3Ac histone modifications on the *EF1α* promoter. ChIP-qPCR was performed on representative clones in presence (gray bars) or absence (black bars) of 3′D4Z4 36 or 37 passages after 3′D4Z4 removal. Two rapidly silenced clones (clones 1C1 and 1B10), two gradually silenced clones (clones 1A2 and 1A5), and one not silenced clone (clone 2H9) were analyzed. qPCR on the *EF1α* promoter (top) was used to evaluate fold enrichment relative to input and normalized to the endogenous positive controls *MageA2* for H3K9me2, *Polmrt* for H3K9me3, and *Nanog* for H3Ac. Chip-qPCRs were performed in triplicate. The negative controls (below) are the endogenous *Nanog* region for H3K9me2 and H3K9me3, and *MageA2* for H3Ac. qPCR, quantitative PCR.

A representative example is clone 2H9 (**Figure 7a**). These data indicate that more than half of the clones contain retroviral vectors at favorable genomic locations that are not subject to silencing. The frequency of this class was likely enriched in the initial sorting for EGFP-expressing cells, which would exclude any cells that never expressed the provirus, or are silenced immediately after infection.

3′D4Z4 does not redirect transgenes to perinuclear positions

The 5′D4Z4 element has been shown to be capable of relocating a telomere to the nuclear periphery.[37] We were interested in whether the D4Z4-C fragment can also relocalize a retroviral transgene to the nuclear periphery. DNA immunoFISH experiments to label a 3 kb transgene are technically challenging and were unsuccessful; hence, we aimed to determine the nuclear localization of the integration sites. For this purpose, we first determined the integration sites of the three classes of HSC1-FRT-D4Z4-GiP by performing ligation-mediated PCR and DNA sequencing of the provirus-genome junction fragments. BLAT analyses (http://www.genome.ucsc.edu) of the ligation-mediated PCR products identify the genomic integration sites (**Supplementary Table S2**) and bacterial artificial chromosome fluorescence *in situ* hybridization (FISH) probes were chosen accordingly for one representative clone from each class. Since the precise integration site of clone 1B10 could not be narrowed down to <3 potential sites, clone 1C1 was chosen as the rapidly silenced clone; clones 1A2 and 2H9 were chosen as representative of gradual and no silencing classes, respectively. Association of the integration site with the nuclear periphery was defined by colocalization of FISH signal with lamin B1.

The rapidly silenced clone 1C1 is integrated far from any transcription unit (590 kb from *Tox* and 650 kb from *Car8*) (**Supplementary Figure S11**) in a lamin-associated domain.[42] Analysis of published ChIP-seq data in ES cells[15,43–46] reveals that active chromatin marks (H3K4me, H3K9Ac, and DNase hypersensitive sites) are absent and repressive marks, including H3K9me2 and H3K9me3, are present (**Supplementary Figure S11**). DNA FISH analysis shows 40% of these alleles are found at the nuclear periphery in non-infected ES cells or infected cells with or without 3′D4Z4 (**Supplementary Figure S14**). This suggests that 3′D4Z4 does not modify the perinuclear localization at this integration site.

The provirus in the gradual silenced clone 1A2 is integrated in a gene desert with the nearest genes being 151 kb upstream or 2.7 Mb downstream (**Supplementary Figure S12**). This integration site has not been associated with a lamin-associated domain in ES cells, but is close (within 2 kb) to a genomic region enriched in ES cells for H3K9me3 histone marks, which are often associated with a repressive chromatin state. Our DNA FISH analysis in non-infected ES cells shows that the site is found at the nuclear periphery, with 50% of the alleles found to be associated with lamin B staining (**Supplementary Figure S14**). After integration of the expressing provirus containing the 3′D4Z4 insulator, the integration site displays moderate, but significant movement away from the periphery with 40% of the alleles found at the periphery. After the removal of 3′D4Z4, the allele moves back towards the nuclear periphery, suggesting that silencing of the transgene correlates with movement of the transgene towards the nuclear periphery.

