



Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure

Ya-Chi Ho,¹ Liang Shan,^{1,5} Nina N. Hosmane,¹ Jeffrey Wang,² Sarah B. Laskey,¹ Daniel I.S. Rosenbloom,³ Jun Lai,¹ Joel N. Blankson,¹ Janet D. Siliciano,¹ and Robert F. Siliciano^{1,4,*}

¹Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

²Health Sciences Center, School of Medicine, Louisiana State University, New Orleans, LA 70112, USA

³Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

⁴Howard Hughes Medical Institute, Baltimore, MD 21205, USA

⁵Present address: Department of Immunobiology, School of Medicine, Yale University, New Haven, CT 06510, USA

*Correspondence: rsiliciano@jhmi.edu

<http://dx.doi.org/10.1016/j.cell.2013.09.020>

SUMMARY

Antiretroviral therapy fails to cure HIV-1 infection because latent proviruses persist in resting CD4⁺ T cells. T cell activation reverses latency, but <1% of proviruses are induced to release infectious virus after maximum *in vitro* activation. The noninduced proviruses are generally considered defective but have not been characterized. Analysis of 213 noninduced proviral clones from treated patients showed 88.3% with identifiable defects but 11.7% with intact genomes and normal long terminal repeat (LTR) function. Using direct sequencing and genome synthesis, we reconstructed full-length intact noninduced proviral clones and demonstrated growth kinetics comparable to reconstructed induced proviruses from the same patients. Noninduced proviruses have unmethylated promoters and are integrated into active transcription units. Thus, it cannot be excluded that they may become activated *in vivo*. The identification of replication-competent noninduced proviruses indicates that the size of the latent reservoir—and, hence, the barrier to cure—may be up to 60-fold greater than previously estimated.

INTRODUCTION

Despite prolonged antiretroviral therapy (ART), HIV-1 persists as transcriptionally inactive proviruses in resting memory CD4⁺ T cells (Chun et al., 1997; Finzi et al., 1997; Wong et al., 1997). This latent reservoir (LR) has a long half-life, preventing cure by ART alone (Finzi et al., 1997; Siliciano et al., 2003; Strain et al., 2003). In resting CD4⁺ T cells, the lack of active forms of key cellular transcription factors (Böhnlein et al., 1988; Duh et al., 1989; Ganesh et al., 2003; Kinoshita et al., 1997; Nabel and Baltimore, 1987; West et al., 2001) and of HIV-1 Tat and its cellular cofactors (Cujec et al., 1997; Herrmann and Rice, 1995; Jones

and Peterlin, 1994; Kao et al., 1987; Selby and Peterlin, 1990; Tyagi et al., 2010) limits the initiation and elongation, respectively, of viral transcription (Lassen et al., 2004; Williams and Greene, 2007). The LR may thus be established when activated CD4⁺ T cells become infected as they revert back to a resting memory state. In addition, DNA methylation and repressive histone modifications may silence proviruses (Blazkova et al., 2009; Coull et al., 2000; He and Margolis, 2002; Kauder et al., 2009; Van Lint et al., 1996; Verdin et al., 1993; Williams et al., 2006).

A major approach to eradicating HIV-1 involves reversing latency in patients on ART (Richman et al., 2009; Deeks, 2012). Cells harboring induced proviruses could then be lysed by HIV-1-specific cytolytic T lymphocytes (CTL) (Shan et al., 2012), while new rounds of infection are blocked by ART. Clinical trials exploring this strategy have used the histone deacetylase inhibitors (Lehrman et al., 2005; Archin et al., 2009, 2012; Contreras et al., 2009).

Accurate measurement of the LR is essential for evaluating eradication strategies. If the LR is eradicated, ART can be discontinued without rebound viremia. Interruption before complete eradication will likely result in rebound (Davey et al., 1999) and repopulation of the LR.

The standard assay for LR size is a viral outgrowth assay (VOA) (Finzi et al., 1997; Siliciano and Siliciano, 2005) measuring the frequency of resting CD4⁺ T cells that produce infectious virus after a single round of maximum *in vitro* T cell activation. Limiting dilutions of resting CD4⁺ T cells are stimulated with the mitogen phytohemagglutinin (PHA), which reverses latency by inducing T cell activation. Released viruses are expanded by addition of CD4⁺ T lymphoblasts from HIV-1-negative donors. Culture supernatants are examined for exponential viral growth by ELISA for HIV-1 p24. With this assay, the mean frequency of latently infected cells in patients on ART is $\sim 1/10^6$ resting CD4⁺ T cells (Eriksson et al., 2013).

It has been assumed that LR size can be assessed with agents like PHA that induce uniform T cell activation (Patel et al., 1988; Hermankova et al., 2003). However, the frequency of latently infected cells detected in the VOA is 300-fold lower than the frequency of resting CD4⁺ T cells that harbor proviruses detectable

by PCR (Eriksson et al., 2013). Thus, at limiting dilution in the VOA, negative wells contain many proviruses, which we designate noninduced proviruses. The noninduced proviruses are generally considered defective but have not been molecularly characterized. The magnitude of the challenge presented by the LR depends on whether noninduced proviruses can be induced in vivo. We present here a molecular characterization of noninduced proviruses.

RESULTS

Transwell VOA Achieves Maximum In Vitro Activation and Outgrowth

To analyze proviruses that did not give rise to infectious virus in the VOA (noninduced proviruses), we first established that the conditions were sufficient to activate 100% of resting CD4⁺ T cells. Resting CD4⁺ T cells from patients on suppressive ART for >6 months were labeled with carboxyfluorescein succinimidyl ester (CFSE) and stimulated with PHA and irradiated allogeneic peripheral blood mononuclear cells (PBMC) under conditions used in the VOA. By day 7, >99.8% of patient cells had divided at least once (Figure S1A available online), confirming that PHA causes uniform T cell activation.

In the VOA, viruses released after reversal of latency replicate in healthy donor CD4⁺ lymphoblasts added to the cultures. To facilitate isolation of noninduced proviruses, we tested whether comparable levels of activation and viral outgrowth could be achieved in transwell cultures in which patient cells were separated from donor lymphoblasts by a cell-impermeable membrane (Figure S1B). In side-by-side comparison with standard VOA cultures from ten patients, transwell cultures showed comparable cellular activation in both p24⁺ and p24⁻ wells, as >95% of patient cells expressed human leukocyte antigen-DR (HLA-DR) and/or CD25 on day 21 (Figure S1C). Transwell cultures also showed viral outgrowth comparable to standard VOA cultures (Figure S1D). Noninduced proviruses were thus isolated from p24⁻ wells of limiting dilution transwell and standard cultures.

Clonal Amplification and Sequencing of Noninduced Proviruses

We obtained near full-length clonal sequences of noninduced proviruses from eight patients on suppressive ART. Patient characteristics are in Table S1. Noninduced proviruses were obtained from wells seeded with 4×10^4 or 2×10^5 resting CD4⁺ T cells that were p24⁻ on day 21. In clonal VOA cultures, wells with replicating virus are p24⁺ by days 10–14 (Laird et al., 2013). Even with a more sensitive real-time (RT)-PCR assay for HIV-1 mRNA (Laird et al., 2013), none of the p24⁻ wells showed exponential growth. Thus, the noninduced proviruses were obtained from wells with no replicating virus, despite maximal T cell activation.

Noninduced proviruses were amplified in limiting dilution PCRs to avoid in vitro recombination. A near-full-length 9.1 kb outer PCR (Li et al., 2007) spanning U5 to U5 (positions 623–9,686, HXB2 coordinates) was followed by nested inner PCRs (Table S2). Aliquots from outer PCRs were first amplified in a nested inner *gag* PCR (Figure 1A). Cell dilutions for which <20%

of the inner reactions were positive were selected because positive wells at these dilutions have >90% probability of being clonal. Aliquots from the *gag*⁺ clonal outer PCRs were amplified using four sets of inner PCR primers to obtain fragments overlapping by 150–3,173 base pairs (bp) (Figure 1A). Of note, instead of cloning PCR products, we directly sequenced them. This dramatically reduces PCR errors, because errors occurring after the first or second cycle are present in too small a fraction of the final products to be observed. Sequences with double peaks or nonidentical overlap regions were discarded. We identified 213 noninduced proviruses from eight patients.

Hypermutation and Large Internal Deletions Render Most Noninduced Proviruses Defective

Most (88.3%) noninduced proviruses had obvious defects precluding replication (Figure 1C). Direct sequencing of the nested *gag* PCR product revealed that ~1/3 (32.4%) of noninduced proviruses had APOBEC3G-mediated G → A hypermutation occurring in the expected sequence context (GG or GGG) (Yu et al., 2004). Hypermuted proviruses are replication defective due to start codon mutations and numerous tryptophan → stop codon mutations (Figure S2). Although the *gag* gene was analyzed here, other regions of the genome show even greater hypermutation (Yu et al., 2004). Of note, it is unlikely that hypermutated proviruses could produce functional viral proteins due to stop codons in most open reading frames (ORFs).

Noninduced proviruses that were not hypermutated were further analyzed by nested amplification of four overlapping sub-genomic fragments (Figure 1A). Of the 144 clonal noninduced proviruses without hypermutation, 97 had large internal deletions identified by smaller amplicon size in electrophoresis (Figure 1B). We mapped the deletion junctions in 58 of these clones (Figure 2A). For example, clone 10CB7_48H1 (Figure 1B) gave a smaller amplicon for the nested C reaction, and amplification of fragments A, B, and D failed due to deletion of nucleotides 4,869–9,533. All 58 mapped deletions would affect expression of the essential regulatory proteins Tat and Rev (Figure 2A) because the deletions encompass the *tat* and *rev* exons, the splice sites, and the Rev-responsive element (RRE).

Deletions are not unexpected. HIV-1 is prone to recombination due to pseudodiploidy (two RNA copies per virion with physical proximity for recombination). Frequent template switching events occur during reverse transcription (Simon-Loriere and Holmes, 2011). Switching between short repeats in a single-genome results in deletion of the intervening sequence and one repeat (Temin, 1993). Large deletions have been observed in unfractionated PBMC from viremic patients (Sanchez et al., 1997). Several lines of evidence suggest that these deletions occur in vivo rather than during in vitro analysis. First, deletions were observed following direct sequencing of uncloned PCR products. Second, for a given provirus, the same deletion junctions were observed in different nested PCRs using different primers. Third, short amplicons were not seen in control experiments with plasmids carrying the reference proviral genomes NL4-3 and BaL. Plasmids were mixed, diluted to eight copies per 10^5 human genome equivalents, and amplified under the same conditions. No deletion or recombination was observed.

