LETTER

https://doi.org/10.1038/s41586-019-0898-8

A quantitative approach for measuring the reservoir of latent HIV-1 proviruses

Katherine M. Bruner[1,8,10], Zheng Wang[1,10], Francesco R. Simonetti[1], Alexandra M. Bender[1], Kyungyoon J. Kwon[1], Srona Sengupta[1], Emily J. Fray[1], Subul A. Beg[1], Annukka A. R. Antar[1], Katharine M. Jenike[1], Lynn N. Bertagnolli[1], Adam A. Capoferri[1], Joshua T. Kufera[1], Andrew Timmons[1], Christopher Nobles[2], John Gregg[2], Nikolas Wada[3], Ya-Chi Ho[1,9], Hao Zhang[4], Joseph B. Margolick[4], Joel N. Blankson[1], Steven G. Deeks[5], Frederic D. Bushman[2], Janet D. Siliciano[1], Gregory M. Laird[6] & Robert F. Siliciano[1,7]*

A stable latent reservoir for HIV-1 in resting CD4$^+$ T cells is the principal barrier to a cure[1–3]. Curative strategies that target the reservoir are being tested[4,5] and require accurate, scalable reservoir assays. The reservoir was defined with quantitative viral outgrowth assays for cells that release infectious virus after one round of T cell activation[1]. However, these quantitative outgrowth assays and newer assays for cells that produce viral RNA after activation[6] may underestimate the reservoir size because one round of activation does not induce all proviruses[7]. Many studies rely on simple assays based on polymerase chain reaction to detect proviral DNA regardless of transcriptional status, but the clinical relevance of these assays is unclear, as the vast majority of proviruses are defective[7–9]. Here we describe a more accurate method of measuring the HIV-1 reservoir that separately quantifies intact and defective proviruses. We show that the dynamics of cells that carry intact and defective proviruses are different in vitro and in vivo. These findings have implications for targeting the intact proviruses that are a barrier to curing HIV infection.

Evaluation of cure strategies requires assays that detect infected cells and distinguish intact proviruses from the vast excess of defective proviruses. We define 'intact proviruses' as those that lack overt fatal defects such as large deletions and hypermutation[7,8], and recognize that some proviruses defined in this way may have minor defects that affect fitness. We analysed 431 near full genome HIV-1 sequences obtained by single genome analysis[8,9] from 28 HIV-1-infected adults. The near full genome sequencing (nFGS) methods used identify defects throughout the genome except the 5′ long terminal repeat (LTR). Consistent with previous reports[8,9], only 2.4% of proviruses were intact (Fig. 1a). The remaining 97.6% had fatal defects including deletions, encompassing on average 49.6% of the genome, and/or G→A hypermutation, which altered start codons and/or introduced stop codons in most open-reading frames (ORFs)[8]. Most defective proviruses had defects in most HIV-1 genes (Fig. 1b, Supplementary Table 1). Ninety-seven per cent of defective proviruses had defects that affect the transcriptional activator Tat (Fig. 1b, Supplementary Table 1) and might not be efficiently transcribed after latency reversal. Thus, cure interventions dependent on viral gene expression[4,5] may affect cells with intact and defective proviruses differently. Hence, separate quantification is essential.

Standard PCR assays use short subgenomic amplicons in conserved regions (Fig. 1c, d) and do not distinguish intact and defective proviruses. Deletions occur throughout the genome (Fig. 1e), affecting not only the fraction of proviruses detected (Fig. 1f) but also the fraction of detected proviruses that are intact (Fig. 1g). For most standard PCR assays, less than 10% of detected sequences are intact (Fig. 1g). Thus, efficacy of a cure intervention causing a selective one log reduction in intact proviruses may be inapparent (Fig. 1h).

Interrogating individual proviruses simultaneously at several positions could differentiate intact from defective proviruses. Analysis of nFGS data revealed that strategically placed amplicons in the packaging signal (Ψ) and *env* regions could jointly identify more than 90% of deleted proviruses as defective (Fig. 2a). Hypermutated proviruses must also be identified (Fig. 1a). Seventy-three per cent of these have mutations in the GG→AG context (Extended Data Fig. 1a). Most also have GA→AA mutations. Only 27% of hypermutated proviruses had only GA→AA mutations, and most of these also had deletions (Extended Data Fig. 1a). Therefore, we focused on GG→AG hypermutation. We identified a conserved region in the Rev-response element (RRE) with adjacent consensus sites (TGGG) for the responsible enzyme, APOBEC3G[10] (Fig. 2b–e). Of the sequences with GG→AG hypermutation, 97% had one or more mutations in this region (Fig. 2c, e), with 13 distinct patterns (Fig. 2f). Using mutant plasmids carrying each pattern, we developed allelic discrimination probes (Extended Data Fig. 1b) that correctly identify 95% of hypermutated sequences as defective (Fig. 2f).

These analyses allowed design of a droplet digital PCR (ddPCR) method that distinguishes most deleted and/or hypermutated proviruses from intact proviruses using two amplicons and hypermutation discrimination probes (Fig. 3a, b). Genomic DNA (gDNA) is isolated using an optimized method to minimize DNA shearing between targeted regions (see below) and partitioned into nanodroplets such that individual droplets rarely contain more than one provirus ($P = 0.00416$). Proviruses within droplets are analysed simultaneously at the Ψ and *env* regions via multiplex PCR, with the *env* PCR also discriminating hypermutated proviruses. Intact proviruses give amplification at both regions (Fig. 3a, b). Intact proviruses per 10^6 cells are calculated using separate amplification of a cellular gene (*RPP30*) after correction for DNA shearing.

The ability of this intact proviral DNA assay (IPDA) to distinguish intact and defective proviruses was verified using plasmid controls that represent proviruses with different defects. These templates gave positive droplets in the expected quadrants: quadrant 1 (Q1) for 3′ deletion and/or hypermutation and Q4 for 5′ deletion (Extended Data Fig. 2). Cultured clonal populations of infected patient cells also give single quadrant patterns (see below). By contrast, uncultured polyclonal patient CD4$^+$ T cell populations gave droplets in all four quadrants, allowing separate quantification of intact and defective proviruses (Fig. 3c).

Validation of the IPDA is described in Extended Data Figs. 3 and 4. DNA shearing is essential for droplet formation but artificially reduces

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [4]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [5]Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [6]Accelevir Diagnostics, Baltimore, MD, USA. [7]Howard Hughes Medical Institute, Baltimore, MD, USA. [8]Present address: Department of Molecular Biosciences, University of Texas, Austin, TX, USA. [9]Present address: Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, CT, USA. [10]These authors contributed equally: Katherine M. Bruner, Zheng Wang. *e-mail: rsiliciano@jhmi.edu

