nature medicine

Defective proviruses rapidly accumulate during acute HIV-1 infection

Katherine M Bruner[1], Alexandra J Murray[1], Ross A Pollack[1], Mary G Soliman[1], Sarah B Laskey[1], Adam A Capoferri[1,2], Jun Lai[1], Matthew C Strain[3], Steven M Lada[4], Rebecca Hoh[5], Ya-Chi Ho[1], Douglas D Richman[3,4,6], Steven G Deeks[5], Janet D Siliciano[1] & Robert F Siliciano[1,2]

Although antiretroviral therapy (ART) suppresses viral replication to clinically undetectable levels, human immunodeficiency virus type 1 (HIV-1) persists in CD4+ T cells in a latent form that is not targeted by the immune system or by ART[1–5]. This latent reservoir is a major barrier to curing individuals of HIV-1 infection. Many individuals initiate ART during chronic infection, and in this setting, most proviruses are defective[6]. However, the dynamics of the accumulation and the persistence of defective proviruses during acute HIV-1 infection are largely unknown. Here we show that defective proviruses accumulate rapidly within the first few weeks of infection to make up over 93% of all proviruses, regardless of how early ART is initiated. By using an unbiased method to amplify near-full-length proviral genomes from HIV-1-infected adults treated at different stages of infection, we demonstrate that early initiation of ART limits the size of the reservoir but does not profoundly affect the proviral landscape. This analysis allows us to revise our understanding of the composition of proviral populations and estimate the true reservoir size in individuals who were treated early versus late in infection. Additionally, we demonstrate that common assays for measuring the reservoir do not correlate with reservoir size, as determined by the number of genetically intact proviruses. These findings reveal hurdles that must be overcome to successfully analyze future HIV-1 cure strategies.

HIV-1 establishes infection in CD4+ T cells[1,2,7], generating a latent reservoir with an extremely long half-life (44 months) that necessitates life-long treatment[4,8]. Efforts to eradicate this reservoir include the 'shock-and-kill' approach in which latent proviruses are induced so that infected cells can be eliminated[9–11]. The latent reservoir was initially defined by using a quantitative viral outgrowth assay (QVOA)[1,4,8,12,13]. PCR can also detect proviral DNA[14–16]; however, the QVOA and PCR assays correlate poorly[17]. We recently examined discrepancies between the assays by sequencing proviruses containing the *gag* gene from QVOA wells that were negative for viral outgrowth[6]. Most proviruses were defective, but 12% were intact, and some of these could be reactivated following a second round of T cell activation[6]. However, these studies were limited to individuals who initiated ART late in the course of infection and only analyzed *gag*+ proviruses in cells that had been expanded for weeks in culture[6], which may have altered the proportions of different proviral populations. Because immediate ART is now recommended for all infected individuals[18] and the size (and perhaps other characteristics) of the reservoir differs in individuals who are treated early versus late in infection[19–21], we sought to define the composition of proviral populations in individuals who were treated during acute infection. As cure efforts advance, it is essential to understand how defective proviruses accumulate, persist in individuals and affect reservoir assay measurements. Therefore, we conducted a proviral analysis on unmanipulated samples from HIV-1-infected adults who were treated at different stages of infection.

We developed a novel, unbiased single-genome-amplification method that captures intact and defective proviruses. Because deletions probably occur during minus strand synthesis before the second strand transfer event of reverse transcription, primers were designed to capture deletions arising during this process (**Fig. 1a**)[22–24]. PCRs were performed at limiting dilution to prevent *in vitro* recombination and competition between short and long templates. Using a nested PCR approach, we performed a near-full-length outer PCR followed by inner PCRs for the *gag* and *env* genes to confirm clonality. Next, six overlapping inner PCRs were performed on all of the wells at limiting dilution (**Fig. 1b**). The PCR products were directly sequenced to minimize PCR-induced error.

We examined proviral DNA in purified resting CD4+ T cells from ten subjects who initiated ART during the chronic phase of infection (CP; ART started >180 d after infection) (**Supplementary Table 1**). Analysis was performed on freshly isolated cells to prevent bias from *in vitro* expansion. Notably, 98% of proviruses were defective (**Fig. 1c**,**d** and **Supplementary Fig. 1**). The most common defects were internal deletions (80%), which varied in size (15 bp to ~8 kb) and in location in the genome. Some proviruses had 3′ deletions affecting the *env*, *tat*, *rev* and *nef* genes. We also identified proviruses

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [2]Howard Hughes Medical Institute, Baltimore, Maryland, USA. [3]Department of Medicine, University of California, San Diego (UCSD), San Diego, California, USA. [4]Veteran's Affairs (VA) San Diego Healthcare System, San Diego, California, USA. [5]Department of Medicine, University of California, San Francisco (UCSF), San Francisco, California, USA. [6]Department of Pathology, University of California, San Diego, San Diego, California, USA. Correspondence should be addressed to R.F.S. (rsiliciano@jhmi.edu).

Received 23 April; accepted 28 June; published online 8 August 2016; doi:10.1038/nm.4156

with 5′ deletions affecting the *gag* and *pol* genes, as well as others with very large deletions (>6 kb) that encompassed most of the HIV-1 genes (**Fig. 1c**,**d**). The 5′ deletions and the very large deletions, which represent 40% of all sequences, were not identified in a previous screen because they contain deletions in the *gag* gene[6]. Proviruses with small (15- to 97-bp) deletions at the packaging signal and the major splice donor site represented 5% of sequences (**Fig. 1d**). These proviruses are probably replication defective owing to a failure to correctly make spliced HIV-1 RNAs or to package genomes into virions[6]. Some clones contained multiple deletions, suggesting multiple template-switching events during reverse transcription (**Fig. 1c**), and others contained sequence inversions or insertions (**Fig. 1d**). For each type of deletion, proviruses with sequence homology at the deletion junctions were found (**Fig. 1e**), consistent with the idea of template switching during reverse transcription[25]. In control experiments, plasmids containing reference (NL4-3) or patient-derived proviruses were diluted into human genomic DNA and amplified by the same procedure with no defects observed (data not shown). Additionally, analysis of the deletions showed nonrandom distribution within the HIV-1 genome, with distinct hotspots (**Supplementary Fig. 2**), and we observed identical deletions in independent amplifications from the same subject, reflecting clonal expansion of infected cells (see below). Taken together, these results indicate that the deletions we observed occurred *in vivo* and did not result from PCR recombination. We also identified 7% of proviruses with guanine-to-adenine (G-to-A) hypermutations that were induced by the cytidine deaminases APOBEC3F and APOBEC3G, which are HIV-1 restriction factors. An additional 8% of proviruses contained both deletions