Finally, the provirus in the no silencing clone 2H9 is located in an intragenic region 41 kb upstream of *Tcf7* and 36 kb downstream of *Vdac1* genes (**Supplementary Figure S13**). Analysis of published ChIP-seq data reveals several regions in ES cells marked by H3K4me1, H3K4me2, and H3K27ac, including two regions about 15.5 and 21.5 kb away from the integration site. These regions correspond to DNase hypersensitive sites and may be active enhancers of nearby genes. This locus is not associated with a lamin-associated domain and DNA FISH shows that only 4% of alleles are found at the nuclear periphery (**Supplementary Figure S14**). Insertion of a transgene with D4Z4-C does not alter its localization, although a very small but significant movement towards the periphery can be observed after the removal of D4Z4-C. Overall, our data from three clones suggest that 3′D4Z4 is unlikely to play a consistent role in the perinuclear localization of retroviral transgenes, and that silencing does not correlate with a dramatic movement of the retroviral transgene into the nucleus.

3′D4Z4 protects transgenes from heterochromatin spread

To determine which epigenetic modifications might be blocked by 3′D4Z4 and explain the difference in proviral expression between clones from each silencing kinetic class, we analyzed proviral DNA methylation and histone modifications 36 or 37 passages (80–90 days) after 3′D4Z4 removal. At that stage of cell culture, only 1–2% of cells are EGFP-positive in clones 1B10 and 1A2, respectively when 3′D4Z4 is deleted, but 75–98% of cells are EGFP-positive in the no silencing 2H9 clone and in all the clones when 3′D4Z4 is present (**Supplementary Figure S15**). All cell cultures are undifferentiated because 96% of the cells are SSEA1-positive. Bisulfite sequencing (**Figure 7b** and **Supplementary Figure S16b**) reveals that the rapid-silencing clones (1B10 and 1C1) have a partially methylated *EF1α* promoter in the presence of 3′D4Z4, but are hypermethylated following deletion. This demonstrates that 3′D4Z4 is capable of partially protecting the transgene from DNA methylation at this site, and the pre-existing partial DNA methylation may promote the rapid silencing observed after 3′D4Z4 deletion.

In contrast, the gradual-silencing clone 1A2 is hypomethylated at the promoter in the presence of 3′D4Z4 and is fully hypermethylated following deletion (**Figure 7b**). Clone 1A5, which has slower gradual silencing than clone 1A2, has only 60% EGFP-positive cells by day 60 (**Supplementary Figure S16a**). In this case, *EF1α* is hypomethylated in the presence of 3′D4Z4, but is only partially methylated in the absence of 3′D4Z4 (**Supplementary Figure S16b**). This shows that 3′D4Z4 can also fully protect the promoter from DNA methylation at some sites, and silencing may be more gradual at the sites that lack pre-existing partial DNA methylation. Not surprisingly, the no silencing clone 2H9 is hypomethylated with or without 3′D4Z4, indicating that these sites may be in locations that are not susceptible to DNA methylation.

Finally, we analyzed histone modifications on the provirus by ChIP-qPCR for repressive chromatin or open chromatin marks on the retroviral *EF1α* promoter in comparison to endogenous loci. In the two rapid clones, the H3K9me2 mark shows low

enrichment compared with the positive control *MageA2* (0.2–0.5) (**Figure 7c**, top panel) but are 2–5 times greater than the negative control *Nanog* (**Figure 7c**, bottom panel). Deletion of 3′D4Z4 increases this enrichment by up to 2.6-fold. In these two clones, the H3K9me3 enrichment levels are similar to the positive control (Polmrt, ref. 45) but increased after 3′D4Z4 deletion only in clone 1B10. Both clones have a low enrichment of active chromatin marks in the presence or absence of 3′D4Z4, consistent with the EGFP expression level in these clones.

In the clones 1A2 and 1A5 which are gradually silenced, the promoter gains modest H3K9me2 enrichment only in clone 1A2 and the constitutive repressive chromatin mark (H3K9me3) does not change after 3′D4Z4 removal in both the clones. In contrast, the greatest fold difference in 1A2 and 1A5 cells is the loss of active chromatin marks when the insulator is deleted. Indeed, in cells containing provirus without 3′D4Z4, enrichment of H3Ac on the *EF1*α promoter is about fivefold lower than enrichment in provirus with 3′D4Z4. It is noteworthy that enrichment on the *EF1*α promoter in clone 1A5 in the presence of 3′D4Z4 is fourfold higher than the endogenous *Nanog* gene and drops to a lower level of enrichment when 3′D4Z4 is deleted (0.8-fold). Finally, clone 2H9, which is not silenced after 3′D4Z4 deletion, does not display any significant chromatin alterations. Overall, we conclude that different repressive histone marks are associated with each silent integration site, indicating that multiple pathways may influence transcription status, depending upon the proviral integration site. However, proviruses within each kinetic class display similar patterns. Rapid silenced clones are characterized by partial DNA methylation of the provirus in the presence of 3′D4Z4 and the highest enrichment in repressive marks combined with the lowest enrichment in active chromatin marks. Gradual-silencing clones are characterized by a dramatic loss in an active chromatin mark when 3′D4Z4 is deleted. Despite the heterogeneity in repressive marks near or on the provirus, 3′D4Z4 protects transgenes from the spread of repressive chromatin marks by blocking DNA methylation and minimizing the intrusion of repressive histone marks.