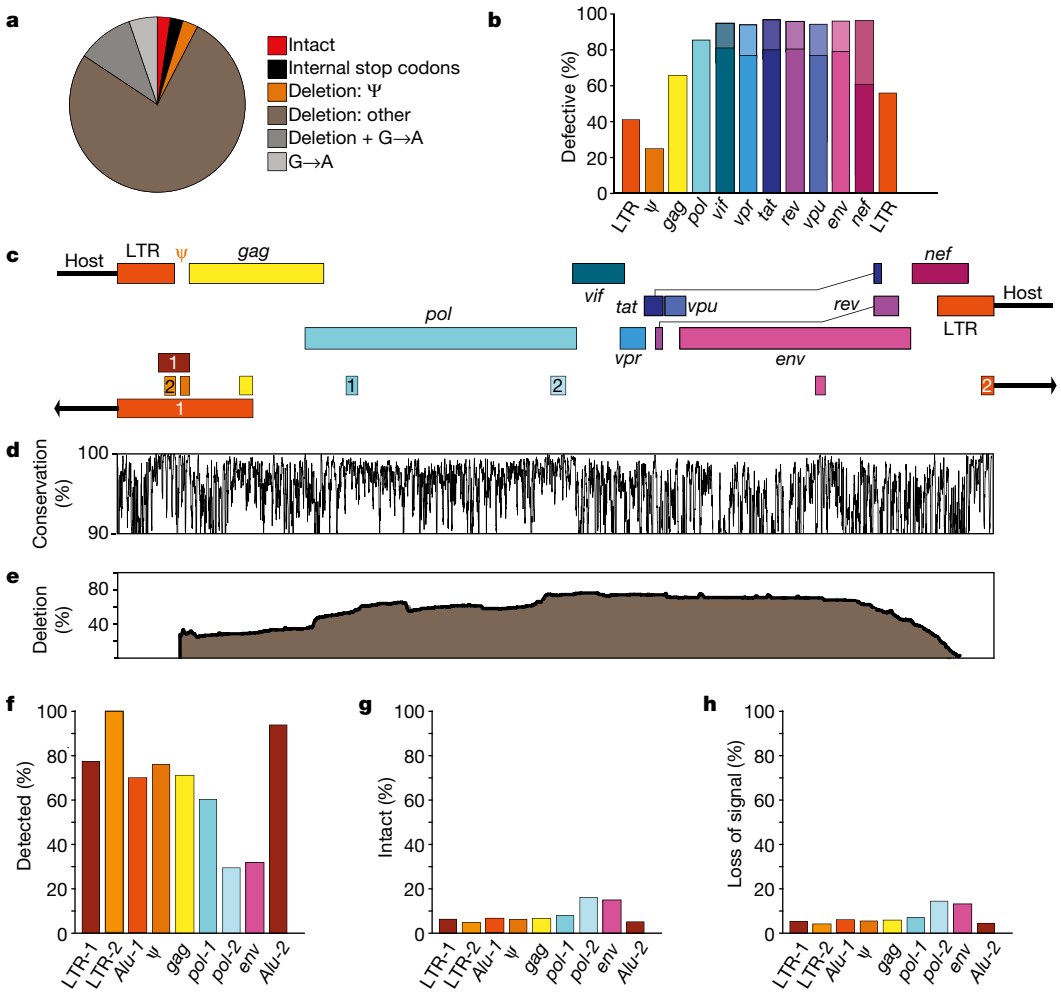


Fig. 1 | DNA PCR assays predominantly measure defective proviruses. **a**, Proviruses persisting in CD4$^+$ T cells of individuals on suppressive ART as detected by nFGS[8,9]. Defects include internal stop codons, deletions not attributable to normal length polymorphisms, and APOBEC3G/F-mediated hypermutation (G→A). Most deletions were large except for those in the packaging signal (ψ) or major splice donor site. Analysis is based on 211 sequences from individuals initiating ART during chronic infection. **b**, Fraction of defective proviruses with defects in the indicated genes or elements. Protein-coding genes were considered defective if any of the following were present: mutated start codon, internal stop codons, frameshifts, or insertions or deletions not representing common length polymorphisms. Splice site mutations further increased the fraction of defective sequences (lighter shaded portions of the bars for spliced genes). LTR sequences were considered defective if mutations in the NF-κB sites and/or deletions were present. See Supplementary Table 1 for details. **c**, Positions of amplicons used in standard DNA and *Alu* PCR assays. See Supplementary Table 2 for coordinates and references. Numbers distinguish distinct assays targeting the same region. *Alu* PCR assays also amplify host genomic sequence (arrows). **d**, Conservation of sequence across the genome based on US clade B sequences in the Los Alamos HIV Sequence Database (https://www.hiv.lanl.gov/). Plotted as the percentage of sequences matching the consensus at each nucleotide. **e**, Position of internal deletions across the HIV-1 genome. Plotted as percentage of total sequences from treated patients deleted at the indicated nucleotide. **f**, Percentage of all proviruses that are amplified by the indicated assay. Analysis is based on the absence of an overlap between deleted regions and the relevant amplicons. **g**, Percentage of the proviruses detected by the indicated assay that are intact. **h**, Percentage loss of assay signal after a selective tenfold reduction in intact proviruses. For **f**–**h**, analysis is based 211 sequences from 19 patients starting ART during chronic infection.

double-positive (Q2) droplets while increasing Q1 and Q4 droplets. To control for shearing, two regions of a cellular gene (*RPP30*) with the same spacing as the Ψ and *env* amplicons are amplified in a separate multiplex ddPCR on each sample. This gives a DNA shearing index (DSI) to correct raw HIV-1 ddPCR results (see Methods). To validate this approach, DNA from JLat cells[11], which contain a single provirus, was subjected to different amounts of shearing and analysed by ddPCR for HIV-1 and *RPP30*. At all levels of shearing, DSI was similar for HIV-1 and *RPP30*. Thus, *RPP30* shearing can be used to correct for HIV-1 shearing (Fig. 3d). At all levels of shearing, corrected results were close to the expected values of one copy per cell for HIV-1 (Fig. 3e) and four copies per cell for *RPP30* (JLat cells are tetraploid for *RPP30*). With correction for shearing, the IPDA gives highly accurate discrimination between molecularly defined standards representing intact and defective proviruses (Extended Data Fig. 5). In acute in vitro infections, high molecular mass DNA from productively infected CD4$^+$ T cells showed a substantial fraction of intact proviruses as assessed by IPDA and confirmed by nFGS (Extended Data Fig. 6). Thus, the assay can reliably quantify intact proviruses in most DNA samples after correction for shearing.

Using samples of only 5×10^6 cells, we measured intact and defective proviruses in 62 infected individuals who had suppression of viraemia on antiretroviral therapy (ART) ($n = 57$) or 'elite' control of HIV-1 without ART ($n = 5$) (Supplementary Table 3). The median DSI was 0.32 (interquartile range ± 0.07; Extended Data Fig. 7), well within the range for which accurate correction is possible. Intact proviruses were rare (approximately 100 per 10^6 resting CD4$^+$ T cells) and greatly outnumbered by defective proviruses (Fig. 3f). These results provide the first demonstration, to our knowledge, using a method not requiring long-distance PCR that the proviral landscape is dominated by defective proviruses. We found strong correlations between IPDA and quantitative viral outgrowth assay (QVOA) measurements on the same

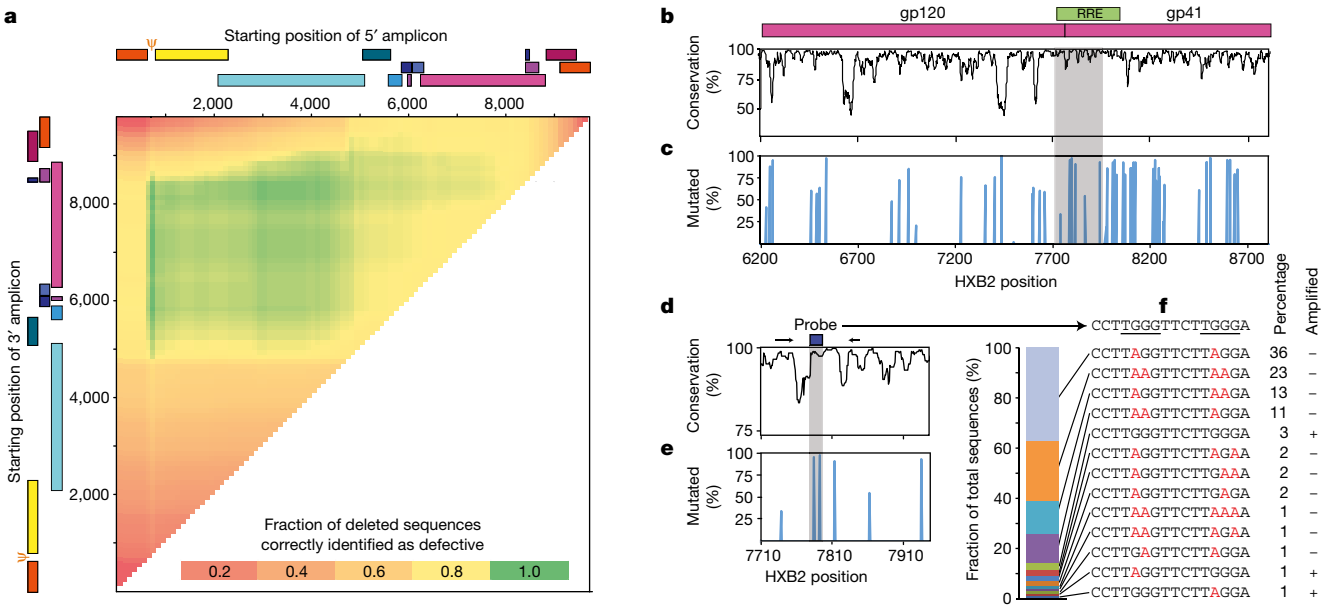


Fig. 2 | Distinguishing intact and defective HIV-1 proviruses. a, Sliding window analysis of optimal amplicon positioning to detect deletions. Optimal discrimination between intact and deleted sequences is obtained with a 5′ amplicon in the Ψ region and a 3′ amplicon in *env*. Ψ is the site of frequent small deletions and is included in larger 5′ deletions. Analysis is based on 431 independent near full genome sequences in the database including 258 that contain mapped deletions. Inclusion of an additional 474 sequences from treated patients did not change optimal positions. **b**, Nucleotide conservation across the *env* gene based on US clade B sequences in the Los Alamos HIV Sequence Database is plotted as the percentage of sequences matching the consensus sequence at each nucleotide. Shaded area is expanded in Fig. 2d. **c**, GG→AG hypermutation across the *env* gene. The percentage of hypermutated proviruses that contain one or more G→A mutations within a given APOBEC3G consensus site is plotted as a function of site position. Shaded area is expanded in Fig. 2e. **d**, Sequence conservation of a region in the RRE containing two APOBEC3G consensus sites. Primer (arrows) and probe (box) positions for the *env* amplicon are shown. **e**, Fraction of hypermutated proviruses with G→A mutations in the probe-binding region (shaded). **f**, Hypermutation patterns at the *env* probe-binding site. The prevalence of 13 observed patterns is indicated in the bar graph on the left and in the 'percentage' column. Mutations in the APOBEC3G consensus sites (underlined) are indicated in red. Analysis is based on 93 independent hypermutated *env* sequences from 18 treated patients. Site-directed mutagenesis was used to modify NL4-3 or a patient-derived proviral construct to generate plasmids containing each pattern. The 'amplified' column indicates that only 5% of hypermutated sequences were amplified by the probe combinations developed to identify intact sequences.

samples ($r = 0.48$, $P = 0.003$, $n = 35$, Fig. 3g), even though the fraction of intact proviruses induced by a single round of T cell activation in the QVOA is small and variable between individuals (Fig. 3h).