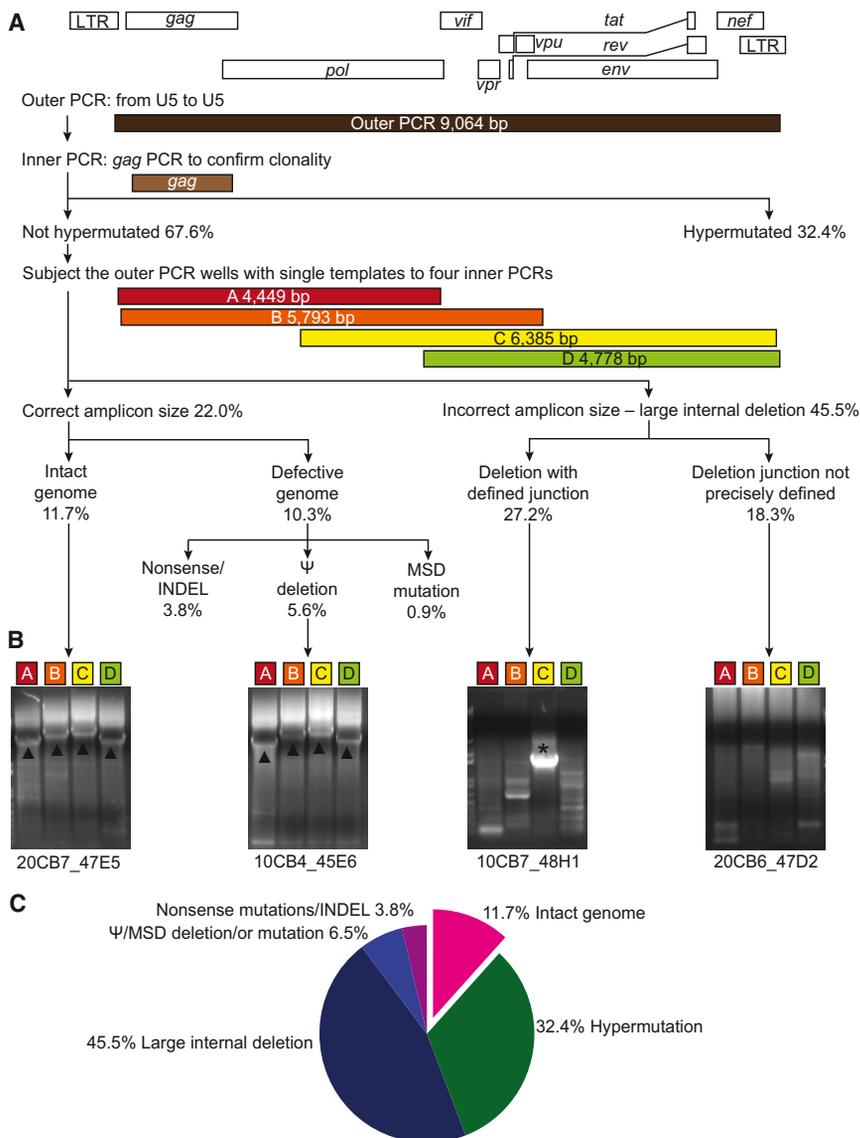


Figure 1. Characterization of Noninduced Proviruses

(A) Strategy and results of analysis of noninduced proviruses. Limiting dilution VOA cultures were established from eight patients. A total of 51 p24⁻ wells containing a total of 8.9×10^6 resting CD4⁺ T cells were analyzed. Two hundred thirteen noninduced proviruses were identified by near-full-length limiting dilution PCR followed by a nested *gag* PCR. The *gag* PCR products were directly sequenced, and proviruses with APOBEC3G-mediated G \rightarrow A hypermutation were identified using the Los Alamos Hypermut algorithm (Rose and Korber, 2000). Nonhypermutated proviruses were analyzed by nested amplification of four overlapping subgenomic fragments (labeled A–D). Amplicons with patient HIV-1 sequence but a size discernibly smaller than expected in 1% agarose gel electrophoresis were considered to contain large internal deletions that, where possible, were mapped by direct sequencing. Other lethal defects including single-nucleotide insertions and deletions (INDELs), packaging signal (Ψ) deletions, and major splice donor (MSD) site mutations were identified by direct sequencing.

(B) Results of 1% agarose gel electrophoresis of four subgenomic amplicons (A–D) from representative noninduced proviruses with different defects. Left to right: intact genome, 16 bp Ψ deletion, deletion with defined junction, and deletion with undefined junction. Clone name is given below each gel. Triangles indicate amplicons with correct size. Asterisks indicate amplicons with a size smaller than expected.

(C) Summary of the analysis of 213 noninduced proviruses.

See also Figure S1 and Table S1.

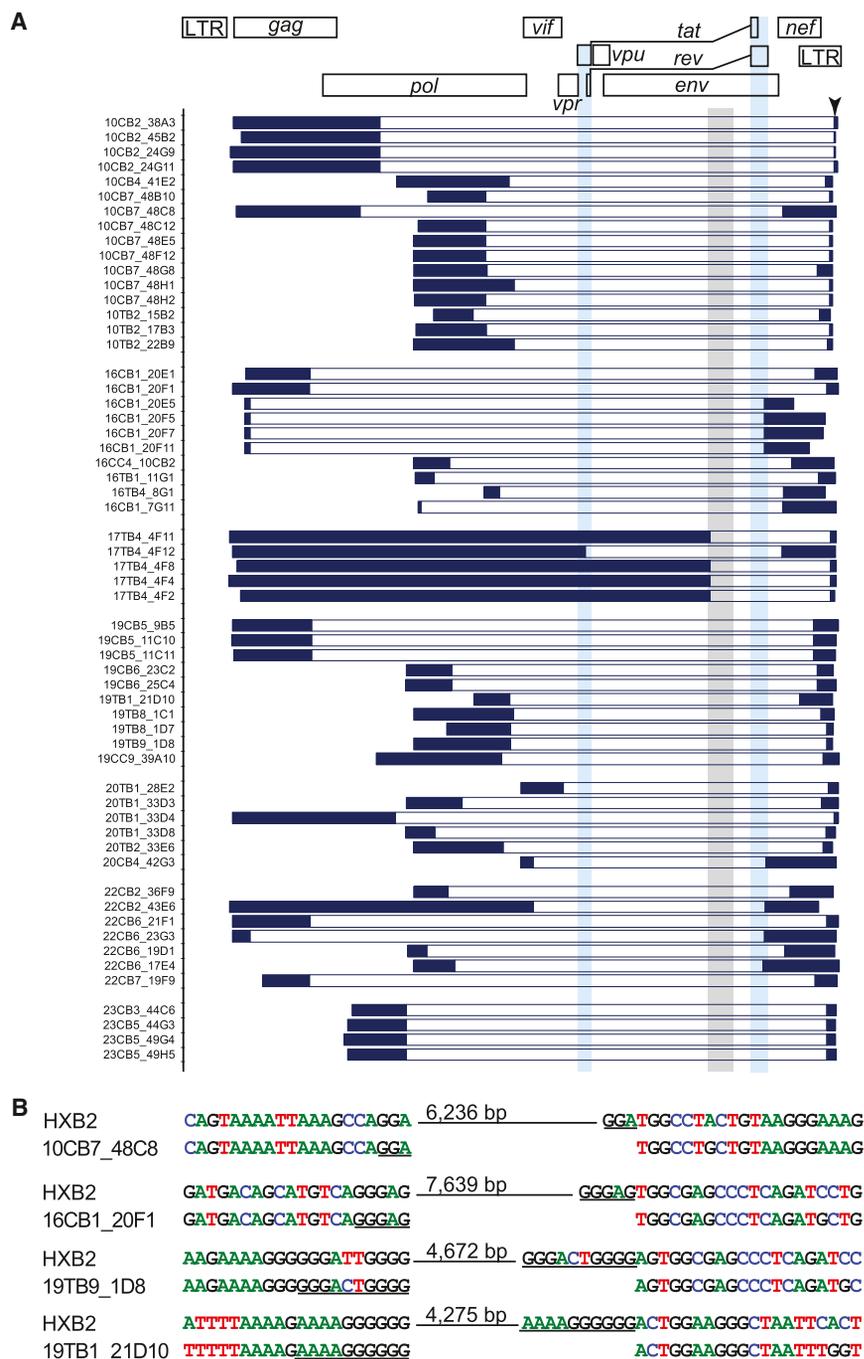
Mutations in *cis* Elements Render Some Noninduced Proviruses Defective

Provirus with correct amplicon size were directly sequenced. A small fraction (8/213) had nonsense mutations and/or

Fourth, short sequence repeats were identified at some deletion junctions (Figure 2B), consistent with a single polymerase jump due to copy choice recombination during reverse transcription of the minus strand (Sanchez et al., 1997). Taken together, these results demonstrate that a large fraction of noninduced proviruses are nonfunctional due to large internal deletions, likely introduced during reverse transcription.

The precise fraction of proviruses with deletions could be underestimated by this analysis because deletions could affect PCR primer binding sites. For 39 clones, the exact deletion junction could not be identified, probably because the deletions encompassed binding sites for primers used in the nested PCRs. For these clones, we obtained at least two sequences to ensure that the amplicons contained nonhypermutated, patient-specific sequences. For 12 mapped deletions, the deletion included the reverse *gag* primer binding site, but mapping was possible because other nested reactions were successful.

frame-shifting insertions or deletions in one or more ORFs. Small deletions (8–98 bp) were found in the packaging signal (Ψ) in 12 sequences (Figure S3A). The deletions encompassed the major splice donor (MSD) site. Point mutations in the MSD site were found in two proviruses. Full genome sequencing showed that seven of the noninduced proviruses with Ψ or MSD mutations were otherwise intact. To determine whether these mutations rendered noninduced proviruses defective, we reconstructed three clones by genome synthesis as described later. The reconstructed proviruses included two clones with short (8 bp and 16 bp) Ψ deletions in packaging stem loop 2 and one with a MSD site mutation (TG|GT \rightarrow TG|GG). Although these clones had intact ORFs, they did not replicate in healthy donor CD4⁺ lymphoblasts (Figure S3B) under conditions in which other reconstructed proviruses replicated well (discussed later). Thus, mutations in *cis* elements render otherwise intact proviruses defective.



A Significant Proportion of Noninduced Proviruses Are Replication Competent

Of 213 noninduced proviruses, 25 (11.7%) had intact ORFs and *cis* elements. When compared with induced proviruses from the same patient, no known lethal mutations were seen. To determine whether these intact noninduced proviruses are replication competent, we used the direct sequencing results to reconstruct full-length noninduced proviral clones by de novo genome synthesis (Figure 3A). This strategy avoids PCR and cloning-induced errors. We reconstructed six noninduced proviruses from four

Figure 2. Mapping of Large Internal Deletions in Noninduced Proviruses

(A) Locations of deletions. Dark blue horizontal bars: continuous sequencing reads interrupted by deletions. White horizontal bars: deletions. Light blue vertical bars, *tat* and *rev* ORFs. Gray vertical bar: RRE element.

(B) Representative sequences at deletion junctions. Short repeats (underlined) at the deletion junctions suggest a single reverse transcriptase jump due to copy choice recombination (Sanchez et al., 1997; Temin, 1993). Hypermutated sequences (see Figure S2) were not analyzed for deletions.

See also Figure S2 and Figure S3 for Ψ deletions.

patients by inserting the synthesized sequence into a plasmid carrying the reference isolate NL4-3. Not captured in our PCR strategy is a 108 bp segment of the provirus, representing nucleotides (nt) 565–672 (HXB2 coordinates). This segment in the 5' untranslated region includes part of U5 and the primer binding site (pbs) (Figure 3A). Although U5 deletions may not affect replication competence (Vicenzi et al., 1994), we took additional steps to make the reconstructed clones fully patient derived, without any NL4-3 sequence. We used limiting dilution PCR to amplify the LTR-*gag* region from cells in $p24^-$ wells. Using a 424 bp segment from the 5' U3-R-U5 region (HXB2 nt 140–564), we constructed phylogenetic trees (Figure S4). Then, using site-directed mutagenesis, we corrected the 108 bp segment from NL4-3 to the phylogenetically closest sequence from the same patient (Figure 3A). This process results in proviruses that are 100% patient derived and 98.2% equivalent to specific proviruses present in vivo, with the remaining 1.8% equivalent to a very closely related provirus from the same patient. Given the high sequence conservation in the 108 bp segment (Figure S4), we estimated that the reconstructed clones could differ from the parent clones by, at most, 3 nt, or 0.03% of the genome. For each patient, we also reconstructed an induced viral clone from a $p24^+$ well.