DISCUSSION

Our findings provide direct evidence that 3′D4Z4 protects retroviral vectors from DNA hypermethylation and deleterious histone modifications that are dependent on the integration site. This activity allows 3′D4Z4 to direct high frequency and long-term expression from HSC1 retroviral vectors in ES cells. When combined with HS4 in the LTRs, cooperation between the two insulators directs unprecedented frequencies of persistent transgene expression in all preselected genomic integration sites that may have important practical applications for genetically manipulating PS cells. Deletion of 3′D4Z4 by FLP recombination induces transgene silencing which can be classified into three distinct profiles defined by different silencing kinetics. This approach allows us to initiate silencing at defined integration sites as a tool to decipher mechanisms that are distinct for each kinetic class.

High frequency retrovirus expression directed by 3′D4Z4

We tested the ability of D4Z4 to improve the frequency of long-term expression of a SIN retroviral vector in ES cells. We found that 3′D4Z4 prevents retroviral vector transgene silencing. This activity was not due to a spacing effect and did not require the *DUX4* promoter or its complete ORF. However, cell populations containing 3′D4Z4 exhibited high cell-to-cell variability, indicating that additional insulator activities are important to obtain consistent expression levels. In contrast, 5′D4Z4 did not prevent the generation of a substantial proportion of EGFP-negative cells in which the vector DNA had been lost. This suggests that 5′D4Z4 sequences are unstable in this retroviral vector over time. However, we noted that the proportion of cells that do express D4Z4-A virus produced high EGFP intensities with very minimal cell-to-cell variability even in the absence of selection. Thus when selection is maintained, 5′D4Z4 may have valuable properties that enhance expression levels. For our purposes of maintaining high frequency and persistent expression from retroviral vectors in the absence of selection, the very efficient activity present in 3′D4Z4 and its stability in unselected cells are more appropriate.

3′D4Z4 and HS4 cooperate to block retrovirus silencing

Since 3′D4Z4 was unable to prevent variable expression, we reasoned that the location upstream of the internal promoter may be insufficient to prevent downstream genomic position effects. To address this concern, we combined a dimer HS4 core in the LTRs with 3′D4Z4 upstream of the *EF1*α gene promoter. After selection, the transduced cell population harboring this vector had high frequency and persistent levels of expression coupled with minimal variability in transgene expression. We conclude that the two insulators cooperate to maintain retroviral transgene expression over time. Since 3′D4Z4 lacks the CTCF site present in 5′D4Z4, it is possible that a more complete silencing activity is produced by combining 3′D4Z4 with HS4 which retains its CTCF site. To our knowledge, this is the first evidence for HS4 cooperation with another insulator resulting in a stronger combined activity that completely neutralizes long-term retrovirus silencing.

Epigenetic analysis of transduced cells confirmed that the silent uninsulated virus is DNA hypermethylated. In contrast, the HS4 virus had partial methylation while 3′D4Z4 virus was hypomethylated on the viral *GAG* and *EF1*α sequences but had partial methylation on the *EGFP* sequences. Cooperation between HS4 and 3′D4Z4 directed hypomethylation on *GAG*, *EF1*α, and *EGFP*, consistent with full barrier or other activities that can block the spread of DNA methylation. In addition, ChIP-qPCR on the *EF1*α promoter demonstrated that H3K9me3 was enriched and H3Ac was depleted on silent uninsulated virus relative to active 3′D4Z4 containing virus in cell populations. Together these data indicate that 3′D4Z4 cooperates with HS4 barrier to block DNA methylation and H3K9me3 spread.