Using nFGS data, we compared the IPDA with the widely used *gag* PCR with respect to desired assay characteristics. IPDA detects a larger fraction of proviruses (Fig. 3i) and separately enumerates intact and two class of defective proviruses. More importantly, the IPDA excludes 97% of proviruses with defects detectable by nFGS (Fig. 3i). The fraction of defective proviruses excluded by standard PCR assays is much lower and almost zero for *Alu* PCR assays. It is only 30% for *gag* PCR. Most proviruses (approximately 70%) classified as intact by IPDA lack defects detectable by nFGS. By contrast, this value is less than 10% for *gag* PCR (Fig. 3i). Thus, the IPDA provides a scalable alternative that is much more selective for intact proviruses. Most defects missed by IPDA are small deletions not overlapping IPDA amplicons. In addition, although nFGS data used to design the assay covers 97.6% of the non-redundant HIV-1 sequence, small deletions in the remaining portion of the genome may be present in some proviruses. The assay was designed for use in treated patients and does not distinguish integrated proviruses from unintegrated linear or circular forms. These are rare in patients on long-term ART[8]. Further sequence data and improved ddPCR technology may allow even greater selectivity for intact proviruses. Importantly, selective intervention-induced reductions in intact proviruses would be apparent with the current IPDA but not with standard assays (Fig. 3i).

The latent reservoir measured by QVOA undergoes slow decay ($t_{1/2} = 44$ months)[12,13]. Intact proviruses measured by IPDA show similar decay (Fig. 4a, b). For most patients, intact proviruses declined with a $t_{1/2}$ of approximately 44 months. Some patients showed even slower decay ($t_{1/2} = 100$–300 months), and for 3 out of 14 patients, there was no decay (Fig. 4b, c). Thus, the IPDA can measure changes consistent with known reservoir dynamics and possibly define subpopulations with slower decay.

The dynamics of defective proviruses in the same patients showed greater variability, with increases over time in some patients (Fig. 4a, c, Extended Data Fig. 8). Although the mean decay slopes did not differ, the standard deviations of decay slopes were much greater for defective proviruses (Extended Data Fig. 9), complicating interpretation of assays dominated by defective forms. Increases in infected cell frequency can reflect the proliferation of infected cells[14–19]. Although the IPDA cannot demonstrate the presence of expanded clones, it can detect increases in infected cell frequency due to clonal expansion. We asked whether the differential dynamics of cells with intact and defective proviruses could be due to uneven distribution in subsets of memory $CD4^+$ T cells[20] with different proliferative potential. Intact and defective proviruses were found in expected ratios in central, transitional and effector memory subsets (Fig. 4d). We then asked whether intact and defective proviruses imposed different constraints on proliferation. We used the IPDA to track individual infected cells from treated patients after in vitro stimulation with anti-CD3/28 (Fig. 4e). Microcultures seeded with around one infected cell per well were subjected to four rounds of stimulation in the presence of antiretroviral drugs. Analysis of more than 1,700 microcultures showed that positive droplets were detected in only one out of three possible quadrants for most positive wells, as expected for cultures initiated with a single infected cell and cultured with antiretroviral drugs (Fig. 4e). Most (97.6%) proviruses were defective (Fig. 4f). Importantly, the IPDA also counts proviruses in each well, allowing the direct quantification of infected cell proliferation. Some cells carrying defective proviruses showed enormous clonal expansion (1,000-fold), whereas cells with intact proviruses were rarely detected and showed little proliferation (Fig. 4f). Despite fewer wells with intact

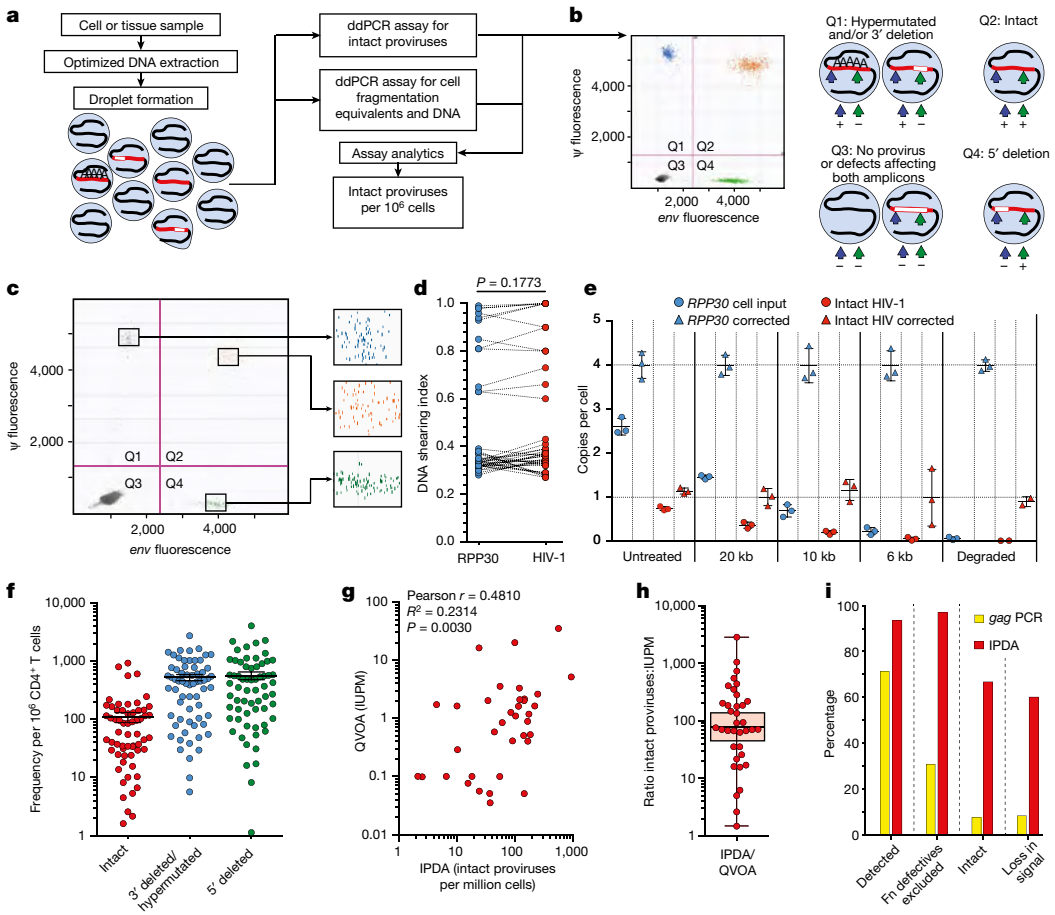


Fig. 3 | IPDA. a, Assay schematic. Multiplex PCRs in droplets amplify Ψ and *env* regions. A separate multiplex PCR targets two regions of the human *RPP30* gene spaced at the same distance as the Ψ and *env* amplicons to provide cell number quantification and DNA shearing correction. See Methods for details. **b**, Representative control ddPCR experiment using proviral constructs[25] with a 5′ deletion (E44E11), a 3′ deletion (4F11), or no defects (NL4-3). A total of 1,000 copies each were mixed with 500 ng of HIV-1-negative DNA to simulate a patient sample. Types of provirus appearing in different quadrants (Q1–Q4) are shown on the right. **c**, Representative IPDA results from a patient $CD4^+$ T cell sample. Boxed areas are expanded to show individual positive droplets. **d**, DNA shearing index (DSI; fraction of templates sheared between targeted regions) measured for *RPP30* and HIV-1 on JLat DNA samples subjected to different levels of shearing ($n = 22$). Compared using two tailed *t*-test for paired non-parametric values. **e**, Use of DSI to correct raw ddPCR output for *RPP30* and HIV. Mean and s.d. of copies per cell of *RPP30* (blue) and HIV (orange) are shown before (circles) and after (triangles) correction for shearing. **f**, IDPA results on $CD4^+$ T cells from infected individuals ($n = 62$) with plasma HIV-1 RNA below the limit of detection. Data are geometric mean ± s.e.m. See Supplementary Table 3 for patient characteristics. Polymorphisms precluding amplification with either primer/probe set were not observed in this cohort and would require triage primer/probe sets incorporating rare polymorphisms. **g**, Correlation between infected cell frequencies measured by QVOA and IPDA on the same samples of $CD4^+$ T cells from treated patients ($n = 36$). IUPM, infectious units per million cells. **h**, IPDA/QVOA ratios for samples from Fig. 3g. Horizontal bars indicate geometric mean and 95% confidence interval. **i**, Bioinformatic comparison of standard *gag* PCR and IPDA with respect to the percentage of proviruses amplified, percentage of defective proviruses excluded, percentage of amplified proviruses that are intact, and percentage loss in assay signal after a selective tenfold reduction in intact proviruses. Fn, fraction. All fractions expressed as percentages. See Fig. 1f–h for details.

proviruses, differences in the mean number of intact and defective proviruses per well remained highly significant ($P = 0.0029$ to $P < 0.0001$ by unpaired *t*-test with Welch's correction for intact proviruses versus proviruses with hypermutation and/or 3′ deletions). These differences are minimal estimates because the fraction of positive microcultures with intact proviruses was lower than expected based on IPDA analysis of the starting population (3.3% versus 7.5%), indicating that some cells with intact proviruses die. These results demonstrate a proliferative defect in response to repetitive in vitro T cell receptor stimulation for cells carrying intact proviruses and show that the IPDA can be a useful initial high-throughput screening tool for mechanistic studies with the caveat that not all Q2 droplets represent intact proviruses (Fig. 3i).

Analysis of proviral integration sites[14,15,21] of expanded clones demonstrated that the cultures were clonal and that integration into genes with cancer association was not required for proliferation (Fig. 4f, Supplementary Table 4). Instead, proliferative potential is related to proviral defects. nFGS confirmed the clonal nature of the proviruses in each well and demonstrated directly that the IPDA correctly identifies the presence and nature of defects in proviruses (Fig. 4g). Of 12 clonally expanded defective proviruses sequenced, 9 were defective in all HIV-1 ORFs, and none was fully competent for the expression of Tat or of Vif, Vpr, Vpu, Env or Nef. Low or absent expression of these genes would allow stimulated cells to escape viral cytopathic effects that could limit proliferation[22,23].