It is striking that all six reconstructed noninduced proviruses from four different patients showed replication fitness comparable to that of the NL4-3 and reconstructed induced proviruses from the same patients (Figure 3B). It is unlikely that all of these intact noninduced proviruses could have actually been defective with inactivating mutations in the 108 bp segment that was not directly sequenced, as we showed that an additional round of

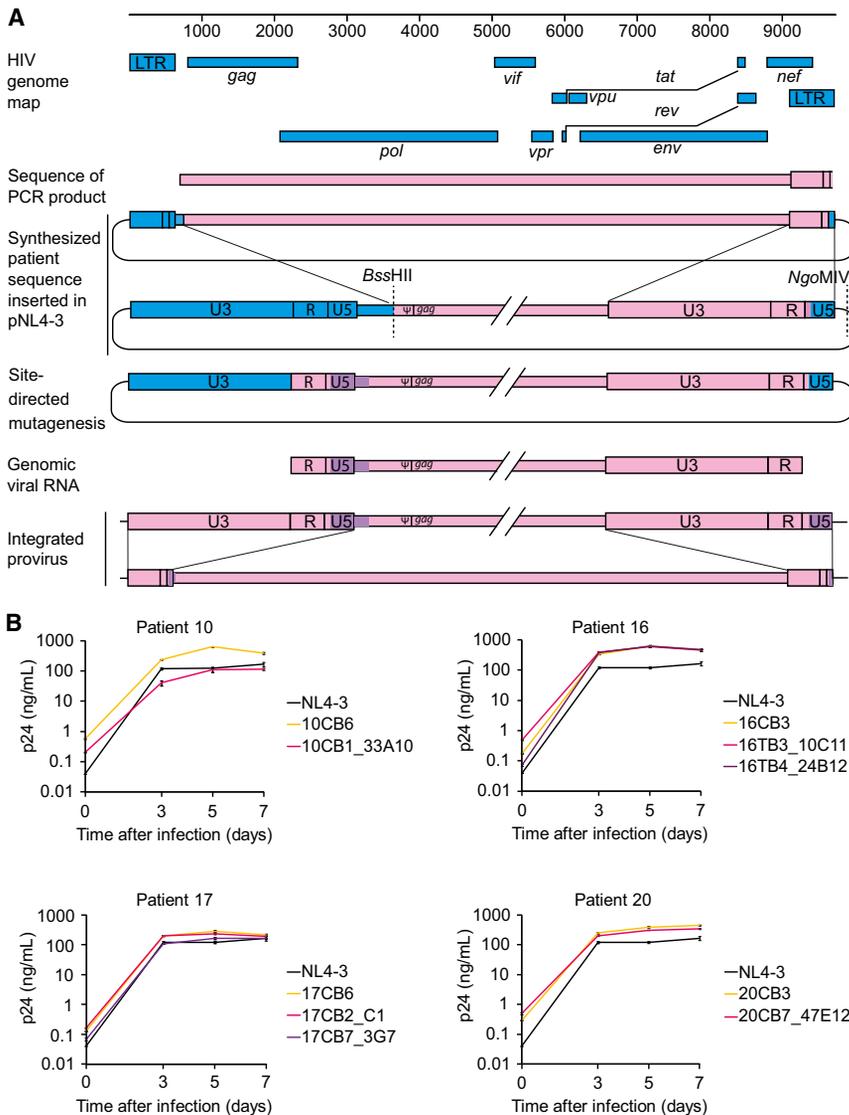


Figure 3. Growth Kinetics of Reconstructed Noninduced Viruses

(A) Reconstruction of full-length noninduced proviruses. Pink indicates clonal patient-derived sequence. Blue indicates pNL4-3. Purple indicates sequence from the most closely related patient-specific virus. (B) Growth kinetics of reconstructed noninduced viruses from p24⁻ wells (pink and purple), reconstructed induced viruses from p24⁺ wells (yellow), and NL4-3 (black). Data are presented as mean ± SEM. See also Figure S4.

lated (Figures 4B and 4D) activity comparable to LTRs from induced proviruses and NL4-3. Decreased LTR function was only observed for hypermutated clones. This likely reflects G → A hypermutation in binding sites for the transcription factors NF-κB and Sp1 (Figure 4E). Thus, most noninduced proviruses have LTRs that are intact at the primary sequence level.

Most Noninduced Proviruses Are Integrated into Active Transcription Units

We next determined the locations of noninduced proviruses in the host genome to understand whether they were integrated into sites unfavorable for transcription. Bushman et al. showed that HIV-1 typically integrates into transcription units (as shown in Schröder et al., 2002). However, in some model systems, integration into regions of heterochromatin is associated with latency (Jordan et al., 2003). Using inverse PCR at limiting dilution,

PHA stimulation causes some noninduced proviruses to produce replication-competent virus (described later). Taken together, these results indicate that a substantial fraction of noninduced proviruses are intact and capable of generating infectious virus if induced in vivo.

Noninduced Proviruses Have Intact Promoter Function unless Hypermutated

The ability of intact noninduced proviruses to produce infectious virus suggests that, at the primary sequence level, their promoters are functional. To confirm this, we cloned long terminal repeats (LTRs) from representative noninduced proviruses into a luciferase reporter construct (Yang et al., 2009). We measured luciferase activity in transfected resting CD4⁺ T cells before and 4 hr after stimulation with phorbol myristate acetate (PMA) and ionomycin. We also analyzed LTRs from induced proviruses from the same patients and NL4-3. In general, LTRs from noninduced proviruses showed basal (Figures 4A and 4C) and stimu-

we found that 92.9% of 70 noninduced proviruses resided in transcription units (Figure 5A), consistent with previous observations in patient resting CD4⁺ T cells (91%) (Han et al., 2004). Based on transcript levels measured in a serial analysis of gene expression (SAGE) library from a primary cell model of latency (Shan et al., 2011), most noninduced proviruses were integrated into genes transcribed at moderate levels in both resting and activated CD4⁺ T cells (Figure 5B). Noninduced proviruses were found in both orientations with respect to the host genes (Figure 5C). Overall, these results indicate that noninduced proviruses are not integrated into chromosomal regions that are repressive for transcription; thus, other factors must have prevented expression.

Lack of CpG Methylation in the LTR of Noninduced Proviruses

We next examined whether noninduced proviruses were silenced by CpG methylation. CpG islands are present in

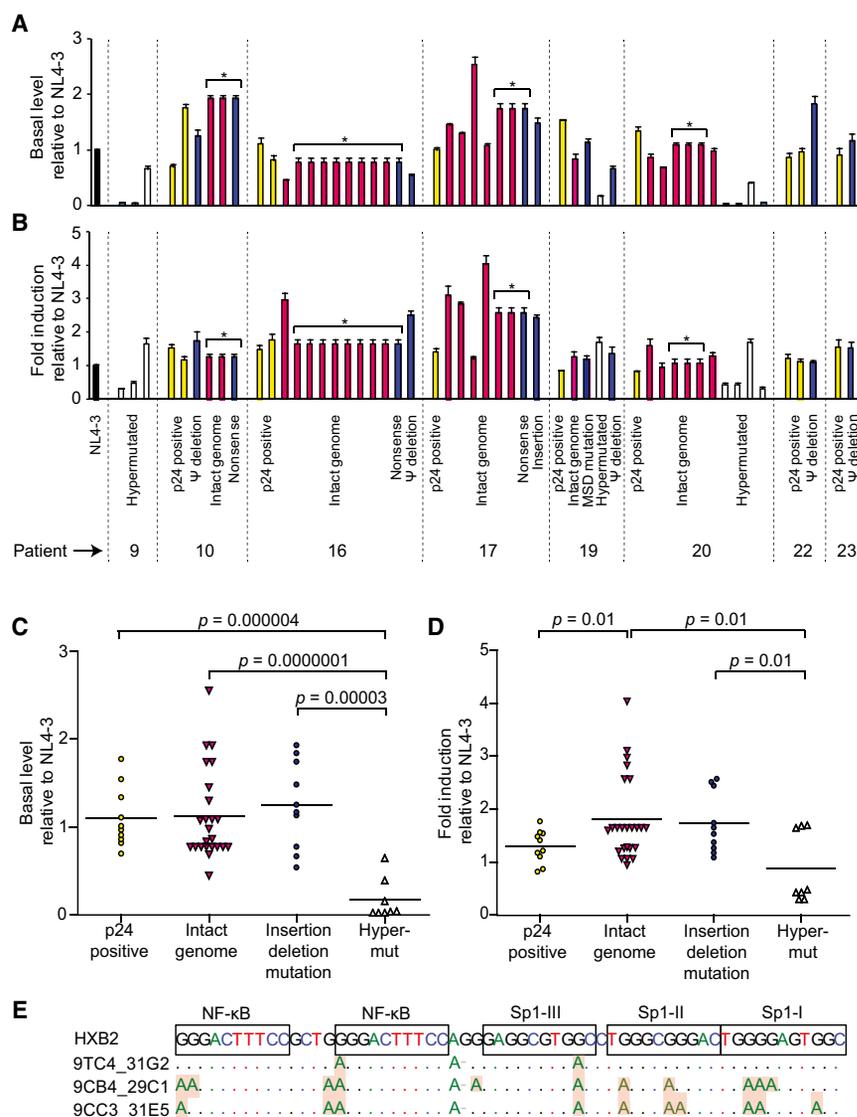


Figure 4. LTR Activity of Noninduced Proviruses

LTR activity measured before, in (A) and (C), and 4 hr after PMA/ionomycin stimulation, in (B) and (D). Resting CD4⁺ T cells were transfected with LTR-firefly luciferase reporter constructs containing LTRs from noninduced proviruses or induced proviruses from p24⁺ wells (yellow bars). Firefly luciferase activity was normalized to an internal control, *Renilla* luciferase driven by a constitutive thymidine kinase promoter. The resulting values were then expressed relative to the NL4-3 control. Noninduced proviruses included those with intact genomes (red bars), insertions or deletions (blue bars), and hypermutation (white bars). Asterisks indicate different clones with the same LTR sequence. Data are presented as mean \pm SEM. (E) Impact of G \rightarrow A hypermutation on transcription factor binding sites. Hypermutated sites in representative clones are shaded.

Intact Noninduced Proviruses May Increase LR Size by \sim 60-Fold

The aforementioned results indicate that although most noninduced proviruses have identifiable lethal defects, a substantial fraction are intact and replication competent at the primary sequence level. Analysis of LTR function, integration sites, and methylation status suggests that these intact noninduced proviruses could be induced in vivo, thereby increasing LR size. We compared the frequency of induced proviruses (defined using the VOA) and intact noninduced proviruses (quantitated as the product of total proviral DNA frequency and the fraction of noninduced proviruses that are intact) among the total pool of proviruses (measured by quantitative PCR).

the HIV-1 genome (Chávez et al., 2011), including one in the U3 region of the LTR, which contains critical transcription factor binding sites (Figure 6). Therefore, DNA from freshly isolated resting CD4⁺ T cells and from cells in p24⁻ wells was treated with bisulfite, and then the LTR region was amplified under limiting dilution conditions (Blazkova et al., 2012). Direct sequencing was used to determine the extent of CpG methylation. Only 3.1% of the LTR CpGs were methylated in resting CD4⁺ T cells from study patients. Even fewer (0.9%) of the CpGs in the LTRs of noninduced proviruses were methylated (Figure 6). In contrast, we readily detected methylation at a CpG island in the *env* region (75.5%), indicating that this method did not selectively amplify nonmethylated sites. Although CpG methylation at the LTR clearly is well documented in some models of HIV-1 latency (Kauder et al., 2009), our results indicate that noninduced proviruses are not silenced through CpG methylation at the 5' LTR.