We interpret the ability of 3′D4Z4 and HS4 to cooperate as an indication that they insulate the vector from different sequences or epigenetic marks. We propose that both insulators function to counteract genomic position effects, and the internally located 3′D4Z4 also acts to block cryptic silencer elements that remain in the SIN LTR, *GAG* or reporter gene sequences. Given that HS4 has an enhancer blocking activity mediated by DNA looping between CTCF-binding sites,[47] we infer that either 3′D4Z4 does not interfere with the presumed looping between the two LTRs,

or enhancer blocking activity by HS4 is not required for its cooperativity with D4Z4. In contrast, HS4 barrier activity is mediated by USF and VEZF1 binding which would not be affected by an internally located 3′D4Z4 element. Overall, our novel retroviral vector design employing both 3′D4Z4 and HS4 is the first to fully escape long-term silencing in ES cells. In situations where selection can be applied, these vectors should ensure persistent long-term expression in PS cells, which is an important criterion for any reporter transgenes designed to mark these stem cells or for potential suicide vectors intended to kill undifferentiated PS cells after cell transplantation. Moreover, expression from the *EF1*α promoter is maintained after spontaneous *in vitro* differentiation into embryoid bodies, although it is not clearly established that D4Z4 function is ubiquitous in all differentiated cell types. Further experiments are required to test the ability of D4Z4 to shield other promoters in differentiated cell types that may be applicable for disease modeling, recombinant protein production, and gene therapy.

Deletion of 3′D4Z4 identifies three kinetic classes of silencing

To determine whether 3′D4Z4 prevents silencing at hostile integration sites or merely acts by redirecting integration into favorable integration sites, we developed an FRT-based system to delete the insulator after integration. Our results show that deletion of 3′D4Z4 allows silencing to occur, and that the continued presence of HS4 also allows silencing to occur but to a lesser degree. These findings are not consistent with the preferred integration site model but rather support a model for 3′D4Z4 function *via* effects on the surrounding chromatin landscape.

To evaluate silencing kinetics, we isolated ES cell clones that express a single copy provirus in the presence of 3′D4Z4 but are silenced after its deletion. This approach revealed three silencing classes distinguished by their kinetics. Rapid-silencing clones were detected after deletion of 3′D4Z4 in the absence of HS4, gradual-silencing clones were found after deletion of 3′D4Z4 in the presence or absence of HS4, and roughly half of all clones were not subject to silencing after 3′D4Z4 deletion. The discovery of these three classes shows that integration sites silence retroviral transgenes with different kinetics, and demonstrate that some integration sites are completely deleterious for expression unless 3′D4Z4 is present to protect the transgene from these rapidly silenced neighborhoods. The different kinetics observed agree well with involvement of multiple-silencing pathways that could be dependent on the integration site.[2,21] Identification of representative integration sites for each class provide examples suggesting that rapid silencing can occur in gene deserts, gradual silencing can occur at an intragenic site near regions with heterochromatin marks, and no silencing can occur at an integration site near expressing endogenous genes enriched in an active chromatin mark. While no general associations of genomic position effects on proviral expression can be made from our limited integration site analysis, bacterial artificial chromosome probes for the integration sites permitted 3D DNA FISH to be performed before and after 3′D4Z4 deletion to examine association with the nuclear periphery. These experiments suggest that the rapid- and gradual-silencing integration sites had preferential localization to the periphery, whereas the no silencing integration site was preferentially excluded from the periphery. Our results indicate that 3′D4Z4 did not consistently alter this localization, and therefore 3′D4Z4 functions differently from 5′D4Z4 that preferentially directs transfected transgenes toward the periphery.