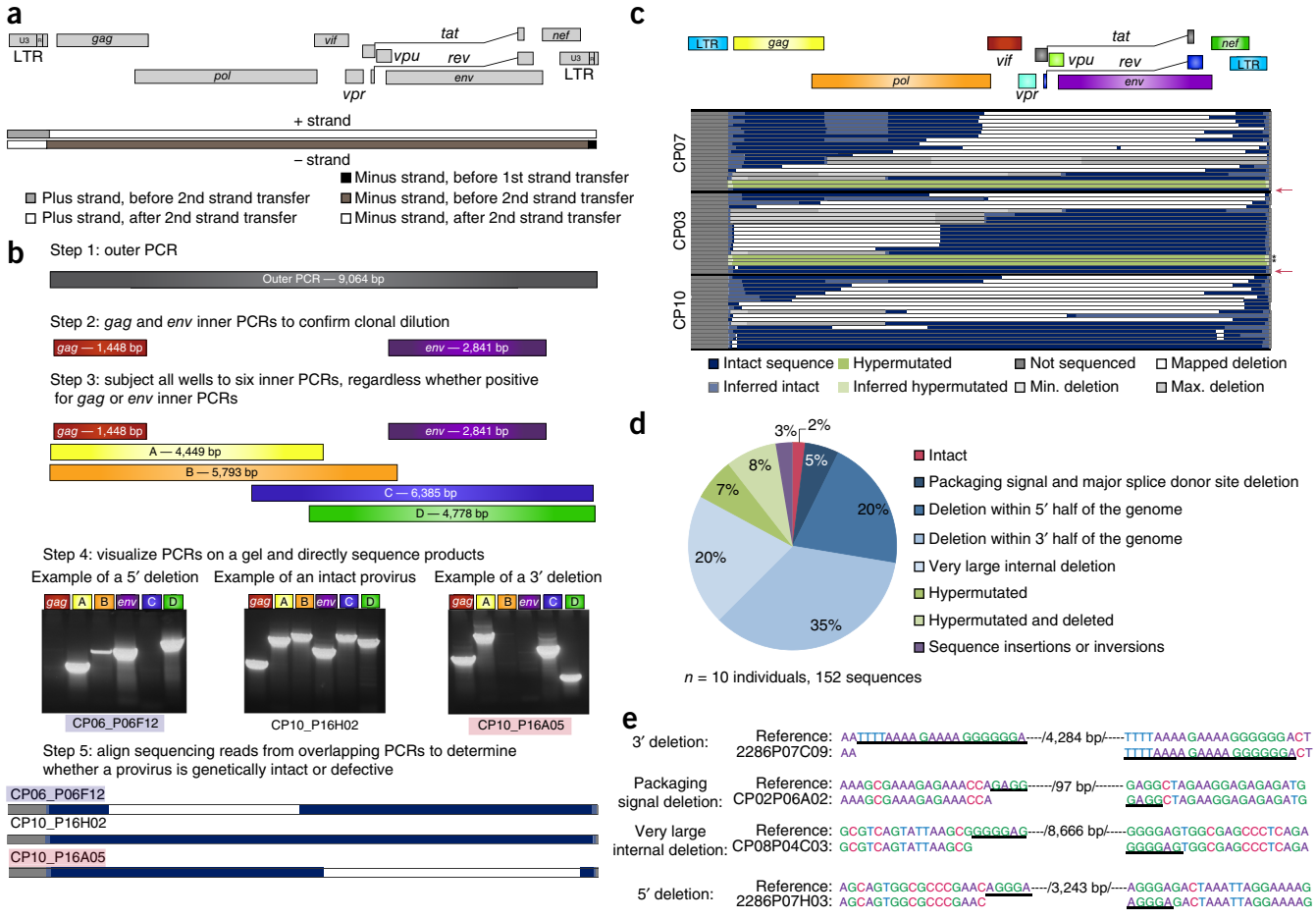

Figure 1 Proviral sequences in chronically treated subjects are highly defective. (**a**) Schematic of strand-transfer events during reverse transcription relative to the HIV-1 genome. Deletions occur during minus strand synthesis before the second strand transfer event by the same copy-choice mechanism that leads to recombination[25]. U3, unique 3′ region; R, repeat region. (**b**) Unbiased full-genome sequencing strategy, based on mechanistic considerations in **a**, to capture the majority of defective proviruses. Bands from representative proviruses with a 5′ deletion (clone CP06_PO6F12), an intact sequence (clone CP10_P16H02) and a 3′ deletion (clone CP10_P16A05) are shown. (**c**) Maps of the HIV-1 genome (top) and proviral clones from three representative subjects treated during the chronic phase of infection (bottom). Each horizontal bar represents an individual clone identified through full-genome sequencing. In cases for which a deletion could not be precisely mapped due to a deletion encompassing multiple forward or reverse primer-binding sites, the possible maximum (max.) and minimum (min.) deletion sizes are plotted. For sequencing data showing a mapped deletion that removes primer-binding sites for other amplicons, the resulting missing sequence was inferred to be present (light blue or green). Pink arrows denote full-length, genetically intact sequences. Black asterisks indicate presumably full-length hypermutated proviruses that were not fully sequenced owing to extreme hypermutation. See **Supplementary Figure 1** for maps of proviruses from seven additional subjects. (**d**) Summary of 152 proviral sequences from ten individuals. (**e**) Proviral sequences showing short repeats (underlined) identified on both ends of the deletion junctions, which are consistent with a copy-choice mechanism of recombination resulting in deletion of the intervening sequence and one homology region (/number of base pairs deleted/).