Bayesian analysis was chosen instead of maximum likelihood estimation because the former provides nonzero point estimates for patients from whom no clones with intact genomes were identified. The positive VOA results in every patient and the successful detection of intact noninduced proviruses in patients for whom >20 clones were analyzed suggest that intact noninduced proviruses could be detected in every patient if enough clones are examined. The fraction of intact noninduced proviruses was calculated as the median of an empirical Bayesian posterior, the most conservative of five models tested (Table S3), with a prior distribution chosen to reflect the observed data. Both the fraction of intact noninduced proviruses and the total number of proviruses per 10⁶ resting CD4⁺ T cells varied dramatically from patient to patient (Figure 7A). There was no correlation between the VOA and the frequency of intact noninduced proviruses (Figure 7B). All statistical models (Table S3) indicated that the median frequency of intact noninduced proviruses was at least \sim 60-fold higher than the frequency of induced proviruses

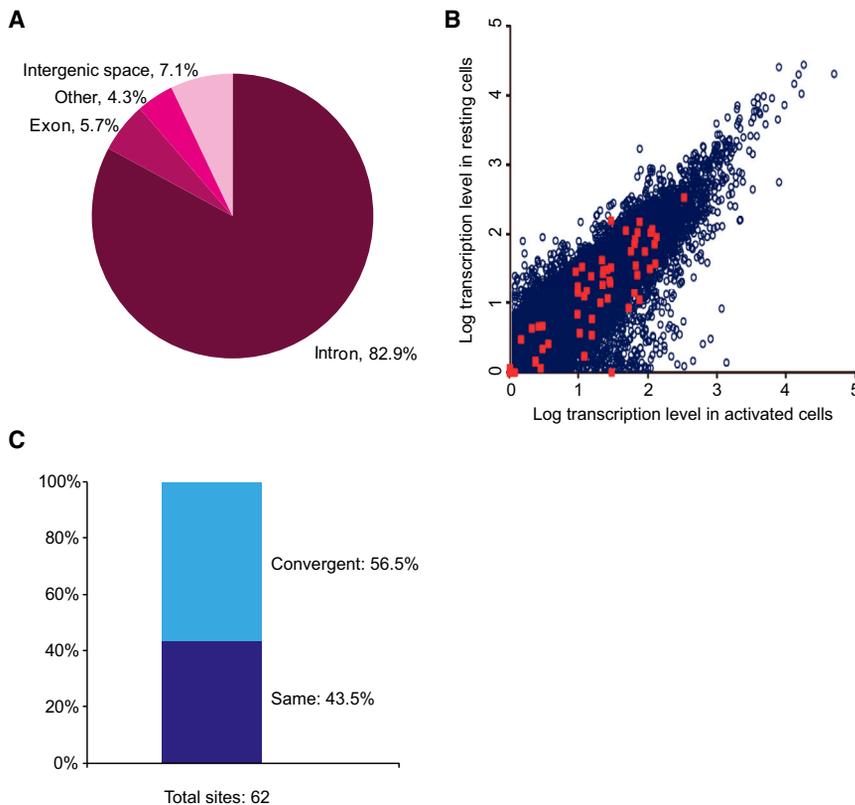


Figure 5. Integration Sites of Noninduced Proviruses

(A) Distribution of integration sites of noninduced proviruses. Intron and exon boundaries are determined as described elsewhere (Shan et al., 2011).

(B) Transcription level of host genes in which noninduced proviruses were integrated. Transcription levels were determined by SAGE analysis of Bcl-2 transduced, activated and resting primary CD4⁺ T cells as described elsewhere (Shan et al., 2011). Blue circles indicate transcript levels in resting and activated CD4⁺ T cells. Red squares indicate transcript levels for the subset of genes in which integration sites of noninduced proviruses were mapped.

(C) Transcriptional orientation of noninduced proviruses relative to the host gene.

detected in the VOA. If the intact noninduced proviruses described here can be induced *in vivo*, then the size of the LR is much greater than previously thought (Figure 7C).

To determine whether intact noninduced proviruses are permanently silenced or potentially inducible under certain conditions, we tested whether repeated PHA stimulation could induce additional noninduced proviruses (Figure S5). We stimulated multiple replicate cultures of 2×10^5 patient resting CD4⁺ T cells with PHA in VOA conditions. We then split each culture well equally into two wells on day 7. As all patient cells have divided by day 7 (Figure S1), each split well contained daughter cells derived from cells activated in the original well. One set of the “split-culture” wells was activated again with PHA, while the other set was cultured without additional stimulation. We then compared supernatant p24 levels after another 14 days of culture. Among 126 p24⁻ wells from four patients, 31 (24.6%) became p24⁺ after the additional round of PHA stimulation, while the paired-culture well that did not receive an additional round of stimulation remained p24⁻ (Figure 7D). It is not yet clear what fraction of the intact noninduced proviruses are inducible *in vivo*. Nevertheless, these results demonstrate that at least some intact noninduced proviruses can be induced under repeated stimulation.

DISCUSSION

The LR in resting CD4⁺ T cells is the major barrier to HIV-1 eradication, and as efforts to cure the infection proceed, accurate

measurement of LR size will be essential. This study provides a molecular basis for understanding measures of the LR. Through an analysis of proviruses that did not give rise to infectious virus following a single round of T cell activation (noninduced proviruses), we have provided a definitive explanation for the large discrepancy between results of PCR and culture assays of LR size. In addition, the identification of intact noninduced

proviruses indicates that the size of the LR may be much greater than previously thought. We reconstructed full-length, intact noninduced proviruses from multiple patients, and all showed growth kinetics comparable to induced proviruses from the same patient and a reference isolate. These intact noninduced proviruses are not detected in standard culture assays but may nevertheless prevent cure. Thus, the present study provides insights into the extent of the challenge posed by the LR and may lead to novel strategies that target intact noninduced proviruses.

We show here that most noninduced proviruses were rendered defective during reverse transcription by APOBEC3G-induced hypermutation (Yu et al., 2004), by internal deletions caused by copy choice recombination during reverse transcription (Sanchez et al., 1997), or by frame-shift or nonsense mutations caused by the error-prone reverse transcriptase (Bebenek et al., 1989). The resulting defective viral genomes can still integrate because only defects at the ends of the genome affect integration. The defective genomes will be detected in most PCR-based assays of proviral DNA, provided that the primer binding sites are intact. Many of the defective proviruses have large internal deletions encompassing the Tat and Rev ORFs and the RRE. Tat-mediated transactivation is required for effective transcriptional elongation (Kao et al., 1987), and the production of virus particles requires that singly spliced and unspliced HIV-1 mRNAs be exported from the nucleus in a Rev-dependent fashion (Malim et al., 1989). Thus, these deleted proviruses may not produce viral proteins, even after successful induction of transcription. The

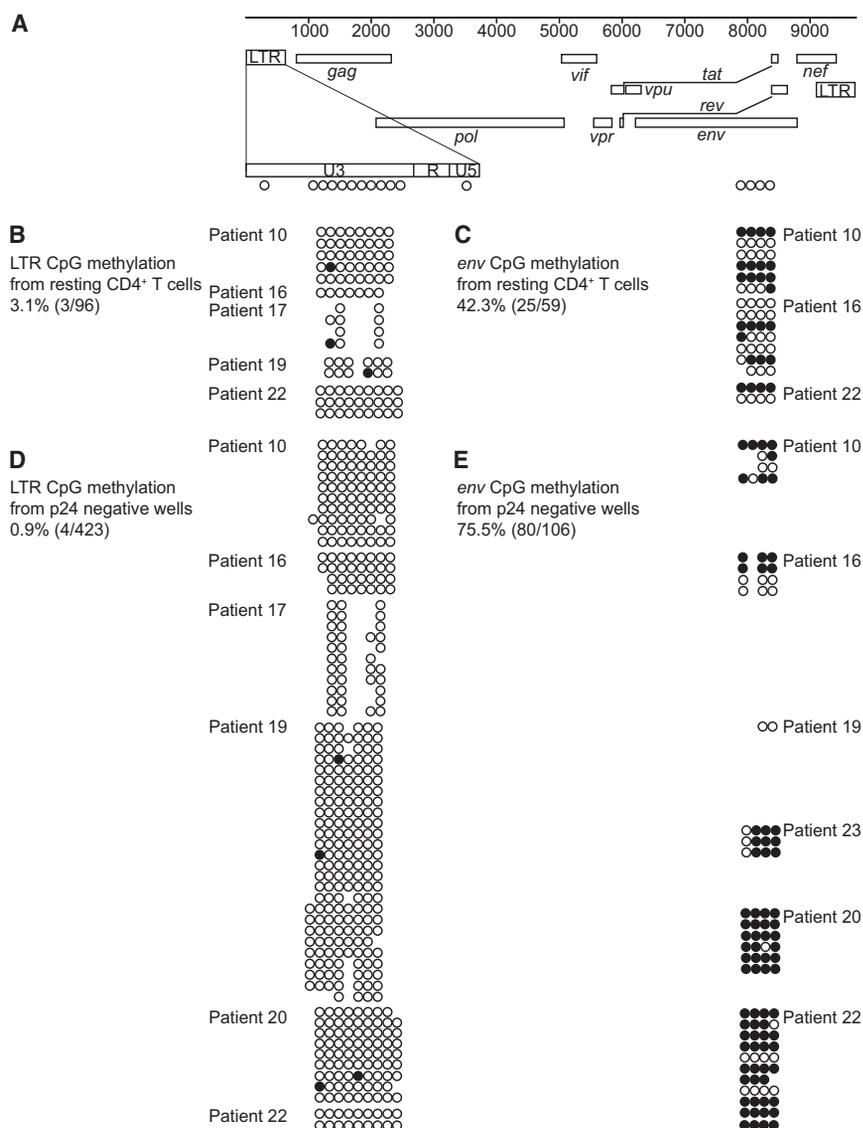


Figure 6. CpG Methylation of Noninduced Proviruses

(A) Positions of CpGs analyzed. (B) CpG methylation in the 5' LTR of proviruses in resting CD4⁺ T cells from the indicated patients. Each row represents a single provirus amplified by limiting dilution PCR after bisulfite treatment. Open circles indicate nonmethylated CpGs. Closed circles indicate methylated CpGs. Missing circles indicate no CpG present due to sequence polymorphism. (C) CpG methylation in the *env* gene of proviruses from resting CD4⁺ T cells. (D) CpG methylation in the 5' LTR of noninduced proviruses from p24⁻ wells. (E) CpG methylation in the *env* gene of noninduced proviruses from p24⁻ wells.

expression. Using prolonged culture and sensitive RT-PCR assays, we also verified that wells from which noninduced proviruses were obtained were truly negative for viral outgrowth. It is also unlikely that these p24⁻ wells remained negative because of reduced viral fitness, as we showed that intact noninduced proviruses had growth kinetics comparable to induced viruses from p24⁺ wells. Taken together, these results confirm that we are examining a population of intact proviruses that were not induced to produce infectious virus after a single round of maximum in vitro activation.

To prove replication competence, we reconstructed six intact, noninduced proviruses from six different p24⁻ wells from four patients. Surprisingly, all reconstructed viruses replicated as well as the standard reference isolate and control viruses reconstructed from p24⁺ wells. A

same is true for hypermutated proviruses, which have stop codons in every ORF. Of note, eradication strategies depend on the production of viral proteins, which allows recognition of the infected cells by HIV-1 specific CTL (Shan et al., 2012). Defective proviruses with large internal deletions and/or APOBEC3G-induced hypermutation may not be eliminated even by strategies that effectively eliminate cells carrying replication-competent virus. These considerations highlight the difficulty of assessing eradication strategies with PCR-based assays.