To examine the 3′D4Z4 activity, we performed epigenetic analyses on the provirus in representative clones of the three silencing classes. The rapid-silencing clones exhibited partial DNA methylation on the *EF1*α promoter that becomes hypermethylated following 3′D4Z4 deletion. The gradual-silencing clones were hypomethylated and also became hypermethylated following D4Z4 deletion. In contrast, the no silencing clone was hypomethylated in the presence or absence of 3′D4Z4. These data unambiguously demonstrate that 3′D4Z4 blocks DNA methylation. This may correlate well with hypomethylation on endogenous D4Z4 elements in facioscapulohumeral dystrophy patients compared with the non-affected population,[48,49] and our data is the first evidence that D4Z4 itself is capable of preventing DNA methylation of a transgene. The ability of D4Z4 on one side of the transgene to create a DNA methylation-free area that extends into the linked promoter is similar to the activity of the UCOE element on the *EF1*α short promoter after lentiviral transduction of iPS and ES cells.[50] In addition, rapid-silencing clones have low levels of H3K9me2 and high levels of H3K9me3 enrichment which correlates with a low level of expression. Deletion of 3′D4Z4 modestly increases H3K9me2 levels in both the clones, but H3K9me3 is only raised in the clone 1B10. In contrast, gradual-silencing clones lost H3Ac marks upon deletion of 3′D4Z4. These results suggest that 3′D4Z4 also contributes to blocking H3K9 methylation marks and enhancing H3 acetylation marks. Our data demonstrate that 3′D4Z4 blocks the spread of repressive epigenetic marks, but the precise factors that mediate this activity are unknown, and took many years to identify with regard to HS4 barrier activity. It is possible they recruit different but complementary factors, and this information may also help to distinguish whether 3′D4Z4 and UCOE are functionally equivalent or distinct. Given that epigenetic marks differ at each integration site and silencing occurs with different kinetics, our results are consistent with the involvement of multiple epigenetic pathways that are dependent at least in part on the surrounding epigenome.

In summary, our use of FLP to delete 3′D4Z4 is a definitive method to identify elements that block transgene silencing and is the first application of this approach in mammalian cells. This tool revealed three classes of silencing kinetics and a diversity of epigenetic alterations on the silent transgene that are governed by multiple-silencing pathways. Better definition of the very earliest epigenetic events that occur on the rapid-silencing clones will require development of a synchronous system for inducing FLP to obtain sufficient cells for epigenetic analyses after 3′D4Z4 deletion. Such a system will not only identify the kinetics of epigenetic events during rapid retrovirus silencing, but also provide a tool for screening chemicals that can inhibit these pathways.

MATERIALS AND METHODS

Cell culture and retroviral infection. Mouse ES cells were cultured in mouse ES medium: Dulbecco's modified Eagle's medium with 15% fetal bovine serum qualified for ES cell culture supplemented with 4 mmol/l

L-glutamine, 0.1 mmol/l minimum essential medium non-essential amino acids, 1 mmol/l sodium pyruvate, 0.55 mmol/l 2-mercaptoethanol (all from Invitrogen, Carlsbad, CA) and purified recombinant leukemia inhibitory factor.[18] J1 mouse ES cells were cultured on gelatine (0.1%)-coated plates; R1 ES cells and mouse iPS cells were grown on feeders. Retroviral production was performed with the Plat-E packaging cell line. Retroviral vector constructions are described in **Supplementary Materials and Methods**. Plasmids described in the paper will be available from Addgene (http://www.addgene.org). All are derived from the HSC1 SIN retroviral vector. One day before infection, ES cells were seeded at 1×10^4 cells per well of a gelatine-coated 24-well plate. For infection, virus was added to the target cells in the presence of 8 µg/ml Polybrene (hexadimethrine bromide; Sigma, St Louis, MO).

Flow cytometry. Single cell suspensions were fixed with 2% formaldehyde in phosphate-buffered saline supplemented by 2% fetal bovine serum for 10 minutes at room temperature. Cells were then suspended in phosphate-buffered saline with 2% fetal bovine serum and filtered through 70 µm nylon membranes. EGFP expression analyses were performed by LSRII flow cytometer (Becton Dickinson, San Jose, CA) using CellQuest Pro software (Becton Dickinson). Cell debris were excluded from analysis by using forward and side scatter gating. Non-infected cells were used as a negative control to adjust EGFP fluorescence.

Slot blot. gDNA was extracted directly from 96-well plates. Cells were digested overnight with proteinase K at 100 µg/ml in lysis buffer (100 mmol/l Tris-HCl, pH 8; 5 mmol/l EDTA, 0.2% SDS, 200 mmol/l NaCl) at 55 °C. gDNA was denaturated with 0.4 N NaOH for 5 minutes at room temperature, neutralized with cold NH_4Ac (final concentration, 1 mol/l), and then transferred onto nitrocellulose membrane. These membranes were processed following standard procedures and hybridized with an EGFP probe to detect integrated provirus and *mThy1*, to detect the endogenous mouse Thy1 gene.