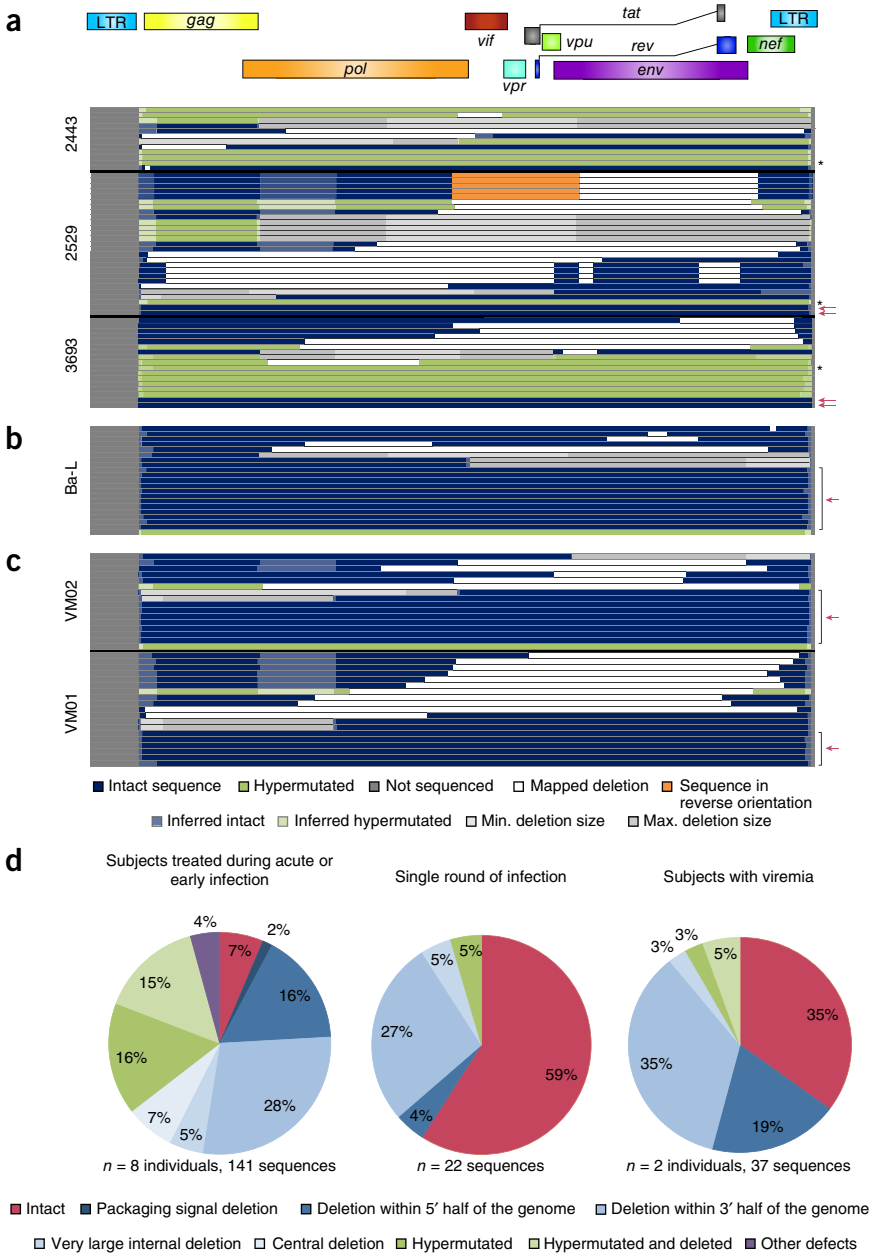

Figure 2 Defective proviruses accumulate rapidly during the course of HIV-1 infection. (**a**–**c**) Maps of independent proviral clones from resting CD4+ T cells of three representative subjects (of nine studied; sequences from additional subjects in **Supplementary Fig. 5**) treated during acute or early infection (**a**), from a single round of *in vitro* infection of healthy donor CD4+ T lymphoblasts with HIV-1 (Ba-L isolate) (**b**), or from CD4+ T cells from two subjects with viremia who are in the chronic phase of infection (**c**). Each horizontal bar represents an individual clone identified through full-genome sequencing. In cases for which a deletion could not be precisely mapped, presumably due to a deletion encompassing multiple forward or reverse primer-binding sites, the maximum (max.) and minimum (min.) deletions sizes are plotted. For sequencing data that shows a mapped deletion removing primer-binding sites for other amplicons, the resulting missing sequence was inferred to be present (light blue or green). A pink arrow denotes full-length, genetically intact sequences. Black asterisks indicate likely full-length hypermutated proviruses that were not fully sequenced owing to extreme hypermutation. (**d**) Summary of all proviral sequences identified from each study population. Subject 2521 was excluded from this analysis because of a single blip during ART.

and hypermutation, indicating that these processes can occur together during reverse transcription (**Fig. 1d**). Almost all of the hypermutated proviruses contained mutated start codons in the *gag*, *gag–pol* and *nef* open-reading frames (ORFs) due to the presence of an obligatory position 2 glycine codon which creates the consensus recognition site for APOBEC3G, ATGGGT[26,27]. Additionally, all of the hypermutated sequences had multiple internal stop codons in the larger ORFs (*gag*, *gag–pol*, *env* and *nef*), and only a small fraction of hypermutated sequences could make functional gene products (**Supplementary Fig. 3**). These defects and common deletions affecting key viral ORFs (**Fig. 1c**,**d**) probably prevent many defective proviruses from being eliminated by eradication strategies that depend on viral protein expression. All of the subjects had undetectable viral loads (<50 copies/ml) for >8 months before sampling, and all but one had low frequencies of 2–long terminal repeat (2-LTR) circles in cells (<25 copies per 10^6 resting CD4+ T cells), which represented less than 8% of the total number of proviruses (**Supplementary Fig. 4a**,**b**). The labile nature of linear unintegrated HIV-1 DNA and the low level of 2-LTR circles suggest that the majority of sequences examined were integrated proviruses[28]. Of 152 near-full-length sequences examined, only three sequences (2%) were intact. These data indicate that defective proviruses are much more common than previously shown (**Fig. 1d**).

Because the dynamics of proviral accumulation remain unclear, we sought to determine how rapidly defective proviruses accumulate. We examined the proviral populations in subjects who were treated during the acute or early phase of infection (AP; ART started within 100 d in all subjects and within 60 d in eight of nine subjects) (**Supplementary Table 1**). Unexpectedly, only 7% of proviruses were intact, even in subjects 2453, 2454, 3693 and 2443, who initiated ART within the first few weeks of infection (**Fig. 2** and **Supplementary Fig. 5**). Although the number of individuals subjected to this extensive analysis was small, the results were very consistent between subjects. The remaining proviral sequences contained defects similar to those in CP-treated subjects (**Fig. 2a**,**d**), and there was no significant difference in the fraction of intact proviruses as compared to that in CP-treated subjects (**Supplementary Fig. 6a**). AP-treated subjects had a higher fraction of hypermutated sequences (35% versus 14%; $P = 0.0036$; **Supplementary Fig. 6b**) and a significantly lower fraction of deleted sequences (57% versus 82%; $P = 0.0028$; **Supplementary Fig. 6c**), suggesting that hypermutation is particularly important during acute infection, perhaps due to upregulation of APOBEC3G and APOBEC3F by type I interferons[29–31]. Notably, defects were even