Although difficult and time consuming, the VOA (Eriksson et al., 2013; Finzi et al., 1997), which has recently been simplified (Laird et al., 2013), does allow detection of cells harboring replication-competent virus. However, the identification of intact, noninduced proviruses raises the possibility that this assay may dramatically underestimate LR size. Several lines of evidence suggest that this is not simply an issue of assay sensitivity. We showed that the PHA stimulation activates all resting CD4⁺ T cells as assessed by proliferation and activation marker

sterilizing cure requires elimination of all replication-competent HIV-1; therefore, the discovery that intact noninduced proviruses are replication competent means that the number of proviruses that must be eliminated is much higher than previously thought. We conservatively estimate that the number may be ~60-fold higher than estimates based on the VOA. Some statistical models suggest an even higher number (medians of 97–273-fold). Of note, there is large interpatient variation in this and other measures of LR size. Overall, our results indicate that the “shock and kill” strategy (Archin et al., 2012; Deeks, 2012) is challenged with a large but unmeasured hidden population of replication-competent proviruses. It is interesting that, despite the intense search for novel latency reversing agents, none of the drugs tested to date reaches the robust level of in vitro HIV-1 induction achieved by PHA. Thus, the finding that the true size of the LR may be ~60-fold greater than that estimated using PHA activation is particularly disturbing. However, it is also important to point out that even a low level of virus gene expression may be

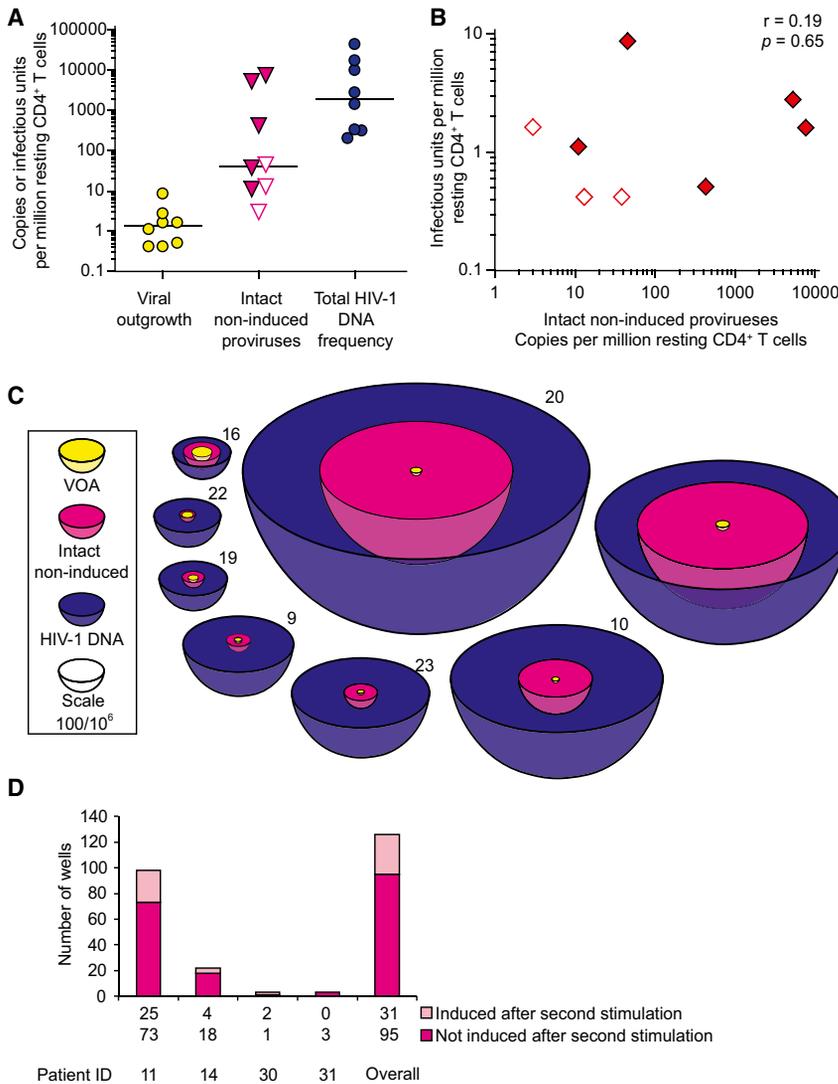


Figure 7. Quantification of Intact Noninduced Proviruses

(A) Comparison of the frequency of cells detected in the VOA, cells carrying intact noninduced proviruses, and cells carrying HIV-1 DNA. The frequency of cells with HIV-1 DNA was measured by quantitative PCR on freshly isolated resting CD4⁺ T cells. The frequency of cells with intact, noninduced proviruses was calculated as frequency of cells with HIV-1 DNA times the proportion of intact noninduced proviruses estimated for each patient using an empirical Bayesian model (Table S3). Open symbols indicate no intact proviruses detected, and empirical Bayesian estimate plotted. Bars indicate median value.

(B) Correlation between the frequency of cells detected in the VOA and the frequency of cells with intact noninduced proviruses. Open symbols, see (A).

(C) Scale representation of the frequencies of the infected resting CD4⁺ T cell populations. Volume reflects population size. Yellow circles indicate minimum size of the LR as measured by VOA. Magenta indicates frequency of cells with intact proviruses. This is the potential size of the LR if intact noninduced proviruses can be induced in vivo. Blue indicates cells with HIV-1 DNA.

(D) Repeated stimulation induces additional proviruses. Bars indicate p24 ELISA results from pairs of split wells cultured with or without a second round of stimulation with PHA and irradiated allogeneic PBMC.

See also Figure S5 and Table S3.

sufficient to allow the elimination of infected cells by an appropriately primed CTL response (Shan et al., 2012) and that the critical variable may be the fraction of latently infected cells induced to express HIV-1 genes.

Understanding why intact noninduced proviruses did not produce infectious virus after maximum in vitro T cell activation is critical for determining their clinical significance. Possible explanations include silencing by repressive chromatin modifications or transcriptional interference. We analyzed CpG methylation of the LTRs of noninduced proviruses at the clonal level. In contrast to some in vitro models of HIV-1 latency (Kauder et al., 2009), we found that in patient CD4⁺ T cells there was little CpG methylation at the LTR, consistent with another recent study (Blazkova et al., 2012). We also examined whether noninduced proviruses are silenced by integration into heterochromatin. We found that most of the noninduced proviruses were integrated into active transcription units, consistent with previous studies showing that most HIV-1 proviruses are integrated into introns of actively transcribed genes in cell lines (Schröder et al., 2002) and patient

resting CD4⁺ T cells (Han et al., 2004). Another potential explanation for the noninduced proviruses is transcriptional interference (Han et al., 2008; Lenasi et al., 2008). Since T cell activation may overcome transcriptional interference due to the high affinity of NF- κ B for its binding sites in the LTR (Lenasi et al., 2008), transcriptional interference may not be a major cause of silencing of the noninduced proviruses.

We propose that, despite maximum T cell activation, the induction of latent proviruses is stochastic. Cellular gene expression levels may follow a digital or analog distribution after T cell receptor activation, as a result of stochastic and dynamic processes (Chakraborty and Das, 2010). Elegant experimental and theoretical studies have shown that HIV-1 proviruses may show stochastic fluctuations in expression depending on levels of Tat (Burnett et al., 2009; Singh et al., 2010; Weinberger et al., 2005; Weinberger et al., 2008). We propose that intact proviruses undergo stochastic induction even after maximum cellular activation. Some will be induced by one round of activation, while others will remain silent but retain the potential to be activated subsequently. These results indicate an increased barrier to cure, as all intact noninduced proviruses need to be eradicated. Underestimation of intact proviruses by VOAs could be

reflected in delayed viral rebound after an apparent “cure,” and overestimation of LR size resulting from detection of defective proviruses by PCR assays could result in prolonged, excessive exposure to toxic latency reversing agents. Thus, the molecular analysis of noninduced proviruses contributes in an important way to HIV-1 eradication efforts.

EXPERIMENTAL PROCEDURES

Study Subjects

Peripheral blood was obtained from healthy volunteers and infected donors who had suppression of viremia to <50 copies HIV-1 RNA/ml for >6 months on ART. This study was approved by the Johns Hopkins Institutional Review Board. Written informed consent was obtained from all participants.

VOA

VOAs were performed as described elsewhere (Finzi et al., 1997; Siliciano and Siliciano, 2005), except that patient cells and donor lymphoblasts were placed in separate chambers in transwell plates. CD8-depleted donor lymphoblasts were added on days 1, 7, and 14. Culture supernatants were examined for p24 by ELISA (PerkinElmer) after 21 days.

Characterization of Full-Length Noninduced Proviruses

Genomic DNA isolated from p24⁻ wells seeded with 4×10^4 or 2×10^5 patient resting CD4⁺ T cells was subjected to limiting dilution prior to amplification with an initial near-full-genome length outer PCR (U5 to U5), followed nested amplification of a segment of the *gag* gene. Aliquots from clonal ($p > 0.9$) positive outer PCR wells were subjected to four inner PCRs to obtain near full-length genome sequence. The procedures are described in [Extended Experimental Procedures](#). PCR conditions are listed in [Table S2](#). PCR products were directly sequenced. Sequences of 213 noninduced clones and 7 induced clones were submitted to GenBank (accession numbers KF526120-KF526339).

Reconstruction of Noninduced Proviruses

The reconstruction of full-length noninduced proviruses is described in [Extended Experimental Procedures](#). Reconstructed plasmids were checked by restriction digestion and sequencing and then transfected into HEK293T cells for virus production. The virus-containing supernatants were adjusted to 200 ng/ml of p24 and used to infect healthy donor CD4⁺ T lymphoblasts for analysis of growth kinetics.

Measurement of LTR Function

Patient-derived LTR sequences were cloned into the wt-LTR-Luc reporter and transfected into resting CD4⁺ T cells from healthy donors using nucleofection (Amaza) (Yang et al., 2009). TK-RLuc was used as internal control. After 48 hr, basal luciferase activity was measured. We also measured luciferase activity 4 hr after stimulation with PMA (10 ng/ml) and ionomycin (1 μ M).

Integration Site, CpG Methylation, and Quantitative RT-PCR

Integration site analysis was carried out by inverse PCR as described elsewhere (Han et al., 2004). Transcription levels of the host genes containing integration sites were determined using a previously described SAGE analysis of Bcl-2-transduced primary resting and activated CD4⁺ T cells (Shan et al., 2011). CpG methylation analysis was carried out as described by Blazkova et al. (2012). Total proviral DNA in resting CD4⁺ T cells was quantified using an RT-PCR assay (Durand et al., 2012). Copy numbers of the human RNaseP gene copies were measured in separate reactions to quantify cell number (TaqMan Copy Number Reference Assay, RNaseP, Invitrogen) (Spivak et al., 2011).

Statistical Analysis

Activation marker expression, viral outgrowth, and LTR function were compared by two-tailed Student's *t* test for independent samples. The correlation between viral outgrowth and intact noninduced provirus frequency was analyzed using MedCalc software. Bayesian inference was used to estimate the proportion of intact noninduced genomes in each patient. An empirical Bayesian prior, among five models ([Table S3](#); [Extended Experimental Proce-](#)

[dures](#)), was chosen to reflect the distribution of the observed data. These proportions were multiplied by total HIV-1 DNA copies per 10^6 resting CD4⁺ T cells to yield an estimate of intact noninduced proviruses per 10^6 cells.