Southern blot. gDNA were extracted by conventional proteinase K digestion and phenol/chloroform extraction. They were digested with *Bgl*II or *Spe*I-*Pvu*I, separated on a 1% agarose gel and transferred onto a membrane (Protran BA85 pure nitrocellulose; Schleicher & Schuell, Dassel, Germany). Blots were probed with a fragment of EGFP labeled with P^{32} by random priming (Megaprime DNA Labelling System, RPN1606; GE Healthcare, Buckinghamshire, UK).

gDNA methylation analysis. gDNA was isolated from ES cells using phenol-chloroform extraction. The DNA (400 ng) was then bisulfite-treated with the EZ DNA methylation-gold kit (Zymo Research, Irvine, CA) to convert unmethylated cytosines into uracil, whereas methylated cytosines remain intact. OSG13/OSG14, Bis-PGKEFGP2f/Bis-PGKEFGP2r, and OSG19/OSG20 primers sets (**Supplementary Table S1**) were used to amplify the *GAG*, *EGFP*, and the *EF1*α gene promoter regions, respectively using Platinum Taq DNA polymerase high fidelity (#11304-011; Invitrogen). PCR products were subcloned in pGEM-T easy vector (A13160; Promega, Madison, WI) and sequenced by The Centre for Applied Genomics facility (Hospital for Sick Children, Toronto, Ontario, Canada).

Chromatin immunoprecipitation assay and qPCR. ChIP was performed using the magna ChIP G kit (#17-611; Millipore, Billerica, MA) following the manufacturer's instructions. Briefly, 10^7 cells were cross-linked in 1% formaldehyde; cells were washed, scraped into a conical tube, and sonicated for 25 minutes in a Bioruptor XL at the high setting with cycles of 30 seconds ON, 30 seconds OFF. Chromatin was subdivided into aliquots containing 10^6 cell equivalents of lysate for each immunoprecipitation. This chromatin was immunoprecipitated with 2 mg of antibodies specific to H3K9me2 (Ab1220; Abcam, Cambridge, MA), H3K9me3 (Ab8898; Abcam), H3Ac (06-599; Millipore), mouse IgG control (Ab18413; Abcam), and rabbit IgG control (ab37415; Abcam). In all, 0.9 µl of purified DNA was analyzed by qPCR using SYBR green (#4309155; Applied Biosystems, Foster City, CA) and ABI7900HT fast real time PCR system (Applied Biosystems). The primer sequences are given in **Supplementary Table S1**.

SUPPLEMENTARY MATERIAL

Figure S1. HSC1 retroviral vectors are silenced over time in iPS cells but not fibroblasts.
Figure S2. Transgene silencing is not due to spontaneous differentiation.
Figure S3. D4Z4 fragments do not contain an enhancer activity in mouse ES cells.
Figure S4. HSC1 D4Z4-A retroviral vector is unstable overtime.
Figure S5. 3′D4Z4 in cooperation with HS4 maintains high frequency and level of expression during differentiation.
Figure S6. 3′D4Z4 cooperates with HS4 to direct high level and low variability of expression.
Figure S7. Deletion frequency for each cell clone.
Figure S8. Integrated copy number of HSC1-FRT-D4Z4-C-GiP and HSC1-HS4-FRT-D4Z4-C-GiP clones.
Figure S9. EGFP expression time course in HSC1-FRT-D4Z4-C-GiP ± flipping.
Figure S10. EGFP expression time course in HSC1-HS4-FRT-D4Z4-C-GiP ± flipping.
Figures S11, S12, and S13. Analysis of genomic integration sites in three representative cell clones.
Figure S14. 3′D4Z4 does not consistently affect retroviral transgene nuclear localization.
Figure S15. Silencing is not due to spontaneous cell differentiation.
Figure S16. 3′D4Z4 protects retroviral transgenes from silencing independent of genomic integration site.
Table S1. List of primers.
Table S2. Integration site of provirus.
Materials and Methods.

ACKNOWLEDGMENTS

We thank Mehdi Karimi (University of British Columbia, Vancouver, British Columbia, Canada) for help with generating the tracks for bioinformatics analysis, Oliver Tam for helpful discussions on ChIP-seq data analysis, and Akitsu Hotta (CiRA, Kyoto, Japan) for critical comments on the manuscript. We acknowledge the assistance of Sherry Zhao for fluorescence-activated cell sorting in the SickKids-UHN Flow Cytometry Facility, Jodi Garner in the SickKids ES Cell Facility, Mike Woodside and Paul Paroutis in the SickKids Imaging Facility and the SickKids TCAG Sequencing Facility. We thank Eric Gilson and Frederique Magdinier (LBMC, ENS, Lyon, France) for providing the C1X plasmid and Elly Tanaka (CRTD, Dresden, Germany) for SKSce-CAGGs-PA plasmid. This work was funded by grants from the Canadian Institutes of Health Research (CIHR MOP-111065 and IG1-102956 to J.E. and RMF-92090 to M.C.L. and J.E.) and a CIHR Frederick Banting and Charles Best Graduate Scholarship to M.Y.M.L. The authors declared no conflict of interest.