The failure of cells carrying intact proviruses to expand after repetitive in vitro T cell receptor stimulation provides insights into the mechanisms that drive in vivo clonal expansion[16–18,24] and into cure strategies that involve T cell activation. The reservoir of intact proviruses undergoes slow decay in vivo, as shown here and previously[12,13]. Thus, observed clonal expansions[14–19] must be more than offset by the death of infected cells. Strong T cell receptor stimulation may induce viral gene expression and cell death or impaired proliferation as shown in our in vitro experiments. Clonal expansion in vivo may be driven by stimuli including homeostatic cytokines that allow proliferation without virus production[19,20].

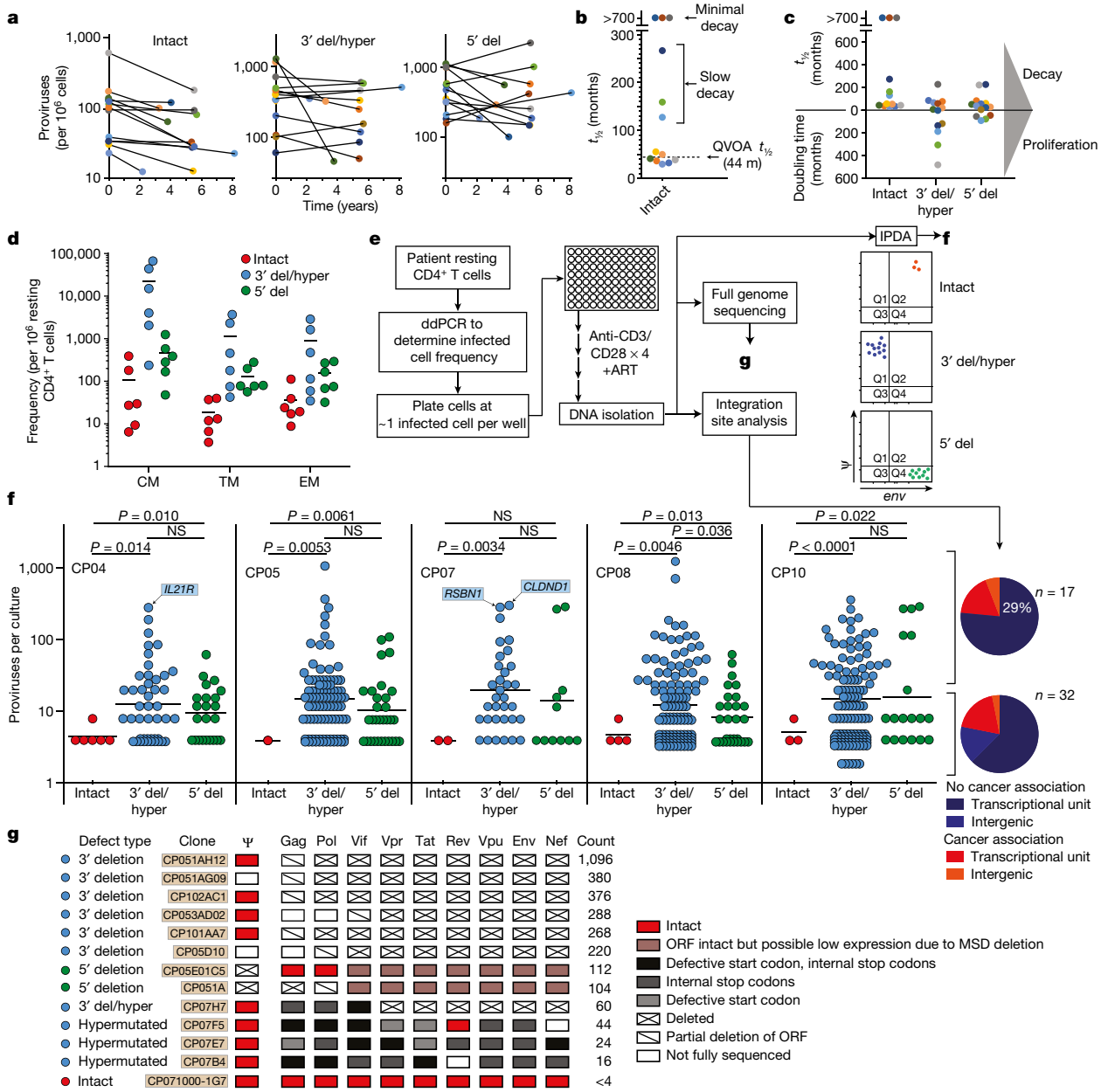


Fig. 4 | IPDA reveals differential dynamics of intact and defective proviruses. a, Decay of intact proviruses (left), proviruses with 3′ deletions and/or hypermutation (3′ del/hyper) (middle), and proviruses with 5′ deletions (5′ del) (right) measured in resting CD4$^+$ T cells from patients on long-term ART. **b**, Half-lives ($t_{1/2}$) of cells carrying intact proviruses assuming exponential decay, from the data in **a**. **c**, Half-lives of populations of cells carrying intact and defective proviruses. Increasing frequencies are plotted as doubling times. **d**, Distribution of intact and defective proviruses in memory CD4$^+$ T cell subsets. CM, central memory; EM, effector memory; TM, transitional memory. **e**, System for examining the proliferation of infected cell clones. Total proviruses in resting CD4$^+$ T cells from patients on ART were measured using a standard *gag* PCR corrected for *gag*-deleted proviruses. Cells were diluted to approximately 1 infected cell per well (2,000–4,000 total resting CD4$^+$ T cells per well) and stimulated 2–4 times with anti-CD3 and anti-CD28 microbeads in the presence of IL-2 and antiretroviral drugs (see Methods) resulting in expansion to an average of 2×10^6 cells per well. DNA was isolated for IPDA and for integration site analysis and nFGS, performed as previously described[8,26]. **f**, IPDA and integration site analysis of 1,731 microcultures from 5 patients. The *y* axis indicates the number of proviruses in each positive culture. Colours indicate type of provirus detected. Bars indicate geometric mean of proviruses/culture for each provirus type. Statistical significance of differences was determined by Welch's *t*-test. Integration sites for some clones with the highest proliferation are indicated in blue boxes. Pie charts indicate the fractions of integration sites in transcriptional units, intergenic regions, or with cancer associations for cultures with high (>20 proviruses, top) or low (<20 proviruses, bottom) proliferation. **g**, nFGS results for cultures showing proliferation. Sequences were analysed for defects affecting expression of each HIV-1 gene as described in Fig. 1b. Count indicates proviruses per culture as detected by IPDA.

Our results show that the small subset of proviruses with the potential to cause viral rebound show different dynamics than the vast excess of defective proviruses captured in standard PCR assays, emphasizing the importance of direct measurement of intact proviruses. The availability of a scalable assay for intact proviruses should accelerate cure research.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0898-8.

Received: 19 April 2018; Accepted: 4 January 2019;
Published online: 30 January 2019

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. Archin, N. M. et al. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* **487**, 482–485 (2012).
5. Borducchi, E. N. et al. Ad26/MVA therapeutic vaccination with TLR7 stimulation in SIV-infected rhesus monkeys. *Nature* **540**, 284–287 (2016).
6. Procopio, F. A. et al. A novel assay to measure the magnitude of the inducible viral reservoir in HIV-infected individuals. *EBioMedicine* **2**, 874–883 (2015).
7. Ho, Y. C. et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
8. Bruner, K. M. et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat. Med.* **22**, 1043–1049 (2016).
9. Imamichi, H. et al. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc. Natl Acad. Sci. USA* **113**, 8783–8788 (2016).
10. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
11. Jordan, A., Bisgrove, D. & Verdin, E. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**, 1868–1877 (2003).
12. 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).
13. Crooks, A. M. et al. Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J. Infect. Dis.* **212**, 1361–1365 (2015).
14. Maldarelli, F. et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
15. Wagner, T. A. et al. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).
16. Bui, J. K. et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog.* **13**, e1006283 (2017).
17. Lorenzi, J. C. et al. Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc. Natl Acad. Sci. USA* **113**, E7908–E7916 (2016).
18. Hosmane, N. N. et al. Proliferation of latently infected $CD4^+$ T cells carrying replication-competent HIV-1: potential role in latent reservoir dynamics. *J. Exp. Med.* **214**, 959–972 (2017).
19. Wang, Z. et al. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc. Natl Acad. Sci. USA* **115**, E2575–E2584 (2018).
20. Chomont, N. et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).
21. Cohn, L. B. et al. HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
22. Ho, D. D. et al. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126 (1995).
23. Wei, X. et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117–122 (1995).
24. 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).
25. Pollack, R. A. et al. Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. *Cell Host Microbe* **21**, 494–506 (2017).
26. Berry, C. C. et al. Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* **28**, 755–762 (2012).

Acknowledgements We thank D. Finzi of NIAID for discussions leading to this work. This work was supported by the NIH Martin Delaney I4C (UM1 AI126603), Beat-HIV (UM1 AI126620) and DARE (UM1 AI12661) Collaboratories, by NIH grant 43222, by the Howard Hughes Medical Institute and the Bill and Melinda Gates Foundation (OPP1115715), and by NIH SBIR grants R43AI124996 and R44AI124996 and NSF grants 1621633 and 1738428 to Accelevir Diagnostics. Samples for some study participants were obtained from the Baltimore-Washington DC Center of the Multicenter AIDS Cohort Study (MACS) supported by NIH grants U01-AI-35042 and UL1-RR025005 (ICTR).