ACCESSION NUMBERS

The GenBank accession numbers for the sequences of 213 noninduced clones and 7 induced clones reported in this paper are KF526120-KF526339.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.09.020>.

ACKNOWLEDGMENTS

We thank all study participants. We thank S. Alireza Rabi, Gregory Laird, and Drs. Stuart Ray and Joel Pomerantz for critical advice; Linda Alston for patient recruitment; David Walker for advice on CpG methylation; and Dr. Gautam Sahu for PCR suggestions. This work was supported by amfAR as part of the amfAR Research Collaboration on HIV Eradication (ARCHE) and also by the Martin Delaney CARE and DARE Collaboratories (National Institutes of Health [NIH] Grants AI096113 and AI096109), by NIH Grant AI043222, by the Johns Hopkins Center for AIDS Research, and by the Howard Hughes Medical Institute. Y.-C.H. is a Howard Hughes Medical Institute International Student Research Fellow.

Received: June 14, 2013

Revised: July 23, 2013

Accepted: August 28, 2013

Published: October 24, 2013

REFERENCES

- Archin, N.M., Espeseth, A., Parker, D., Cheema, M., Hazuda, D., and Margolis, D.M. (2009). Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. *AIDS Res. Hum. Retroviruses* 25, 207–212.
- Archin, N.M., Liberty, A.L., Kashuba, A.D., Choudhary, S.K., Kuruc, J.D., Crooks, A.M., Parker, D.C., Anderson, E.M., Kearney, M.F., Strain, M.C., et al. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* 487, 482–485.
- Bebenek, K., Abbotts, J., Roberts, J.D., Wilson, S.H., and Kunkel, T.A. (1989). Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* 264, 16948–16956.
- Blazkova, J., Trejbalova, K., Gondois-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J., and Hirsch, I. (2009). CpG methylation controls reactivation of HIV from latency. *PLoS Pathog.* 5, e1000554.
- Blazkova, J., Murray, D., Justement, J.S., Funk, E.K., Nelson, A.K., Moir, S., Chun, T.W., and Fauci, A.S. (2012). Paucity of HIV DNA methylation in latently infected, resting CD4⁺ T cells from infected individuals receiving antiretroviral therapy. *J. Virol.* 86, 5390–5392.
- Böhnelein, E., Lowenthal, J.W., Siekevitz, M., Ballard, D.W., Franza, B.R., and Greene, W.C. (1988). The same inducible nuclear proteins regulates mitogen activation of both the interleukin-2 receptor- α gene and type 1 HIV. *Cell* 53, 827–836.
- Burnett, J.C., Miller-Jensen, K., Shah, P.S., Arkin, A.P., and Schaffer, D.V. (2009). Control of stochastic gene expression by host factors at the HIV promoter. *PLoS Pathog.* 5, e1000260.
- Chakraborty, A.K., and Das, J. (2010). Pairing computation with experimentation: a powerful coupling for understanding T cell signalling. *Nat. Rev. Immunol.* 10, 59–71.
- Chávez, L., Kauder, S., and Verdin, E. (2011). In vivo, in vitro, and in silico analysis of methylation of the HIV-1 provirus. *Methods* 53, 47–53.

- Chun, T.W., Stuyver, L., Mizell, S.B., Ehler, L.A., Mican, J.A., Baseler, M., Lloyd, A.L., Nowak, M.A., and Fauci, A.S. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* *94*, 13193–13197.
- Contreras, X., Schwenecker, M., Chen, C.S., McCune, J.M., Deeks, S.G., Martin, J., and Peterlin, B.M. (2009). Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells. *J. Biol. Chem.* *284*, 6782–6789.
- Coull, J.J., Romerio, F., Sun, J.M., Volker, J.L., Galvin, K.M., Davie, J.R., Shi, Y., Hansen, U., and Margolis, D.M. (2000). The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J. Virol.* *74*, 6790–6799.
- Cujec, T.P., Okamoto, H., Fujinaga, K., Meyer, J., Chamberlin, H., Morgan, D.O., and Peterlin, B.M. (1997). The HIV transactivator TAT binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II. *Genes Dev.* *11*, 2645–2657.
- Davey, R.T., Jr., Bhat, N., Yoder, C., Chun, T.W., Metcalf, J.A., Dewar, R., Natarajan, V., Lempicki, R.A., Adelsberger, J.W., Miller, K.D., et al. (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl. Acad. Sci. USA* *96*, 15109–15114.
- Deeks, S.G. (2012). HIV: Shock and kill. *Nature* *487*, 439–440.
- Duh, E.J., Maury, W.J., Folks, T.M., Fauci, A.S., and Rabson, A.B. (1989). Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. *Proc. Natl. Acad. Sci. USA* *86*, 5974–5978.
- Durand, C.M., Ghiaur, G., Siliciano, J.D., Rabi, S.A., Eisele, E.E., Salgado, M., Shan, L., Lai, J.F., Zhang, H., Margolick, J., et al. (2012). HIV-1 DNA is detected in bone marrow populations containing CD4+ T cells but is not found in purified CD34+ hematopoietic progenitor cells in most patients on antiretroviral therapy. *J. Infect. Dis.* *205*, 1014–1018.
- Eriksson, S., Graf, E.H., Dahl, V., Strain, M.C., Yuki, S.A., Lysenko, E.S., Bosch, R.J., Lai, J., Chioma, S., Emad, F., et al. (2013). Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog.* *9*, e1003174.
- Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* *278*, 1295–1300.
- Ganesh, L., Burstein, E., Guha-Niyogi, A., Louder, M.K., Mascola, J.R., Klomp, L.W., Wijmenga, C., Duckett, C.S., and Nabel, G.J. (2003). The gene product Murr1 restricts HIV-1 replication in resting CD4+ lymphocytes. *Nature* *426*, 853–857.
- Han, Y., Lassen, K., Monie, D., Sedaghat, A.R., Shimoji, S., Liu, X., Pierson, T.C., Margolick, J.B., Siliciano, R.F., and Siliciano, J.D. (2004). Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J. Virol.* *78*, 6122–6133.
- Han, Y., Lin, Y.B., An, W., Xu, J., Yang, H.C., O'Connell, K., Dordai, D., Boeke, J.D., Siliciano, J.D., and Siliciano, R.F. (2008). Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe* *4*, 134–146.
- He, G., and Margolis, D.M. (2002). Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. *Mol. Cell. Biol.* *22*, 2965–2973.
- Hermankova, M., Siliciano, J.D., Zhou, Y., Monie, D., Chadwick, K., Margolick, J.B., Quinn, T.C., and Siliciano, R.F. (2003). Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes in vivo. *J. Virol.* *77*, 7383–7392.
- Herrmann, C.H., and Rice, A.P. (1995). Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J. Virol.* *69*, 1612–1620.
- Jones, K.A., and Peterlin, B.M. (1994). Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* *63*, 717–743.
- Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* *22*, 1868–1877.
- Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* *330*, 489–493.
- Kauder, S.E., Bosque, A., Lindqvist, A., Planelles, V., and Verdin, E. (2009). Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog.* *5*, e1000495.
- Kinoshita, S., Su, L., Amano, M., Timmerman, L.A., Kaneshima, H., and Nolan, G.P. (1997). The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. *Immunity* *6*, 235–244.
- Laird, G.M., Eisele, E.E., Rabi, S.A., Lai, J., Chioma, S., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* *9*, e1003398.
- Lassen, K., Han, Y., Zhou, Y., Siliciano, J., and Siliciano, R.F. (2004). The multifactorial nature of HIV-1 latency. *Trends Mol. Med.* *10*, 525–531.
- Lehrman, G., Hogue, I.B., Palmer, S., Jennings, C., Spina, C.A., Wiegand, A., Landay, A.L., Coombs, R.W., Richman, D.D., Mellors, J.W., et al. (2005). Depletion of latent HIV-1 infection in vivo: a proof-of-concept study. *Lancet* *366*, 549–555.
- Lenasi, T., Contreras, X., and Peterlin, B.M. (2008). Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe* *4*, 123–133.
- Li, B., Gladden, A.D., Altfeld, M., Kaldor, J.M., Cooper, D.A., Kelleher, A.D., and Allen, T.M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J. Virol.* *81*, 193–201.
- Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* *338*, 254–257.
- Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* *326*, 711–713.
- Patel, S.S., Duby, A.D., Thiele, D.L., and Lipsky, P.E. (1988). Phenotypic and functional characterization of human T cell clones. *J. Immunol.* *141*, 3726–3736.
- Richman, D.D., Margolis, D.M., Delaney, M., Greene, W.C., Hazuda, D., and Pomerantz, R.J. (2009). The challenge of finding a cure for HIV infection. *Science* *323*, 1304–1307.
- Rose, P.P., and Korber, B.T. (2000). Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics* *16*, 400–401.
- Sanchez, G., Xu, X., Chermann, J.C., and Hirsch, I. (1997). Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. *J. Virol.* *71*, 2233–2240.
- Schröder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* *110*, 521–529.
- Selby, M.J., and Peterlin, B.M. (1990). Trans-activation by HIV-1 Tat via a heterologous RNA binding protein. *Cell* *62*, 769–776.
- Shan, L., Yang, H.C., Rabi, S.A., Bravo, H.C., Shroff, N.S., Irizarry, R.A., Zhang, H., Margolick, J.B., Siliciano, J.D., and Siliciano, R.F. (2011). Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J. Virol.* *85*, 5384–5393.
- Shan, L., Deng, K., Shroff, N.S., Durand, C.M., Rabi, S.A., Yang, H.C., Zhang, H., Margolick, J.B., Blankson, J.N., and Siliciano, R.F. (2012). Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity* *36*, 491–501.
- Siliciano, J.D., Kajdas, J., Finzi, D., Quinn, T.C., Chadwick, K., Margolick, J.B., Kovacs, C., Gange, S.J., and Siliciano, R.F. (2003). Long-term follow-up

- studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med.* **9**, 727–728.
- Siliciano, J.D., and Siliciano, R.F. (2005). Enhanced culture assay for detection and quantitation of latently infected, resting CD4+ T-cells carrying replication-competent virus in HIV-1-infected individuals. *Methods Mol. Biol.* **304**, 3–15.
- Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626.
- Singh, A., Razoooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010). Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys. J.* **98**, L32–L34.
- Spivak, A.M., Salgado, M., Rabi, S.A., O'Connell, K.A., and Blankson, J.N. (2011). Circulating monocytes are not a major reservoir of HIV-1 in elite suppressors. *J. Virol.* **85**, 10399–10403.
- Strain, M.C., Günthard, H.F., Havlir, D.V., Ignacio, C.C., Smith, D.M., Leigh-Brown, A.J., Macaranas, T.R., Lam, R.Y., Daly, O.A., Fischer, M., et al. (2003). Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence. *Proc. Natl. Acad. Sci. USA* **100**, 4819–4824.
- Temin, H.M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl. Acad. Sci. USA* **90**, 6900–6903.
- Tyagi, M., Pearson, R.J., and Karn, J. (2010). Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. *J. Virol.* **84**, 6425–6437.
- Van Lint, C., Emiliani, S., Ott, M., and Verdin, E. (1996). Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J.* **15**, 1112–1120.
- Verdin, E., Paras, P., Jr., and Van Lint, C. (1993). Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J.* **12**, 3249–3259.
- Vicenzi, E., Dimitrov, D.S., Engelman, A., Migone, T.S., Purcell, D.F., Leonard, J., Englund, G., and Martin, M.A. (1994). An integration-defective U5 deletion mutant of human immunodeficiency virus type 1 reverts by eliminating additional long terminal repeat sequences. *J. Virol.* **68**, 7879–7890.
- Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182.
- Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat. Genet.* **40**, 466–470.
- West, M.J., Lowe, A.D., and Karn, J. (2001). Activation of human immunodeficiency virus transcription in T cells revisited: NF-kappaB p65 stimulates transcriptional elongation. *J. Virol.* **75**, 8524–8537.
- Williams, S.A., and Greene, W.C. (2007). Regulation of HIV-1 latency by T-cell activation. *Cytokine* **39**, 63–74.
- Williams, S.A., Chen, L.F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E., and Greene, W.C. (2006). NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.* **25**, 139–149.
- Wong, J.K., Hezareh, M., Günthard, H.F., Havlir, D.V., Ignacio, C.C., Spina, C.A., and Richman, D.D. (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295.
- Yang, H.C., Shen, L., Siliciano, R.F., and Pomerantz, J.L. (2009). Isolation of a cellular factor that can reactivate latent HIV-1 without T cell activation. *Proc. Natl. Acad. Sci. USA* **106**, 6321–6326.
- Yu, Q., König, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **11**, 435–442.