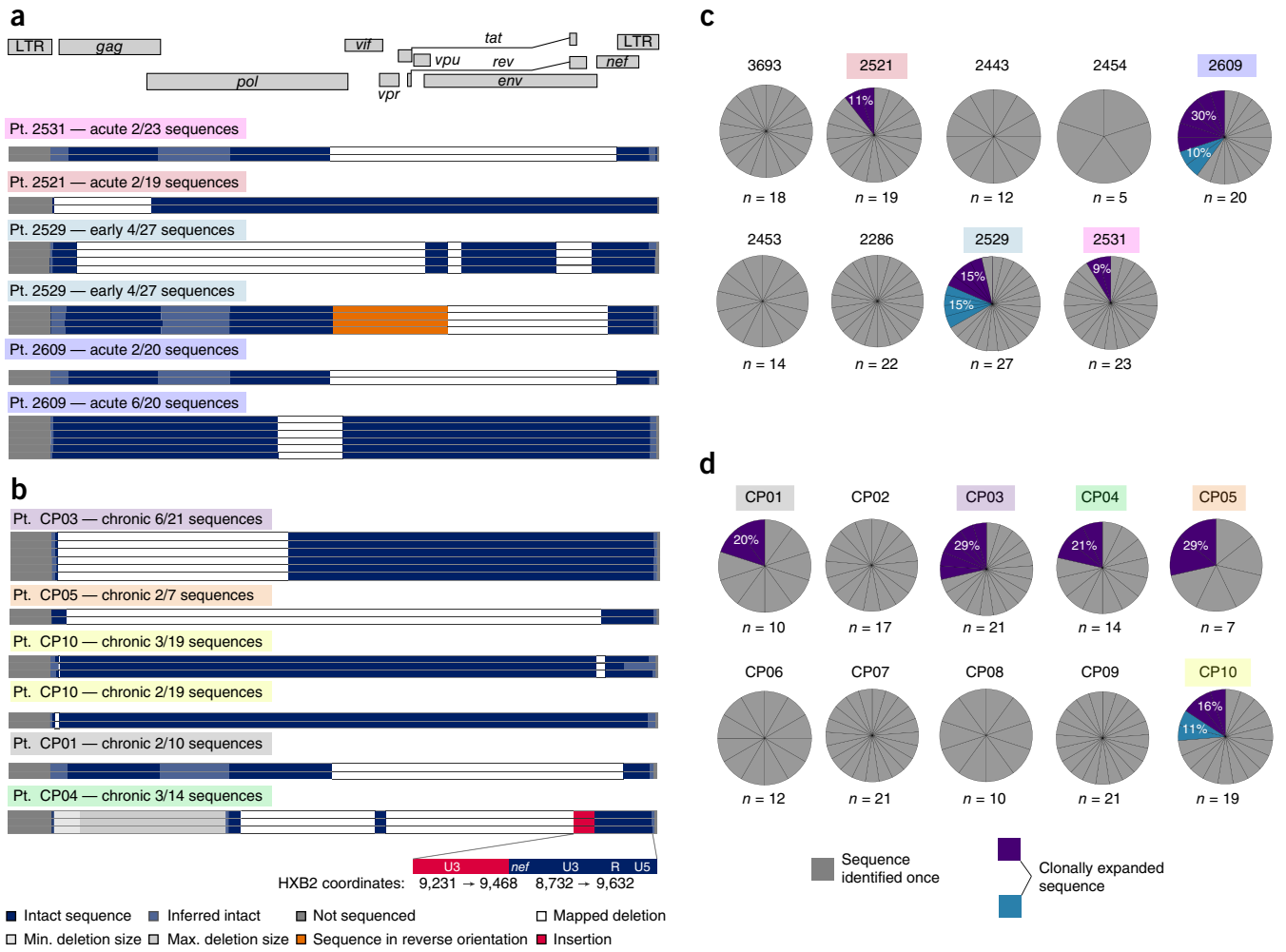

Figure 3 Expanded clones identified in chronically and acutely treated subjects are grossly defective. (**a**,**b**) Maps of expanded HIV-1 clones identified in resting CD4+ T cells from subjects treated during acute or early (**a**) or during chronic (**b**) infection. Expanded clones are defined as clones amplified in completely independent PCR reactions from a single subject that are identical at every nucleotide. The frequency of each clone is shown relative to the total number of clones identified in that subject. Colored boxes denote the subject in which each expanded clone was identified and that corresponds to the subjects indicated in **c**,**d**. Pt., patient. (**c**,**d**) Proportion of sequences that are expanded clones, from subjects treated during acute or early (**c**) or during chronic (**d**) infection. The number of sequences examined for each subject (*n*) is noted. Expanded clones (purple or blue) are shown as a percentage of total sequences from the relevant subject.

detectable in a single round of *in vitro* infection (41% of sequences; **Fig. 2b**,**d**). Consistent with this result, we readily detected defective proviruses in unfractionated CD4+ T cells from subjects with viremia who are in the chronic phase of infection, representing 65% of sequences (**Fig. 2c**,**d**). In these subjects, the provirus populations include both proviruses in newly infected cells and archived proviruses in the reservoir. These results indicate that defective proviruses are probably generated during the initial rounds of replication after transmission, as defective proviruses are generated at a high frequency from the process of reverse transcription (**Fig. 2b**). Because many of the proviruses identified have defects that would preclude high-level expression of viral genes, cells carrying these proviruses would presumably be less susceptible to elimination through viral cytopathic effects or through lysis by cytotoxic T lymphocytes. Thus, even individuals treated very early after infection have large numbers of defective proviruses. Taken together, these results demonstrate that defective proviruses arise commonly, accumulate rapidly within 2–3 weeks after infection and persist *in vivo*.