Author contributions K.M.B., Z.W., G.M.L. and R.F.S. designed the study. K.M.B., Z.W., G.M.L., F.R.S., A.M.B., K.J.K., S.S., E.J.F., A.A.R.A., K.M.J., L.N.B., J.T.K. and Y.-C.H. performed experiments. K.M.B., Z.W., F.R.S., A.T., N.W., J.D.S., G.M.L. and R.F.S. analysed data. C.N., J.G. and F.D.B. provided integration site analysis. H.Z. performed cell sorting. S.A.B., A.A.C., J.B.M., J.N.B. and S.G.D. provided patient samples. K.M.B. and R.F.S. wrote the manuscript.

Competing interests Aspects of IPDA are subject of a patent application PCT/US16/28822 filed by Johns Hopkins University. K.M.B. and R.F.S. are inventors on this application. Accelevir Diagnostics holds an exclusive license for this patent application. G.M.L. is an employee of and shareholder in Accelevir Diagnostics. R.F.S. holds no equity interest in Accelevir Diagnostics. R.F.S. is a consultant on cure-related HIV research for Merck and Abbvie.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0898-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0898-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to R.F.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Study participants. Characteristics of study participants are given in Supplementary Table 3. The Johns Hopkins Institutional Review Board and the UCSF Committee on Human Research approved this study. All participants provided written consent before enrolment. Except where indicated, participants were HIV-1-infected adults on suppressive ART with undetectable plasma HIV-1 RNA levels (<50 copies per ml) for more than 6 months. Chronic phase (CP)-treated subjects are defined as subjects starting ART more than 180 days from the estimated date of infection. Acute phase-treated subjects started ART less than 100 days after the estimated date of infection. For longitudinal analysis, additional peripheral blood mononuclear cell (PBMC) samples were obtained from 10 HIV-1-infected men followed in the Baltimore-Washington DC centre of the Multicentre AIDS Cohort Study (MACS)[27] who had undetectable plasma HIV RNA (less than 20 copies per ml by Roche Taqman assay) at all semiannual study visits for 5 years or more, with no blips or missed visits. PBMC cryopreserved at visits at least 5 years apart, and viably stored as per MACS protocols, were studied. Characteristics of these 10 men are given in Supplementary Table 3 (CP31–CP49).

CD4$^+$ T cell isolation. PBMCs were isolated by density centrifugation using Ficoll-Paque PLUS (GE Healthcare Life Sciences) per the manufacturer's instructions. Untouched total CD4$^+$ T cells were then enriched from PBMCs using negative immunomagnetic selection using the EasySep Human CD4$^+$ T-Cell Enrichment Kit (StemCell Technologies). In some experiments, resting CD4$^+$ T cells (CD4$^+$, CD69$^-$, CD25$^-$ and HLA-DR$^-$) were isolated using a second negative selection step (CD25-Biotin; Anti-Biotin MicroBeads; CD69 MicroBead Kit II; Anti–HLA-DR MicroBeads, all from Miltenyi Biotec). Resting CD4$^+$ T cell purity was consistently greater than 95% as assessed using flow cytometry.

Bioinformatic analysis. We constructed an alignment of 431 published[8,9] full-length proviral sequences obtained from 28 HIV-1-infected adults by single genome analysis. Of these individuals, 19 started suppressive ART during chronic infection, and 9 started suppressive ART during acute infection. Twenty-four individuals had prolonged suppression of viraemia (>6 months) when studied, and five were viraemic when studied (one individual was studied at two time points, before and after ART). Clinical characteristics of these individuals are given in the relevant publications[8,9]. The proportion of different types of defects in HIV-1 proviruses varies with the stage of disease at which ART is started and the level of viraemia[8]. For most analyses, we used sequences from patients who initiated ART during chronic infection and who had sustained suppression of viraemia to below 50 copies HIV-1 RNA per ml of plasma for >6 months. These sequences ($n = 211$) are thus representative of the most common group of infected individuals who would be eligible for cure interventions. The proviral landscape in patients starting ART during acute infection is also dominated by defective sequences, with a higher fraction showing hypermutation[8]. In the analysis of deletions, common length polymorphisms were not included. Terminal deletions preclude integration and are not seen. Internal deletions in the 5′ LTR are not evaluable because the nFGS methods do not capture this region. Hypermutation occurred with or without deletions. For analyses of specific types of defect, we used all sequences in the database with that defect. For example, for analysis of hypermutation, we used all database sequences with hypermutation ($n = 100$) and additional hypermutated sequences obtained by single genome *env* sequencing. Hypermutation in the GG or GA context was confirmed using the Hypermut algorithm[28]. Positions of primers for ddPCR analysis were evaluated using a sliding window analysis of two hypothetical 100-base-pair amplicons, scoring for lack of significant overlap (>5%) with mapped deletions in the database sequences with deletions. Optimal discrimination between intact and deleted sequences is obtained with a 5′ amplicon in the Ψ region and a 3′ amplicon in *env*. Ψ is the site of frequent small deletions and is included in larger 5′ deletions. Sequence conservation was evaluated using separate alignments of US clade B sequences from the Los Alamos HIV sequence database (https://www.hiv.lanl.gov/).

IPDA. A variety of experimental conditions were used in the development of the IPDA. The following is an optimized recommended procedure. In general, the IPDA is performed on DNA from 5×10^6 CD4$^+$ T cells. Genomic DNA is extracted using the QIAamp DNA Mini Kit (Qiagen) with precautions to avoid excess DNA fragmentation. DNA concentrations are determined using the Qubit3.0 and Qubit dsDNA BR Assay Kit (ThermoFisher Scientific). Quantification of intact, 5′ deleted, and 3′ deleted and/or hypermutated proviruses is carried out using primer/probe combinations optimized for subtype B HIV-1 (Accelevir Diagnostics). The primer/probe mix consists of oligonucleotides for two independent hydrolysis probe reactions that interrogate conserved regions of the HIV-1 genome to discriminate intact from defective proviruses (Supplementary Table 5). HIV-1 reaction A targets the packaging signal (Ψ) that is a frequent site of small deletions and is included in many large deletions in the proviral genome. The Ψ amplicon is positioned at HXB2 coordinates 692–797. This reaction uses forward and reverse primers, as well as a 5′ 6-FAM-labelled hydrolysis probe. Successful amplification of HIV-1 reaction A produces FAM fluorescence in droplets containing Ψ, detectable in channel 1 of the droplet reader. HIV-1 reaction B targets the RRE of the proviral genome, with the amplicon positioned at HXB2 coordinates 7736–7851. This reaction uses forward and reverse primers, as well as two hydrolysis probes: a 5′ VIC-labelled probe specific for wild-type proviral sequences, and a 5′ unlabelled probe specific for APOBEC-3G hypermutated proviral sequences (Supplementary Fig. 2). Successful amplification of HIV-1 reaction B produces a VIC fluorescence in droplets containing a wild-type form of RRE, detectable in channel 2 of the droplet reader, whereas droplets containing a hypermutated form of RRE are not fluorescent.

Droplets containing HIV-1 proviruses are scored as follows. Droplets positive for FAM fluorescence only, which arises from Ψ amplification, score as containing 3′ defective proviruses, with the defect attributable to either APOBEC-3G-mediated hypermutation or 3′ deletion. Droplets positive for VIC fluorescence only, which arises from wild-type RRE amplification, score as containing 5′ defective proviruses, with the defect attributable to 5′ deletion. Droplets positive for both FAM and VIC fluorescence score as containing intact proviruses (Fig. 3a, b). Double-negative droplets contain no proviruses or rare proviruses (approximately 3.8%) with defects affecting both amplicons.

An important aspect of the quantification of intact proviruses is correction for DNA shearing between amplicons, which artificially reduces Q2 droplets while increasing Q1 and Q4 droplets (Fig. 3b). This can be done accurately through ddPCR analysis of a host gene, which also provides a measure of input cell number. In principle, any host gene can be used, with two ddPCR amplicons spaced at the same distance as the HIV-1 amplicons described above. As demonstrated in Fig. 3d, e, this approach allows for accurate correction for DNA shearing. Simultaneous quantification of DNA shearing and input human genome equivalents is performed using another aliquot of the same DNA sample. For the studies described here, oligonucleotides for two independent hydrolysis probe reactions that interrogate the human *RPP30* gene (chromosome 10: 90,880,081 on GRCh38) were used (Accelevir Diagnostics). In this primer/mix, the total distance between duplex reactions and individual amplicon sizes are equivalent to those of the HIV-1 proviral discrimination reactions described above. Optimal reaction melting temperatures are equivalent for all reactions. Comprehensive oligonucleotide analysis predicted no potential off-target nucleotide binding or amplification (BLASTn) and no self-dimerization, hetero-dimerization, or primer hairpin formation. *RPP30* reaction A uses a 5′ 6-FAM-labelled hydrolysis probe, and *RPP30* reaction B uses a 5′ HEX-labelled hydrolysis probe. Droplets positive for both 6-FAM and HEX fluorescence score as containing an unsheared *RPP30* gene fragment, whereas droplets positive for only a single fluorescence score as containing part of a sheared *RPP30* gene fragment. The ratio of dual fluorescent to single fluorescent droplets is used to calculate a DSI, and this index is applied to both genome copy number input reactions and HIV-1 proviral discrimination reactions to correct for experimentally observed DNA shearing.

ddPCR was performed on the Bio-Rad QX200 AutoDG Digital Droplet PCR system using the appropriate manufacturer supplied consumables and the ddPCR Supermix for probes (no dUTPS) (Bio-Rad Laboratories). For HIV-1 proviral discrimination reactions, 700 ng of genomic DNA was analysed in each reaction well. For DNA shearing and copy number reference reactions, 7 ng of genomic DNA was analysed in each reaction well. Multiple replicate wells were performed for each reaction type to ensure consistent quantification, and replicate wells were merged during analysis to increase IPDA dynamic range. The thermal cycling program used for all reactions, with a 2 °C ramp rate, is given in Supplementary Table 6. In general, parallel processing and analysis of uninfected donor CD4$^+$ T cells was performed for each IPDA run as a negative control while qualified JLat6.3 genomic DNA was analysed in each IPDA run as a positive control. After correction for DNA shearing, results are expressed as proviral copies per 10^6 CD4$^+$ T cells.