EXTENDED EXPERIMENTAL PROCEDURES

Viral Outgrowth Assay

Viral outgrowth assays (VOAs) were performed as described (Finzi et al., 1997; Siliciano and Siliciano, 2005). Briefly, resting CD4⁺ T cells were isolated from aviremic patients using magnetic microbead selection (Miltenyi Biotec). Standard and transwell cultures were performed side by side using the same population of patient resting CD4⁺ T cells, irradiated allogeneic healthy donor PBMC and CD8-depleted healthy donor lymphoblasts. The number of cells per well and the volume of culture medium were equal for these two culture conditions. In transwell cultures, patient resting CD4⁺ T cells and irradiated PBMC from healthy donors were incubated at the upper chamber on a cell-impermeable polyester membrane with 0.4 μ M pores. The irradiated PBMCs are necessary for optimal PHA activation but die during the culture period, leaving the upper chamber to contain a pure population of patient CD4⁺ T cells. CD8-depleted allogeneic lymphoblasts were added to the lower chamber of the transwell culture in the same numbers as for standard cultures.

Characterization and Reconstruction of Noninduced Proviruses

Genomic DNA was extracted from cells of p24 negative wells using Gentra Puregene Cell Kit (QIAGEN). Aliquots were serially diluted and distributed into 96 well plates. Each well was filled up to a final volume of 50 μ l with PCR reaction mixture solutions including outer PCR primers (BLOuterF and BLOuterR). We chose this outer PCR primer pair as it has been previously used in patient samples for near full-length provirus (Li et al., 2007) and shows superior efficiency in the amplification of near full-length proviruses from patient cells compared with other primer sets (data not shown). Touchdown cycling conditions with decreasing annealing temperature were used to increase specificity (Table S2). Then, 2 μ l aliquots from each outer PCR well were subjected to nested *gag* PCR and 1% agarose gel electrophoresis. If < 20% of the *gag* PCR wells were positive, the corresponding outer PCR wells contained one template with > 90% of possibility, based on the Poisson distribution. The inner PCR primer set described by Li et al. results in a 275 bp missing fragment. We further explored different primer combinations for nested inner PCRs, and using the primers (A, B, C, and D) shown in Figure 1, we were able to obtain sequence for all but 108 base pairs of the provirus. This fragment was reconstructed in separate reactions (see below). From these corresponding positive outer PCR wells of such plates, 2 μ l aliquots were taken for each of four nested PCRs (A, B, C and D). The products were directly sequenced without cloning. Sequences were aligned using CodonCode Sequence Assembly and Alignment Software. The presence of double peaks on sequencing chromatograms was taken as evidence that more than one template may have been present initially, and the sequence was discarded. Direct sequencing results were analyzed by the Los Alamos Hypermut program and HIVAAlign program to identify hypermutation and nonsense mutations.

Direct sequencing results were used for de novo genome synthesis. The resulting sequences were cloned into pNL4-3 vector (GenScript Inc USA and Bio Basic Inc) (see text). Reconstructed plasmids were transformed into Stbl3 chemically competent cells (Invitrogen) and cultured at 30°C to reduce recombination.

To correct the 5'LTR to 100% patient derived, the plasmid fragments between the *Aat*II and *Bss*HIII sites were cloned into Zero Blunt TOPO vector (Invitrogen) and subjected to site directed mutagenesis by QuikChange Site-Directed Multi Kit (Agilent Technologies, Inc).

The reconstructed plasmids were confirmed by restriction digestion and sequencing. Plasmids were transfected to HEK293T cells using Lipofectamine 2000 (Invitrogen). Viruses released into the supernatants were collected by ultracentrifugation 48 – 72 hr post-transfection. Viruses were titered and normalized to 200 ng p24/ml for growth kinetics experiments. Healthy donor CD4⁺ lymphoblasts were infected using spinoculation (1,200 g, 25°C for 2 hr), and growth was monitored by p24 ELISA of culture supernatants.

Estimate the Frequency of Intact Noninduced Proviruses

The number of intact proviral genomes per million cells was estimated for each patient by Bayesian inference. The clones examined were modeled as binomial data, with an unknown probability of being intact. The probability was estimated for each patient based on five different prior distributions, in order to emphasize the effect of the choice of prior distribution on results. The beta distribution is the conjugate prior for a binomial distribution and describes the probabilities of failure and success for a binomial experiment, *i.e.*, the probability of an examined clone being intact versus defective, so all five priors are beta distributions with varying parameters.

The uniform and Jeffreys prior distributions are uninformative priors, chosen to minimize the effect of initial assumptions (Brown et al., 2001). The empirical Bayesian prior distribution was determined by calculating the sample mean and variance of the observed proportions of intact noninduced clones and calculating the beta distribution that fits those parameters. The weighted empirical Bayesian prior has a weighted mean and sample variance, using the number of clones tested for each patient as the weight. The weighting decreases the sample variance and makes the prior stronger. By using the weighting, we acknowledge the consequently strong prior and claim that if we were to double or triple the number of clones tested, we would probably find at least one intact non-induced provirus in every patient. The three patients for whom no intact clones were found are also the three patients in whom the fewest clones were tested. The iterative median approach started with an arbitrary prior, used the posterior median to create an empirical Bayes prior for the next round, and iterated until convergence. (Using the empirical Bayes prior as the initial prior, 40 iterations were sufficient for convergence.)

The empirical Bayesian priors reflect the distribution of the observed data, so these are not fully Bayesian analyses. However, empirical Bayesian analysis can be understood as an approximation of the fully Bayesian hierarchical approach (Gelman et al., 2004).

The median and 95% credible intervals for each approach are reported in [Table S3](#). To estimate the number of cells harboring intact noninduced proviruses per million resting CD4⁺ T cells, we multiplied the median and 95% interval by the number of cells carrying HIV-1 DNA per million resting CD4⁺ T cells.

SUPPLEMENTAL REFERENCES

Brown, L.D., Cai, T.T., and DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Stat. Sci.* 16, 101.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis, Second Edition* (Boca Raton, FL: Chapman and Hall/CRC).

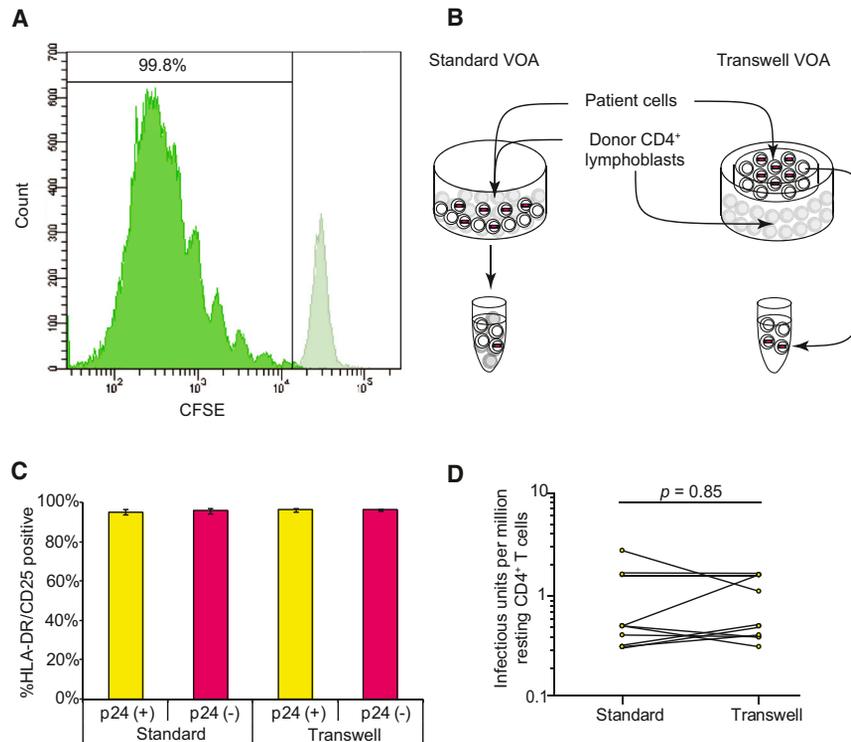


Figure S1. VOA Conditions Achieve Maximum In Vitro Activation and Outgrowth, Related to Figure 1

(A) CFSE fluorescence in labeled patient resting CD4⁺ T cells on day 0 (light green) and day 7 (dark green) of the VOA.

(B) Separation of patient cells from healthy donor CD4⁺ lymphoblasts in a transwell version of the VOA. Patient cells (white) were activated with PHA and irradiated allogeneic PBMC (not shown). To expand the viruses released from infected cells, CD4⁺ T lymphoblasts from healthy donors (gray) were added directly to the culture (standard VOA) or were separated from patient cells by a cell-impermeable polyester membrane with 0.4 μm pores (transwell VOA).

(C) Comparison of activation status of patient cells in standard and transwell VOA cultures, and in p24 ELISA positive and negative wells on day 21. HLA-A2 negative patient cells were cocultured with HLA-A2 positive healthy donor CD4⁺ lymphoblasts and irradiated allogeneic PBMC. The viable HLA-A2 negative population was gated as patient cells for surface activation marker analysis. Data are presented as mean \pm SEM.

(D) Comparison of VOA results by p24 ELISA of supernatants from parallel standard and transwell VOA cultures from 10 patients.



Figure S2. APOBEC3G-Mediated G to A Hypermutation Renders Some Noninduced Proviruses Defective, Related to Figure 2
 (A) Nucleotide sequences of the *gag* gene (representing nucleotides 790 – 2,292, HXB2 coordinates) of hypermutated noninduced proviruses from a representative patient (#20). Sequence changes representing probable G to A hypermutation are shaded. Sequences are aligned to the reference sequence HXB2 and an intact, noninduced provirus from the same patient.
 (B) Amino acid sequences of Gag from the hypermutated noninduced proviruses from (A). Sequence changes resulting from probable G to A hypermutation are shaded. Sequences are aligned to the reference sequence HXB2 and an intact, noninduced provirus from the same patient. Note that the methionine start codon of Gag is frequently mutated to isoleucine (ATG → ATA). This is because the second amino acid of Gag is glycine (GGX), which results in an APOBEC3G recognition site ATGGGX, allowing mutation of the third position of the start codon. Hypermutation also affects many tryptophan codons that contain the APOBEC3G recognition site. Mutation of either or both of the G nucleotides to A generates a stop codon (TGG → TGA, TAG, or TAA, designated with an *).