Proliferation of cells carrying replication-competent proviruses could be a major barrier to eradication of the virus[32–34]. Integration-site analysis, pioneered by Schröder *et al.*[35], has provided critical evidence for the proliferation of infected cells *in vivo* but does not reveal whether the proviruses are replication competent[33,34,36]. Our unbiased analysis allowed us to detect expanded cellular clones *in vivo* and evaluate the presence of defects. Among 312 sequences from nine AP-treated (**Fig. 3a**) and ten CP-treated subjects (**Fig. 3b**), 38 were from expanded cellular clones that were identified to be identical sequences arising from independent amplifications. Notably, all of these expanded clones contained gross defects that preclude replication. In subjects 2529, 2609, CP03, CP05 and CP10, expanded clones represented over 25% of all sequences (**Fig. 3c**,**d**). Subject 2521 contained an expanded clone that represented 11% of all proviral sequences (**Fig. 3a**,**c**) after just 17 months of infection and 8 months on suppressive ART, indicating that expanded clones do not require years to accumulate to large proportions. Our results indicate that the majority of expanded

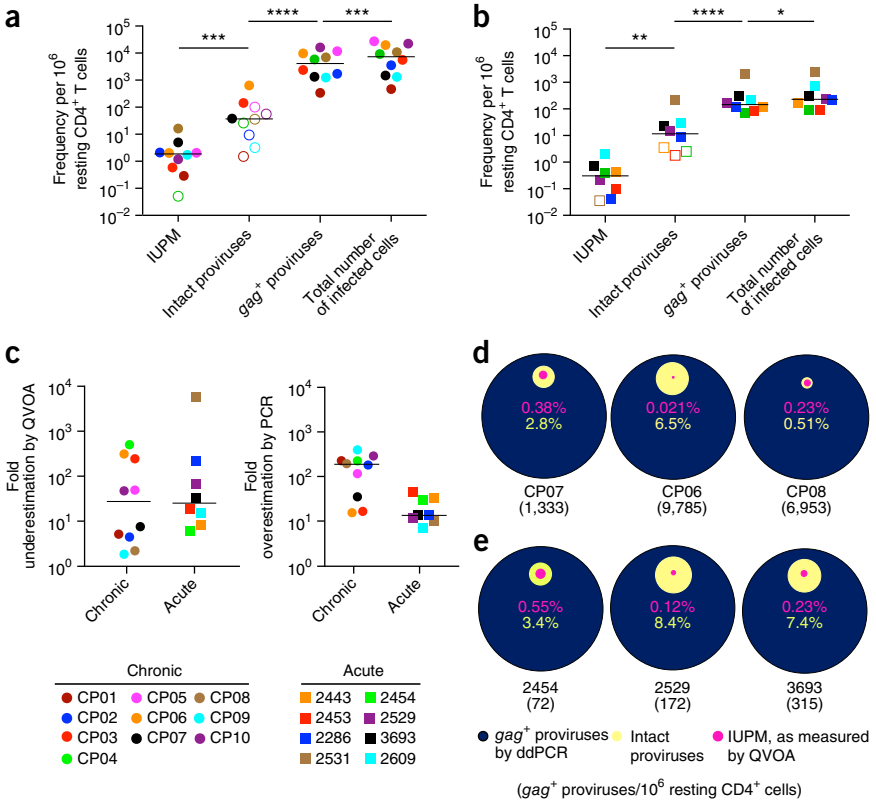

Figure 4 Current assays substantially underestimate or overestimate the size of the latent reservoir. (**a**,**b**) Comparison of different reservoir measurements in resting CD4+ T cells from subjects who initiated ART during chronic (**a**) or during acute or early (**b**) infection. The frequency of infected cells was measured by QVOA, which detects cells that release infectious virus after one round of T cell activation and reports them as the number of infectious units per million resting CD4+ T cells (IUPM). Circles indicate individuals who initiated ART during the chronic phase of infection; squares indicate individuals who initiated ART during the acute or early phase of infection. For individuals in whom the IUPM was below the limit of detection, the median posterior estimate of infection frequency was plotted instead (open symbols). The predicted total number of infected cells was calculated for each subject by correcting the ddPCR result (*gag*+ proviruses) for the fraction of proviruses with deletions in *gag*. The frequency of cells with intact proviruses was calculated as the frequency of infected cells multiplied by the fraction of intact proviruses estimated for each subject using an empirical Bayesian model. Open symbols indicate subjects in which no intact proviruses were detected. Horizontal bars indicate median values. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$ by two-tailed paired *t*-test. The variance was similar between groups. Subject 2521 was excluded from this and subsequent analyses because of a single blip during ART. (**c**) Underestimation of the latent reservoir size by the QVOA (left) and overestimation of the latent reservoir size by DNA PCR (right). Underestimation of the latent reservoir size by the QVOA was calculated by dividing the number of intact proviruses by the QVOA result for each subject. Overestimation of the latent reservoir size by DNA PCR was calculated by dividing the *gag*+ provirus PCR values by the number of intact proviruses. Circles indicate individuals who initiated ART during the chronic phase of infection; squares indicate individuals who initiated ART during the acute or early phase of infection. Horizontal bars indicate median values. (**d**,**e**) Comparison of frequencies of infected cells as determined by QVOA (pink), analysis of intact proviruses (yellow) and ddPCR for *gag*+ proviral DNA (blue) for three representative CP-treated (**d**) or AP-treated (**e**) subjects. All values are plotted as a percentage of the frequency of cells with *gag*+ proviral DNA. The number of *gag*+ proviruses per million resting CD4+ T cells, as measured by ddPCR, is indicated in parentheses for each subject. For comparison plots of additional subjects, see **Supplementary Figure 7d**,**e**.

clones are defective. However, considering that <7% of proviruses are intact, rare expanded clones may carry replication-competent virus and contribute to HIV-1 persistence, as was seen in a recent clinical case study[37]. Although cells harboring defective proviruses can undergo clonal expansion with minimal consequences to the host cells, cells harboring replication-competent viruses could presumably expand by either activation of the cell but not the provirus[38] or by proliferation and survival of the cells despite viral gene expression and virion production.