Cell lines. The JLat full-length clone (clone 6.3) from E. Verdin was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH. Cells were authenticated through short tandem repeat analysis and tested negative for mycoplasma.

QVOA. The QVOAs were performed as previously described[1,29]. MOLT-4/CCR5 cells[30] were added on day 2 of the culture and the culture supernatants were examined for the p24 viral capsid protein by ELISA (PerkinElmer) after 14 and 21 days. Results were expressed as infectious units per million cells (IUPM) CD4$^+$ T cells calculated using maximal likelihood as described by previously (IUPMStats[31]).

T cell subset analysis. Resting CD4$^+$ T cells from HIV-1-infected donors on suppressive ART were isolated as described above. To sort resting memory subsets, we incubated cells with FcgR block (BD Pharmingen) for 10 min before staining with a FITC-labelled antibody to CD3 (Biolegend; clone HIT3a), phycoerythrin (PE)-Cy7-labelled antibody to CD4 (Biolegend; clone RPA-T4), allophycocyanin (APC)-labelled antibody to CD45RO (Biolegend; clone UCHL1), BV421-labelled antibody to CD27 (Biolegend; clone O323) and PE-labelled antibody to CCR7 (Biolegend; clone G043H7). Dead cells were excluded using propidium iodide.

PE mouse IgG2aκ, BV421 mouse IgG1κ, and APC IgG2aκ isotype antibodies were used in fluorescence-minus-one controls to set sorting gates. Memory cells were distinguished from naive cells by the presence or absence of CD45RO staining, respectively. Central memory cells were distinguished from effector and transitional memory subsets by the presence of CCR7 as described previously[32]. $CCR7^-$ cells were subdivided into effector memory (T_{EM}) ($CD45RO^+CCR7^-$) or transitional memory (T_{TM}) ($CD45RO^+CCR7^-CD27^+$) cells as described previously[20]. Central, effector and transitional resting memory subsets were sorted using a Beckman Coulter MoFlo XDT Cell sorter.

Clonal microcultures. Purified resting $CD4^+$ T cells from HIV-1-infected donors on suppressive ART were analysed for proviral DNA copies using *gag* quantitative PCR as previously described[33]. The resulting values were corrected for gag^- proviruses, and cells were plated at approximately one infected cell per well (2,000–4,000 total resting $CD4^+$ T cells) in 96-well plates. The cells were then stimulated with anti-CD3/CD28 Dynabeads (25 μl per million cells; Thermo Fisher Scientific) in RPMI containing 10% fetal bovine serum and 100 U ml^{-1} IL-2 (Novartis) for 7 days in the presence of tenofovir disoproxil fumarate (10 μM) and emtricitabine (10 μM). Half of the medium in each well was removed and replaced with fresh medium, anti-CD3/CD28 Dynabeads and antiretroviral drugs weekly for 2–3 weeks. DNA isolation was performed on cells from each well using a Quick-DNA 96 Kit (Zymo Research Corporation). One-quarter of the extracted DNA was analysed by the IPDA to determine the type of provirus and proviral copy number in each well. The remaining DNA was used for integration site analysis and full genome sequencing.

Integration site analysis. Integration site analysis was performed using previously described linker ligation method[26,34,35]. Sites for which both the 5′ and 3′ junctions were captured are shown. Cancer associations were determined as previously described[35].

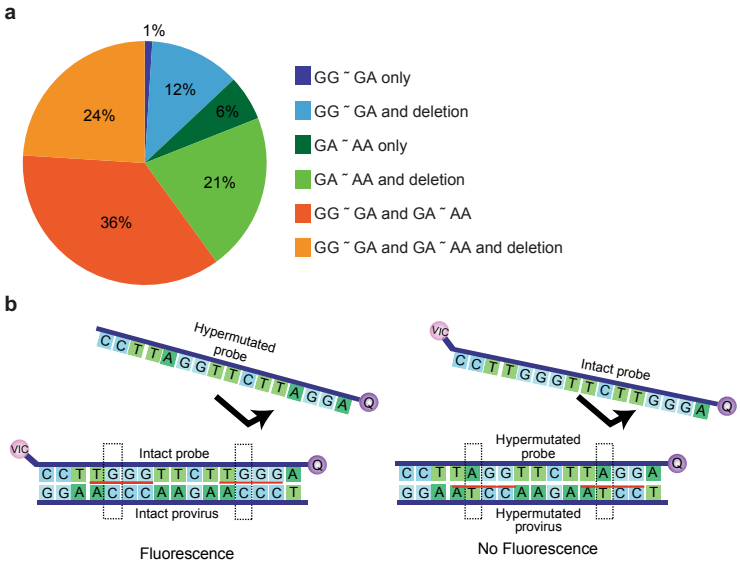
Full genome sequencing. Full genome sequencing was performed at the single molecule level as previously described[8].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

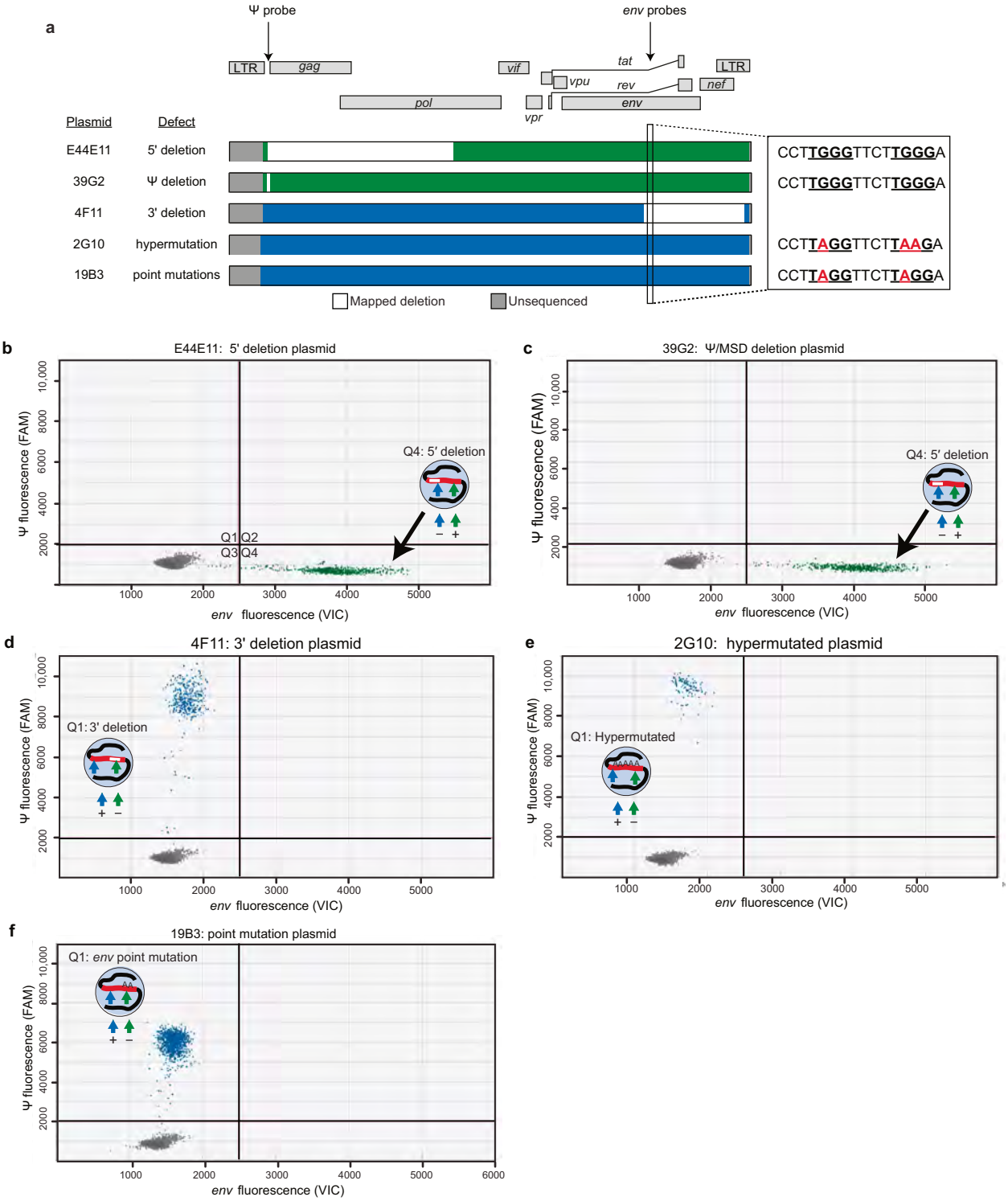
Data availability

The IPDA was developed through an analysis of published near full genome HIV-1 sequences (refs [8,9], GenBank accession numbers KX505390–KX505744 and KU677989–KU678196, respectively). All other data are available from the corresponding author upon reasonable request.