REFERENCES

1. Barklis, E, Mulligan, RC and Jaenisch, R (1986). Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* **47**: 391–399.
2. Pannell, D and Ellis, J (2001). Silencing of gene expression: implications for design of retrovirus vectors. *Rev Med Virol* **11**: 205–217.
3. Osborne, CS, Pasceri, P, Singal, R, Sukonnik, T, Ginder, GD and Ellis, J (1999). Amelioration of retroviral vector silencing in locus control region beta-globin-transgenic mice and transduced F9 embryonic cells. *J Virol* **73**: 5490–5496.
4. Ryan, RF, Schultz, DC, Ayyanathan, K, Singh, PB, Friedman, JR, Fredericks, WJ *et al.* (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol Cell Biol* **19**: 4366–4378.
5. Schultz, DC, Ayyanathan, K, Negorev, D, Maul, GG and Rauscher, FJ 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**: 919–932.
6. Schultz, DC, Friedman, JR and Rauscher, FJ 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form

a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev* **15**: 428–443.
7. Wolf, D and Goff, SP (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**: 46–57.
8. Wolf, D and Goff, SP (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**: 1201–1204.
9. Dodge, JE, Ramsahoye, BH, Wo, ZG, Okano, M and Li, E (2002). De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289**: 41–48.
10. Challita, PM and Kohn, DB (1994). Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc Natl Acad Sci USA* **91**: 2567–2571.
11. Dalle, B, Rubin, JE, Alkan, O, Sukonnik, T, Pasceri, P, Yao, S *et al.* (2005). eGFP reporter genes silence LCRbeta-globin transgene expression via CpG dinucleotides. *Mol Ther* **11**: 591–599.
12. Lorincz, MC, Schübeler, D and Groudine, M (2001). Methylation-mediated proviral silencing is associated with MeCP2 recruitment and localized histone H3 deacetylation. *Mol Cell Biol* **21**: 7913–7922.
13. Fuks, F, Hurd, PJ, Wolf, D, Nan, X, Bird, AP and Kouzarides, T (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* **278**: 4035–4040.
14. Pannell, D, Osborne, CS, Yao, S, Sukonnik, T, Pasceri, P, Karaiskakis, A *et al.* (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *EMBO J* **19**: 5884–5894.
15. Karimi, MM, Goyal, P, Maksakova, IA, Bilenky, M, Leung, D, Tang, JX *et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**: 676–687.
16. Dong, KB, Maksakova, IA, Mohn, F, Leung, D, Appanah, R, Lee, S *et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* **27**: 2691–2701.
17. Leung, DC, Dong, KB, Maksakova, IA, Goyal, P, Appanah, R, Lee, S *et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc Natl Acad Sci USA* **108**: 5718–5723.
18. Vedadi, M, Barsyte-Lovejoy, D, Liu, F, Rival-Gervier, S, Allali-Hassani, A, Labrie, V *et al.* (2011). A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat Chem Biol* **7**: 566–574.
19. Roth, SL, Malani, N and Bushman, FD (2011). Gammaretroviral integration into nucleosomal target DNA *in vivo*. *J Virol* **85**: 7393–7401.
20. Wu, X, Li, Y, Crise, B and Burgess, SM (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**: 1749–1751.
21. Senigl, F, Auxt, M and Hejnar, J (2012). Transcriptional provirus silencing as a crosstalk of de novo DNA methylation and epigenomic features at the integration site. *Nucleic Acids Res* **40**: 5298–5312.
22. Dickson, J, Gowher, H, Strogantsev, R, Gaszner, M, Hair, A, Felsenfeld, G *et al.* (2010). VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* **6**: e1000804.
23. Bell, AC, West, AG and Felsenfeld, G (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447–450.
24. Chung, JH, Whiteley, M and Felsenfeld, G (1993). A 5′ element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* **74**: 505–514.
25. Ghirlando, R, Giles, K, Gowher, H, Xiao, T, Xu, Z, Yao, H *et al.* (2012). Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta* **1819**: 644–651.
26. Phillips, JE and Corces, VG (2009). CTCF: master weaver of the genome. *Cell* **137**: 1194–1211.