Our analyses show that the fraction of intact proviruses is much lower than originally thought, even in individuals who were treated early. We have previously shown that intact proviruses replicate well *in vitro* when reconstructed[6]. Furthermore, when wells that were negative for viral outgrowth in the QVOA were reactivated, additional replication-competent virus was isolated[6]. Taken together these data suggest that some, if not all, of these intact proviruses are competent for viral replication and that the number of intact proviruses is probably the closest estimate to the true size of the latent reservoir. Given these results, we sought to more accurately estimate the true reservoir size in AP- and CP-treated individuals, as defined by the number of intact proviruses, and compare that result to current assays, such as the QVOA and DNA PCR. We used a validated droplet digital method to measure proviral DNA with *gag* primers[14,39] (*gag*+ proviruses) and corrected for proviruses that were deleted in *gag* (total number of infected cells). The DNA PCR assay gave frequencies of infected cells that were dramatically higher than the frequency of cells with intact proviruses ($P < 0.0001$ for CP- or AP-treated subjects). The QVOA gave frequencies of infected cells that were dramatically lower than the frequency of cells with intact proviruses ($P < 0.001$ for CP-treated subjects; $P < 0.01$ for AP-treated subjects) (**Fig. 4a**,**b**). We found that the QVOA potentially underestimates the latent reservoir by a median of 27-fold in CP-treated subjects and 25-fold in AP-treated subjects (**Fig. 4c**), whereas DNA PCR overestimated the reservoir size by a median of 188-fold in CP-treated subjects and 13-fold in AP-treated subjects (**Fig. 4c**). Notably, there was no correlation between the number of intact proviruses and either the QVOA or the DNA PCR results (**Supplementary Fig. 7a–c**). Additionally, the relationships between the QVOA, the number of intact proviruses and proviral DNA varied greatly from person to person in both CP-treated subjects (**Fig. 4d** and **Supplementary Fig. 7d**) and AP-treated subjects (**Fig. 4e** and **Supplementary Fig. 7e**). This precludes the use of either the QVOA or DNA PCR as a surrogate marker for the number of intact proviruses, which is probably the closest estimate of the true size of the latent reservoir.

This unbiased screen demonstrates that the fraction of intact proviruses is considerably smaller than was previously shown and that

defects accumulate as early as 2–3 weeks after infection. Although unexpected, the rapid accumulation is consistent with the observation that defects can be observed in 40% of proviruses generated in a single-round *in vitro* infection (**Fig. 2d**) and the fact that cells harboring defective proviruses may be less susceptible to immune clearance or cytopathic effects. With these new data, we were able to revise our estimates of the reservoir size in both CP- and AP-treated individuals. If all of the intact proviruses could be induced *in vivo*, then the true size of the latent reservoir would be, on average, 12 infectious proviruses per million resting $CD4^+$ T cells (median of AP-treated) in individuals who were treated in the acute or early phase of infection and 37 infectious proviruses per million resting $CD4^+$ T cells (median of CP-treated) in individuals who were treated in the chronic phase of infection, with substantial person-to-person variation. Furthermore, our results change our understanding of the effect of early ART on the proviral landscape. Although early ART initiation limits the size of the latent reservoir, it does not have a profound effect on the composition of proviral populations, with the vast majority of proviruses in both CP- and AP-treated adults containing defects. Finally, these results have implications for assessing HIV-1 cure strategies. Because there is no correlation between the QVOA or DNA PCR and the number of intact proviruses, these assays cannot be used to accurately predict the true reservoir size, even for individuals who are treated early in infection. Indeed, any analysis of subgenomic regions of proviruses to evaluate viral evolution or reservoir reduction must be reconsidered in this context, as the majority of sequences studied will be from defective proviruses[40]. Notably, because the nature of the defects described here indicates that many defective proviruses may not be eliminated by eradication strategies, defective proviruses could obscure the measurement of real changes in the rarer population of intact proviruses that must be eliminated to achieve a cure.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The sequences reported in this paper have been deposited in the GenBank database (354 sequences, with accession nos. KX505390 to KX505744).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the study participants who have made this research possible. We also thank G. Laird for critical advice and discussion, L. Alston, H. McHugh and D. Xu for assistance with study participants, V. Walker-Sperling for providing the Ba-L virus, C. Pohlmeyer for providing a sample from a viremic subject, and all of the members of the Siliciano laboratory for valuable discussion and advice. This work was supported by the Genomics and Sequencing Core at the UCSD Center for AIDS Research (P30AI036214 (D.D.R.)), the Pendleton Charitable Trust (D.D.R.), the VA San Diego Healthcare System (D.D.R.), the Veterans Medical Research Foundation (D.D.R.), the Martin Delaney CARE and DARE Collaboratories (US National Institutes of Health (NIH) grants AI096113 (M.C.S., D.D.R. and R.F.S.) and 1U19AI096109 (S.G.D. and R.F.S.)), an ARCHE Collaborative Research Grant from the Foundation for AIDS Research (amFAR 108165-50-RGRL (R.F.S.)), the Johns Hopkins Center for AIDS Research grant P30AI094189 (R.F.S.), the NIH grants 43222 (R.F.S.) and R21AI113147-02 (J.D.S.), the Howard Hughes Medical Institute (R.F.S.), and the Bill and Melinda Gates Foundation (R.F.S.).