27. Detels, R. et al. The multicenter AIDS Cohort Study, 1983 to … *Public Health* **126**, 196–198 (2012).
28. Rose, P. P. & Korber, B. T. Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics* **16**, 400–401 (2000).
29. Laird, G. M., Rosenbloom, D. I., Lai, J., Siliciano, R. F. & Siliciano, J. D. Measuring the frequency of latent HIV-1 in resting $CD4^+$ T cells using a limiting dilution coculture assay. *Methods Mol. Biol.* **1354**, 239–253 (2016).
30. Laird, G. M. et al. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398 (2013).
31. Rosenbloom, D. I. 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).
32. Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).
33. Durand, C. M. et al. 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 (2012).
34. Lewinski, M. K. et al. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**, 6610–6619 (2005).
35. Sherman, E. et al. INSPIIRED: a pipeline for quantitative analysis of sites of new DNA integration in cellular genomes. *Mol. Ther. Methods Clin. Dev.* **4**, 39–49 (2017).

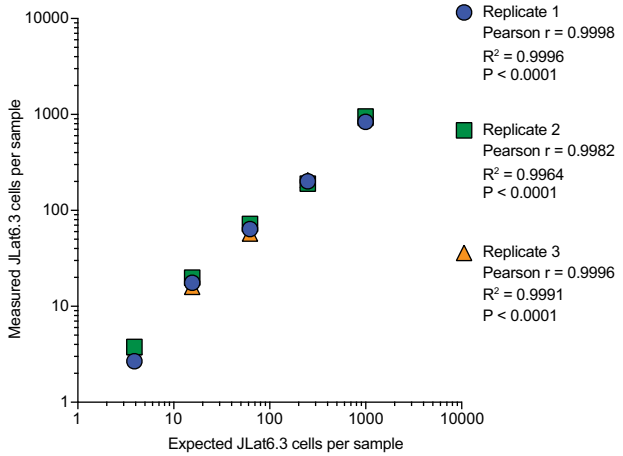


Extended Data Fig. 1 | Analysis of hypermutation. a, Most hypermutated proviruses show both GG→AG and GA→AA patterns of hypermutation. Analysis based on hypermutated full-genome sequences in the database ($n = 100$). Sequences were analysed for GG→AG and GA→AA patterns using all available sequences for each clone and the Los Alamos Hypermut algorithm[17]. **b**, Hypermutation discrimination using two probes in the RRE of the *env* gene. The intact probe hybridizes with a region containing two adjacent APOBEC3G consensus sites (red underline) in intact proviruses. It is labelled with a fluorophore (VIC) and a quencher (Q). Also present in the reaction is a hypermutated probe that lacks the fluorophore and does not bind (arrow) to intact proviral sequences owing to G→A mutations at both APOBEC3G consensus sites. Dashed boxes indicate the nucleotide positions of sequence differences between the intact and hypermutated probes. The hypermutated probe preferentially binds to the same region in hypermutated proviruses. It lacks a fluorophore and prevents binding of the fluorophore-labelled intact probe (arrow). Therefore, no fluorescent signal is generated for 95% of hypermutated proviruses (Fig. 2f).



Extended Data Fig. 2 | Plasmid controls show the specificity of the IPDA. a, Maps of proviral plasmid control templates. Plasmids E44E11, 39G2 and 4F11 have deletions in the indicated regions (white). Plasmid 2G10 is a heavily hypermutated patient-derived sequence with G→A mutations in the probe-binding region of the *env* amplicon (enlarged region). Plasmid 19B3 has GA point mutations in this region. These G→A mutations (red) occur at two APOBEC3G consensus sites (TGGG, underlined) in this region. These plasmids have been previously described[12,34]. **b**–**e**, IPDA on plasmids representing the indicated defective proviruses showing positive droplets only in the expected quadrants.

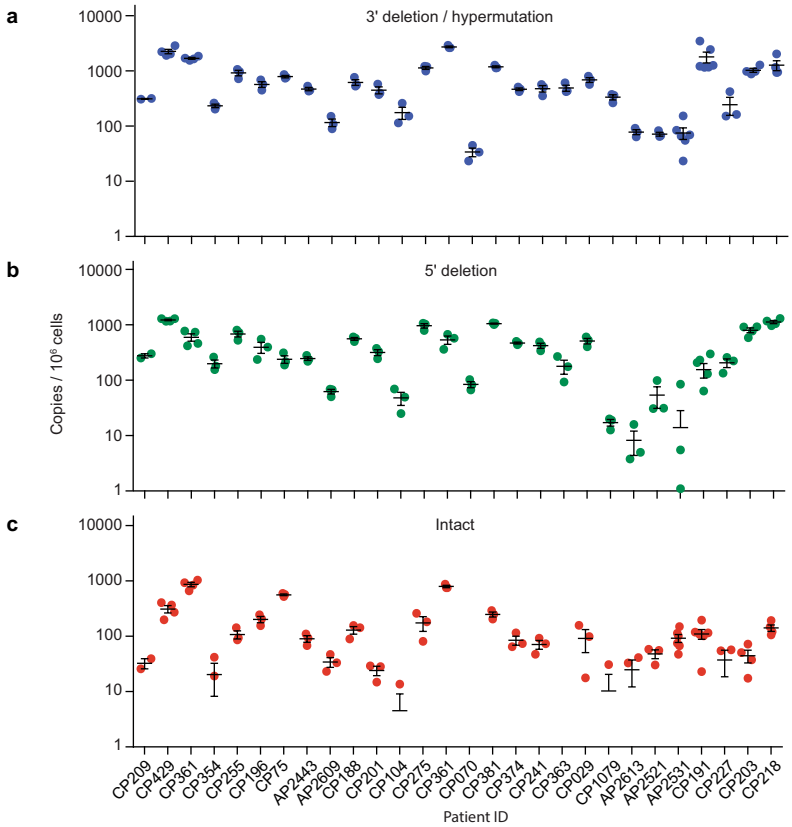
a



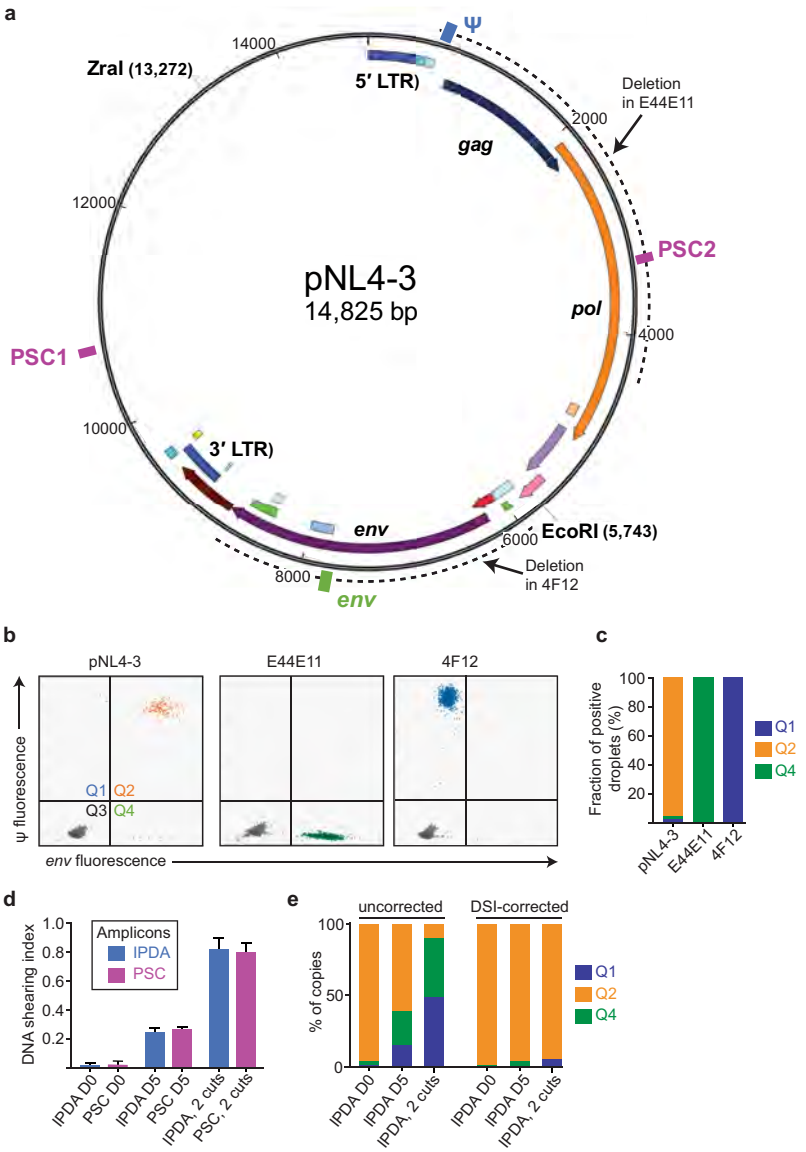
b

Expected intact proviruses per 10^6 $CD4^+$ T cells	Intact Proviruses per 10^6 input $CD4^+$ T-cells by IPDA					
	Run 1	Run 2	Run 3	Average	StDev	% CV
998.0	836.4	953.8	925.1	905.1	61.2	7
249.5	200.7	190.2	206.0	199.0	8.0	4
62.4	64.0	72.3	57.5	64.6	7.4	12
15.6	17.6	19.9	16.0	17.8	1.9	11
3.9	2.7	3.8	3.8	3.4	0.6	18