27. Bell, AC, West, AG and Felsenfeld, G (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387–396.
28. West, AG, Huang, S, Gaszner, M, Litt, MD and Felsenfeld, G (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell* **16**: 453–463.
29. Emery, DW (2011). The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. *Hum Gene Ther* **22**: 761–774.
30. Giraldo, P, Rival-Gervier, S, Houdebine, LM and Montoliu, L (2003). The potential benefits of insulators on heterologous constructs in transgenic animals. *Transgenic Res* **12**: 751–755.
31. Aker, M, Tubb, J, Groth, AC, Bukovsky, AA, Bell, AC, Felsenfeld, G *et al.* (2007). Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum Gene Ther* **18**: 333–343.
32. Moltó, E, Fernández, A and Montoliu, L (2009). Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Brief Funct Genomic Proteomic* **8**: 283–296.
33. Truffinet, V, Guglielmi, L, Cogné, M and Denizot, Y (2005). The chicken beta-globin HS4 insulator is not a silver bullet to obtain copy-number dependent expression of transgenes in stable B cell transfectants. *Immunol Lett* **96**: 303–304.
34. Yao, S, Osborne, CS, Bharadwaj, RR, Pasceri, P, Sukonnik, T, Pannell, D *et al.* (2003). Retrovirus silencer blocking by the cHS4 insulator is CTCF independent. *Nucleic Acids Res* **31**: 5317–5323.
35. Wijmenga, C, Hewitt, JE, Sandkuijl, LA, Clark, LN, Wright, TJ, Dauwerse, HG *et al.* (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* **2**: 26–30.
36. Lemmers, RJ, van der Vliet, PJ, Klooster, R, Sacconi, S, Camaño, P, Dauwerse, JG *et al.* (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**: 1650–1653.
37. Ottaviani, A, Rival-Gervier, S, Boussouar, A, Foerster, AM, Rondier, D, Sacconi, S *et al.* (2009). The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet* **5**: e1000394.
38. Petrov, A, Allinne, J, Pirozhkova, I, Laoudj, D, Lipinski, M and Vassetzky, YS (2008). A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity *in vivo*: implications for the facio-scapulo-humeral dystrophy. *Genome Res* **18**: 39–45.
39. Ding, H, Beckers, MC, Plaisance, S, Marynen, P, Collen, D and Belayew, A (1998). Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. *Hum Mol Genet* **7**: 1681–1694.
40. Dixit, M, Ansseau, E, Tassin, A, Winokur, S, Shi, R, Qian, H *et al.* (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci USA* **104**: 18157–18162.
41. Gabriëls, J, Beckers, MC, Ding, H, De Vriese, A, Plaisance, S, van der Maarel, SM *et al.* (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* **236**: 25–32.
42. Peric-Hupkes, D, Meuleman, W, Pagie, L, Bruggeman, SW, Solovei, I, Brugman, W *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**: 603–613.
43. Creyghton, MP, Cheng, AW, Welstead, GG, Kooistra, T, Carey, BW, Steine, EJ *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**: 21931–21936.
44. Meissner, A, Mikkelsen, TS, Gu, H, Wernig, M, Hanna, J, Sivachenko, A *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766–770.
45. Mikkelsen, TS, Ku, M, Jaffe, DB, Issac, B, Lieberman, E, Giannoukos, G *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–560.
46. Stadler, MB, Murr, R, Burger, L, Ivanek, R, Lienert, F, Schöler, A *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–495.
47. Gaszner, M and Felsenfeld, G (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**: 703–713.
48. Tsien, F, Sun, B, Hopkins, NE, Vedanarayanan, V, Figlewicz, D, Winokur, S *et al.* (2001). Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. *Mol Genet Metab* **74**: 322–331.
49. van Overveld, PG, Lemmers, RJ, Sandkuijl, LA, Enthoven, L, Winokur, ST, Bakels, F *et al.* (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* **35**: 315–317.
50. Pfaff, N, Lachmann, N, Ackermann, M, Kohlscheen, S, Brendel, C, Maetzig, T *et al.* (2013). A ubiquitous chromatin opening element prevents transgene silencing in pluripotent stem cells and their differentiated progeny. *Stem Cells* **31**: 488–499.