AUTHOR CONTRIBUTIONS

K.M.B., R.A.P., S.G.D. and R.F.S. designed the experiments; K.M.B., A.J.M., R.A.P., M.G.S., A.A.C., J.L., M.C.S. and S.M.L. performed the experiments; K.M.B., A.J.M., S.B.L., Y.-C.H., D.D.R., J.D.S. and R.F.S. analyzed the data; R.H. and A.A.C. managed study participant recruitment; S.G.D. provided patient samples; and K.M.B. and R.F.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Finzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
2. Chun, T.W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* **94**, 13193–13197 (1997).
3. Wong, J.K. *et al.* Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295 (1997).
4. Siliciano, J.D. *et al.* 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 (2003).
5. Ruelas, D.S. & Greene, W.C. An integrated overview of HIV-1 latency. *Cell* **155**, 519–529 (2013).
6. Ho, Y.-C. *et al.* Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
7. Finzi, D. *et al.* Latent infection of $CD4^+$ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
8. Crooks, A.M. *et al.* Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J. Infect. Dis.* **212**, 1361–1365 (2015).
9. Archin, N.M. & Margolis, D.M. Emerging strategies to deplete the HIV reservoir. *Curr. Opin. Infect. Dis.* **27**, 29–35 (2014).
10. Deeks, S.G. HIV: shock and kill. *Nature* **487**, 439–440 (2012).
11. Archin, N.M. *et al.* Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* **487**, 482–485 (2012).
12. Siliciano, J.D. & Siliciano, R.F. 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 (2005).
13. Laird, G.M. *et al.* Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398 (2013).
14. Strain, M.C. *et al.* Highly precise measurement of HIV DNA by droplet digital PCR. *PLoS One* **8**, e55943 (2013).
15. Rouzioux, C., Mélard, A. & Avéttand-Fénoël, V. in *Human Retroviruses* (eds. Vicenzi, E. & Poli, G.) **1087**, 261–270 (Humana Press, 2013).
16. Henrich, T.J., Gallien, S., Li, J.Z., Pereyra, F. & Kuritzkes, D.R. Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR. *J. Virol. Methods* **186**, 68–72 (2012).
17. Eriksson, S. *et al.* Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog.* **9**, e1003174 (2013).
18. Günthard, H.F. *et al.* Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society–USA Panel. *J. Am. Med. Assoc.* **312**, 410–425 (2014).
19. Archin, N.M. *et al.* Immediate antiviral therapy appears to restrict resting $CD4^+$ cell HIV-1 infection without accelerating the decay of latent infection. *Proc. Natl. Acad. Sci. USA* **109**, 9523–9528 (2012).
20. Jain, V. *et al.* Antiretroviral therapy initiated within 6 months of HIV infection is associated with lower T cell activation and smaller HIV reservoir size. *J. Infect. Dis.* **208**, 1202–1211 (2013).
21. Deng, K. *et al.* Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517**, 381–385 (2015).
22. Delviks-Frankenberry, K. *et al.* Mechanisms and factors that influence high-frequency retroviral recombination. *Viruses* **3**, 1650–1680 (2011).
23. Jetzt, A.E. *et al.* High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**, 1234–1240 (2000).
24. Yu, H., Jetzt, A.E., Ron, Y., Preston, B.D. & Dougherty, J.P. The nature of human immunodeficiency virus type 1 strand transfers. *J. Biol. Chem.* **273**, 28384–28391 (1998).
25. Hwang, C.K., Svarovskaia, E.S. & Pathak, V.K. Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of *in vivo* template switching. *Proc. Natl. Acad. Sci. USA* **98**, 12209–12214 (2001).
26. Yu, Q. *et al.* Single-strand specificity of APOBEC3G accounts for minus strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **11**, 435–442 (2004).
27. Kieffer, T.L. *et al.* G–>A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting $CD4^+$ T cells *in vivo*. *J. Virol.* **79**, 1975–1980 (2005).
28. Pierson, T.C. *et al.* Molecular characterization of preintegration latency in human immunodeficiency virus type 1 infection. *J. Virol.* **76**, 8518–8531 (2002).
29. Pillai, S.K. *et al.* Role of retroviral restriction factors in the interferon-α-mediated suppression of HIV-1 *in vivo*. *Proc. Natl. Acad. Sci. USA* **109**, 3035–3040 (2012).

30. Harper, M.S. *et al.* Interferon-α subtypes in an *ex vivo* model of acute HIV-1 infection: expression, potency and effector mechanisms. *PLoS Pathog.* **11**, e1005254 (2015).
31. Stacey, A.R. *et al.* Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J. Virol.* **83**, 3719–3733 (2009).
32. Kearney, M.F. *et al.* Origin of rebound plasma HIV includes cells with identical proviruses that are transcriptionally active before stopping of antiretroviral therapy. *J. Virol.* **90**, 1369–1376 (2016).
33. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
34. Wagner, T.A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).
35. Schröder, A.R.W. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002).
36. Cohn, L.B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
37. Simonetti, F.R. *et al.* Clonally expanded CD4+ T cells can produce infectious HIV-1 *in vivo*. *Proc. Natl. Acad. Sci. USA* **113**, 1883–1888 (2016).
38. Bosque, A., Famiglietti, M., Weyrich, A.S., Goulston, C. & Planelles, V. Homeostatic proliferation fails to efficiently reactivate HIV-1 latently infected central memory CD4+ T cells. *PLoS Pathog.* **7**, e1002288 (2011).
39. Massanella, M., Gianella, S., Lada, S.M., Richman, D.D. & Strain, M. Quantification of total and 2-LTR (long terminal repeat) HIV DNA, HIV RNA and herpesvirus DNA in PBMCs. *Bio Protoc.* **5**, e1492 (2015).
40. Lorenzo-Redondo, R. *et al.* Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* **530**, 51–56 (2016).

ONLINE METHODS

Study subjects. The Johns Hopkins Institutional Review Board and the UCSF Committee on Human Research approved this study. All participants provided written consent before enrollment. Nineteen HIV-1-infected individuals who met the criteria of suppressive ART and undetectable plasma HIV-1 RNA levels (<50 copies/ml) for a minimum of 8 months were enrolled. Ten of these participants were recruited from the SCOPE and OPTIONS cohorts at the University of California, San Francisco. CP-treated subjects are defined as subjects starting ART >180 d from the estimated date of infection. AP-treated subjects started ART <100 d after the estimated date of infection. Subjects with viremia were either untreated or had viral loads >4,000 copies/ml during ART. **Supplementary Table 1** details the characteristics of the study participants.

Isolation of resting CD4$^+$ T lymphocytes. Peripheral blood mononuclear cells (PBMCs) were isolated, using density centrifugation, on a Ficoll–Hypaque gradient. CD4$^+$ T cells were isolated from PBMCs using a negative-selection method (CD4$^+$ T cell Isolation Kit II, Miltenyi Biotec). Resting CD4$^+$ lymphocytes (CD4$^+$CD69$^-$CD25$^-$HLA-DR$^-$) were enriched by a second negative depletion using biotin-conjugated anti-CD25, anti-biotin MicroBeads, CD69 MicroBead Kit II, and anti–HLA-DR MicroBeads (all from Miltenyi Biotec). The purity of the resting CD4$^+$ cells was consistently >95%, as assessed using flow cytometry.

DNA extraction and limiting-dilution PCR. DNA was extracted from 2×10^6 resting CD4$^+$ cells by using the Qiagen Gentra Purgene Cell Kit A, which allows for extraction of large fragments (~200 kb in size) so as to minimize fragmentation of HIV-1 genomes. DNA was subjected to a nested limiting-dilution PCR protocol modified from Ho *et al.* using Platinum Taq HiFi Polymerase (Life Technologies)[6]. The outer PCR was nearly full-length from HIV-1 reference genome HXB2 coordinates 623 to coordinates 9,686 and employed a touch-down-PCR protocol. The outer-PCR wells were diluted 1:3 with deionized water and 1 µl of outer-PCR DNA was used for nested amplification of both *gag* and *env* to determine clonality. Clonality was determined using Poisson statistics and at dilutions that gave a high probability of clonality ($P > 0.85$); all outer-PCR wells, including those that were negative for *gag* and *env*, were subjected to six inner PCRs. 1 µl of outer-PCR DNA was used for each nested inner PCR reaction. A detailed PCR protocol and primer sequences and locations, can be found in **Supplementary Table 2**. PCR products were visualized on 1% agarose gels. The products were directly sequenced to minimize PCR-induced error, using the QIAquick Gel Extraction Kit followed by Sanger sequencing. Sequencing reads from the six overlapping nested PCRs were aligned and compared to the reference HIV-1 genome HXB2 using CodonCode Aligner software to reconstruct near-full-length sequences and identify defects. Hypermutation was determined by using the full-length sequence for each clone and the Los Alamos hypermut algorithm[41]. Some hypermutated clones had extensive hypermutation that prevented full sequencing of the entire genome. Because the PCR products were correctly sized on an agarose gel, the sequences were inferred to be full-length hypermutated. We were able to directly map the majority of deletions by direct sequencing of PCR products, which were generated with primers flanking the deletion junction. Some very large internal deletions were identified despite deletion of some inner primer-binding sites. The observed bands likely resulted from exponential amplification in the outer PCR and linear amplification in the inner PCRs. These deletions were confirmed by sequencing bands from multiple inner PCRs, all of which showed the exact same deletion product.