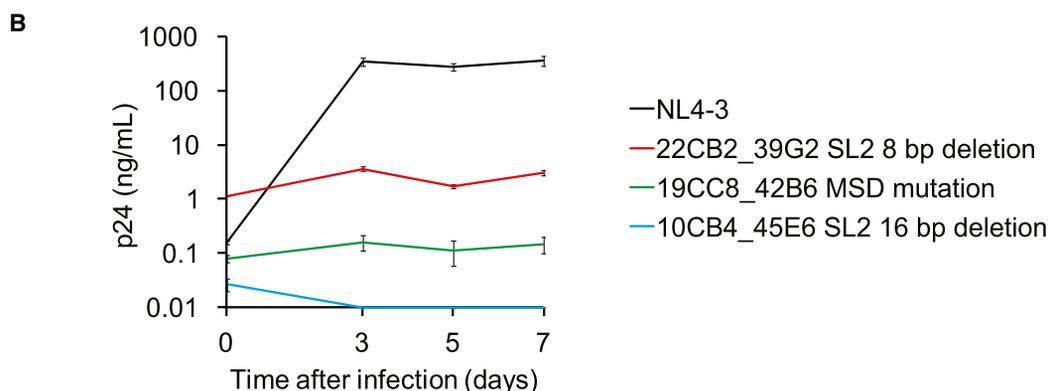
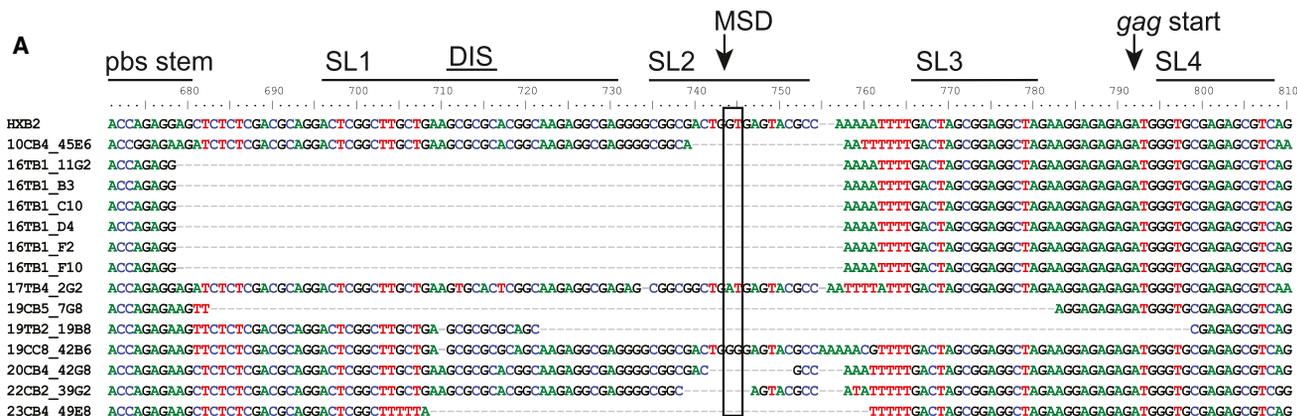


Figure S3. Deletions and Mutations in *cis* Elements, Related to Figure 2

(A) Location of *cis* element deletions and mutations in 14 clones from 5 patients. pbs, tRNA primer binding site; SL, packaging stem loop; DIS, dimerization initiation site; MSD, major splice donor site, typically TG|GT (arrow and box).

(B) Growth kinetics of 3 reconstructed clones from 3 patients with MSD site deletion or mutation but intact ORFs for all viral proteins, in comparison with the NL4-3 reference strain. Each viral inoculum was normalized to 200 ng p24/ml. Data are presented as mean ± SEM.

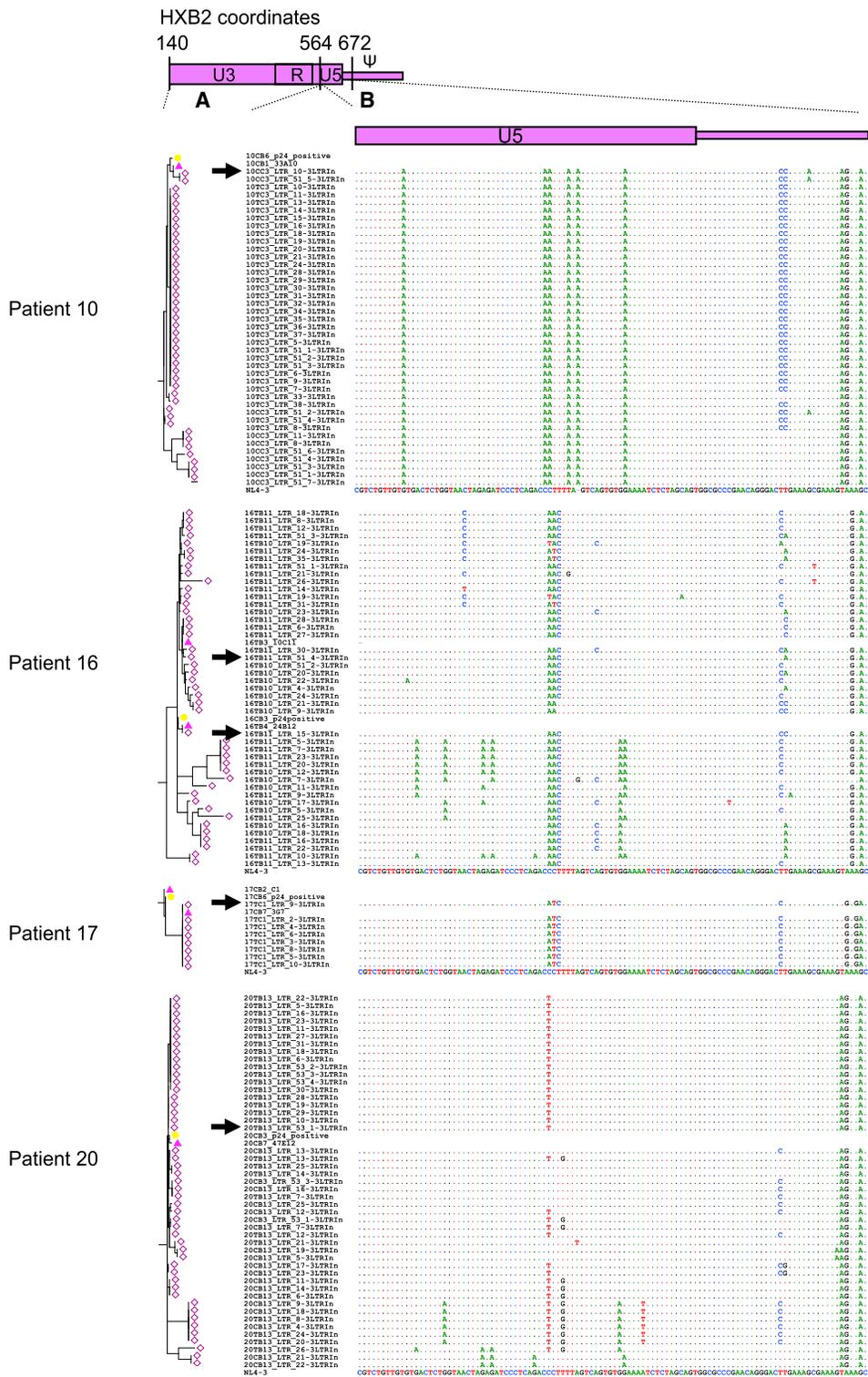


Figure S4. Phylogenetic Analysis for Reconstruction of the 108 bp Sequence, Related to Figure 3
(A) Use of phylogenetic analysis to reconstruct full-length patient-derived proviruses for 6 noninduced proviral clones and 4 induced proviral clones. We corrected the 108 bp segment (565 – 672, HB2 coordinates) missing from the original PCRs (Figure 3A) to the most closely related sequence from the same patient. For this purpose, we used limiting dilution PCR to amplify the LTR-gag region from cells of p24 negative wells. An average of 31 clonal sequences were obtained for each patient. Using a 424 bp segment from the U3-R-U5 region (HB2 positions 140 – 564), we constructed phylogenetic trees and calculated the phylogenetic distance between reconstructed clones and each of these LTR-gag PCR sequences using a maximum likelihood estimate. Pink triangles, intact noninduced

(legend continued on next page)

proviruses to be reconstructed. Yellow circles, induced proviruses from p24 positive wells from the same patient. Open diamonds, noninduced proviral LTR sequences from the same patient. Arrows, exact sequences used for correction, chosen based on smallest pairwise distance in a maximum likelihood phylogenetic analysis using MEGA 5.10 software.

(B) Nucleotide sequences of the 108 bp segment (nt 565 – 672). Nucleotide differences between NL4-3 (bottom sequence) and the patient-derived sequence that was most closely related to the reconstructed clone (arrow) were corrected by site directed mutagenesis in the reconstructed clones (blank sequences).

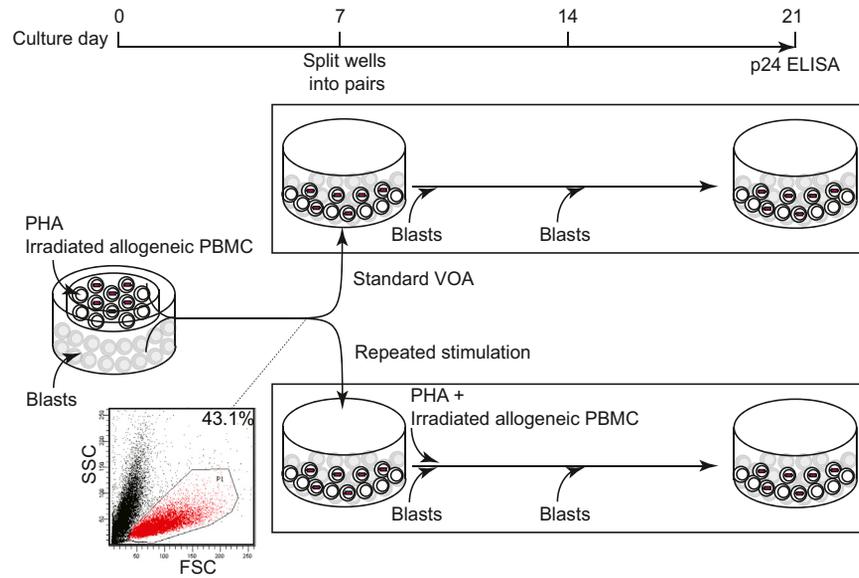


Figure S5. Stochastic Activation of Noninduced Proviruses, Related to Figure 7

Scheme for repeated stimulation. To understand whether noninduced proviruses may be induced by repeated stimuli, resting CD4⁺ T cells from four aviremic donors were plated at 2×10^5 /well and subjected to activation with PHA and irradiated allogeneic PBMC for 1 day. Patient cells were then cultured in a transwell VOA, with patient cells in the upper chamber along with irradiated allogeneic PBMC and donor lymphoblasts in the lower chamber. On day 7, viability of cells in the upper chamber was examined. On day 7, the cells in the upper and lower chambers of each well were combined and then separated equally into duplicate culture wells. As all patient cells had divided at least once by day 7, each pair of split wells should contain an approximately equal number of daughter cells derived from the patient cells in the original well. One set of the split culture wells was subjected to an additional round of stimulation with PHA and irradiated allogeneic PBMC. The other set of wells was not stimulated. Both sets of wells were cultured for an additional 14 days with two more additions of donor lymphoblasts. Levels of HIV-1 p24 antigen were measured in the supernatant on day 21.