***In vitro* infections.** HIV-1-negative donor CD4$^+$ T cells were isolated as described above and activated with phytohemagglutinin (PHA, Fisher Scientific) (0.5 µg/ml) for 3 d in enriched T cell medium (TCM) consisting of RPMI supplemented with 10% fetal calf serum, interleukin (IL)-2 and cytokine-rich supernatant, as previously described[42]. Cells were infected with replication-competent HIV-1 Ba-L virus (300 ng p24 viral capsid protein per 10^6 cells) by spinoculation (2 h, 37 °C, 1200*g*). Cells were then suspended at 3×10^6/ml in TCM and incubated for 6 d at 37 °C. Following the incubation, supernatant was collected and stored at −80 °C. 25×10^6 HIV-1-negative donor CD4$^+$ T cells were activated by anti-CD3 and anti-CD28 stimulation (BD Biosciences) for 72 h in TCM, as previously described[42]. Activated cells were suspended in 2.5 ml of TCM and distributed in 100-µl aliquots into a 96-well V-bottom plate (Corning). A 500-ng inoculum of HIV-1 Ba-L was added to each well, and the cells were infected by spinoculation (2 h, 37 °C, 1200*g*). Following the spinoculation, cells were suspended in 40 ml of TCM supplemented with enfuvirtide (10 µM) (Genentech) to prevent additional rounds of viral replication. Following a 2-d incubation at 37 °C, genomic DNA was isolated and analyzed as described above.

Quantitative viral outgrowth assay (QVOA). The QVOAs were performed as previously described[1,12,13]. MOLT-4 cells transfected with CCR5 (MOLT4/CCR5; obtained from the NIH AIDS reagent program) were added on day 2 of the culture[13], and the culture supernatants were examined using the standard protocol, which involves an ELISA for the p24 viral capsid protein (PerkinElmer) after 21 d (ref. 12). HIV-1 RNA in the culture supernatant was not measured. MOLT-4/CCR5 cells were negative for mycoplasma contamination and were stained to verify expression of CD4 and the HIV-1 co-receptors CXCR4 and CCR5. Infectious units per million cells (IUPM) values were calculated using IUPMStats[43]. If all culture wells were negative for viral outgrowth, then the median posterior estimate of infection frequency was used.

Droplet digital PCR (ddPCR). DNA was extracted from 5×10^6 to 10×10^6 resting CD4$^+$ T cells with the QIAamp DNA Blood Mini Kit. DNA was fragmented and a duplex droplet digital PCR was performed with primers in the *gag* gene, as well as primers for 2-LTR circles, as previously described[14,39]. For patient 2609, primers in the *pol* gene were used instead, as the *gag* PCR primers and probes contained multiple mismatches with the patient sequence and the *gag* gene did not get amplified[39]. Measurements of the cellular gene ribonuclease P/MRP subunit p30 (*RPP30*) in a replicate well by ddPCR were used to calculate the frequency of resting CD4$^+$ T cells. The frequencies of *gag*$^+$ DNA and 2-LTR circles were plotted as a frequency per 10^6 resting CD4$^+$ T cells. In the case of a negative 2-LTR circle measurement, the lower limit of detection was determined by using the total cellular input for that sample. Because many of the proviruses examined were *gag* negative, we also estimated the total number of infected cells in study subject resting CD4$^+$ T cells as the copies of *gag*$^+$ proviral DNA divided by the fraction of proviruses detected that were *gag*$^+$. For patient 2609 the total number of infected cells was estimated as the copies of *pol*$^+$ proviral DNA divided by the fraction of proviruses detected that were *pol*$^+$, as the *gag* PCR primers did not amplify the gene.

Bayesian analysis to calculate number of intact proviruses. The proportion of intact genomes for each subject was estimated by using an empirical Bayesian analysis, with prior distributions generated separately for the acute and chronic subject cohorts. The proportion of clones identified as intact was calculated for each subject, and the mean and variance of this measurement were used to generate beta-prior distributions with the same mean and variance. The proviral genome sequences were treated as Bernoulli trials, with intact clones counted as successes. The final estimate for the percentage of intact genomes for each subject was calculated as the expected value of the posterior distribution. The number of intact proviruses was calculated as the frequency of cells with total HIV-1 DNA multiplied by the frequency of intact proviruses estimated for each subject using an empirical Bayesian model.

Correlation plots and statistical analysis. All data sets, except the percentage of intact proviruses, followed a log distribution and were log-transformed. The log-transformed data met the D'Agostino–Pearson test for normal distribution, and Pearson correlations were performed on log-transformed data. For the QVOA data in which two of the samples were below the limit of detection, the median posterior estimate of infection frequency was plotted instead. $P < 0.05$ was considered statistically significant. Prism software (GraphPad) was used to perform all statistical tests and to generate correlation plots.

41. Rose, P.P. & Korber, B.T. Detecting hypermutations in viral sequences with an emphasis on G–> A hypermutation. *Bioinformatics* **16**, 400–401 (2000).
42. Kim, M. *et al.* A primary CD4$^+$ T cell model of HIV-1 latency established after activation through the T cell receptor and subsequent return to quiescence. *Nat. Protoc.* **9**, 2755–2770 (2014).
43. Rosenbloom, D.I.S. *et al.* Designing and interpreting limiting dilution assays: general principles and applications to the latent reservoir for human immunodeficiency virus-1. *Open Forum Infect. Dis.* **2**, ofv123 (2015).

doi:10.1038/nm.4156