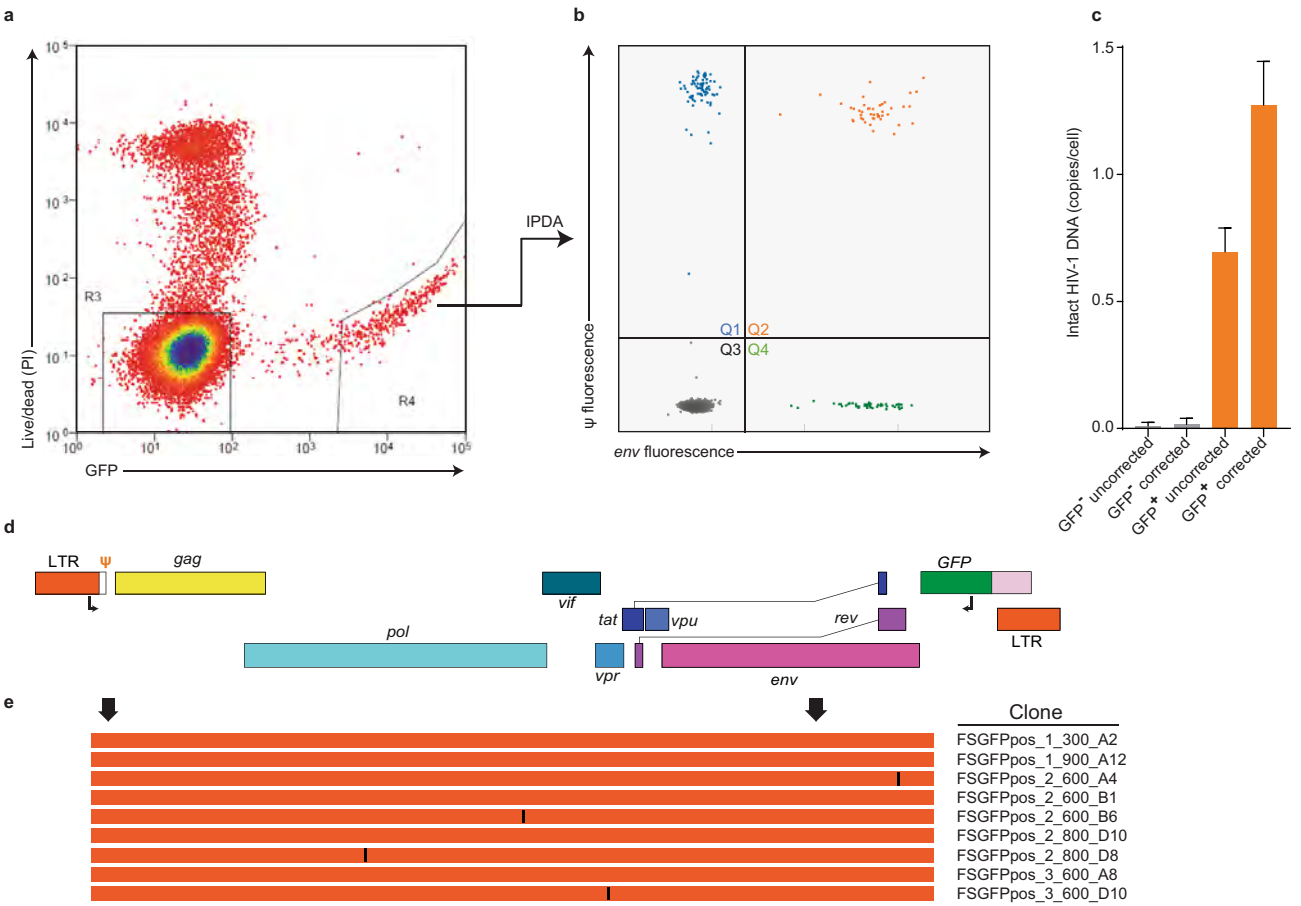
Extended Data Fig. 3 | IPDA accuracy, reproducibility and limit of quantification. a, Correlation between expected and IPDA-measured frequencies of intact proviruses per 10^6 cells. Genomic DNA from uninfected donor $CD4^+$ T cells was spiked with JLat6.3 DNA cell equivalents and subjected to a serial fourfold dilution. This material was then analysed by the IPDA, and the IPDA-measured frequencies of intact proviruses per million cells were compared to the expected frequencies (998, 249.5, 62.4, 15.6 and 3.9 intact proviruses per 10^6 $CD4^+$ T cells). This experiment was performed independently three times. The agreement between the expected and IPDA-measured frequencies of intact proviruses was determined using Pearson correlation. **b**, Reproducibility of the IPDA across independent assay runs. Reproducibility was assessed by determining the coefficient of variation (CV) across three independent assay measurements of genomic DNA from uninfected donor $CD4^+$ T cells spiked with JLat6.3 cell equivalents and subject to serial fourfold dilution, as described in **a**.



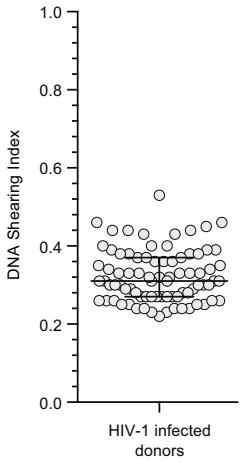
Extended Data Fig. 4 | IPDA reproducibility. a–**c**, Frequencies of cells containing proviruses with 3′ deletions and/or hypermutation (**a**), 5′ deletions (**b**), or no defects (intact; **c**) in $CD4^+$ T cells from 28 treated patients. Each data point represents a replicate IPDA determination from a single sample from the indicated patient. The mean and s.e.m. of the replicates are plotted. The variability between patients is much greater than the variation between replicates from a single patient. Technical replicates are shown to indicate low intrinsic variability of the IPDA.



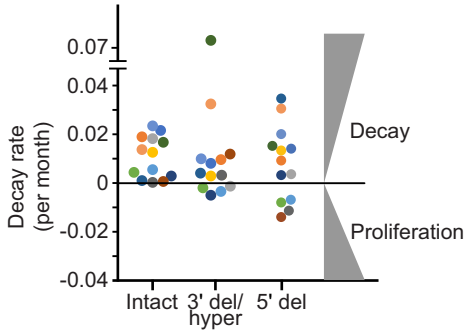
Extended Data Fig. 5 | Plasmid controls confirm specificity of the IPDA. a, Map of the plasmid pNL4-3 carrying an intact HIV-1 provirus. Positions of the Ψ (blue) and *env* (green) IPDA amplicons and of a distinct set of plasmid shearing control (PSC, magenta boxes) amplicons are indicated. Spacing between PSC amplicons is equal to spacing between Ψ and *env* amplicons. Dotted lines indicate positions of deletions in plasmids carrying previously described[12] defective proviruses E44E11 and 4F12 with 5′ and 3′ deletions, respectively. **b**, IPDA analysis of the indicated ZraI-cut plasmids representing intact, 5′-deleted and 3′-deleted proviruses. **c**, Summary of droplet counts for the experiment shown in **b**. E44E11 and 4F12 give positive droplets only in quadrants 4 and 1 (Q4 and Q1), respectively. For pNL4-3, more than 95% of droplets are in Q2, with the remainder attributable to shearing between the Ψ and *env* amplicons. **d**, Analysis of shearing. For IPDA analysis of patient samples, shearing was measured using amplicons in the *RPP30* gene (Fig. 3a, d, e). For plasmid control experiments, shearing of ZraI-cut plasmids was analysed using two sets of amplicons, the Ψ and *env* IPDA amplicons and the equally spaced PSC amplicons shown in **a**. ddPCR analysis was done on fresh (D0) maxipreps of pNL4-3 linearized with ZraI at a concentration mimicking patient samples. To assess the effects of higher levels of DNA fragmentation, IPDA analysis was also done on pNL4-3 DNA that had been incubated at 4 °C for 5 days (D5), and on pNL4-3 DNA cut with both ZraI and EcoRI (two cuts). The mean and range of duplicate determinations of the DSI is shown for each set of amplicons. The DSI was the same for the IPDA and PSC amplicons at three different levels of shearing. The DSI was used to correct the IPDA droplet counts in **e**. Negative values were set to 0. **e**, Uncorrected and DSI-corrected IPDA analysis of the intact proviral construct pNL4-3 at different levels of fragmentation. After correction, positive droplets were almost exclusively in Q2 even at higher levels of fragmentation.



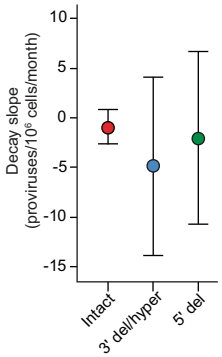
Extended Data Fig. 6 | Sequence analysis of Q2 proviruses. a, Sorting of productively infected CD4$^+$ T cells. Cell preparations with a high fraction of intact proviruses were obtained by infecting CD4$^+$ T lymphoblasts with a replication-competent HIV-1 carrying GFP in the *nef* ORF (R7-GFP37). After 48 h, GFP$^+$ cells were collected by sorting. Genomic DNA was isolated, subjected to pulse field electrophoresis to remove unintegrated intermediates, and analysed by IPDA. **b**, IPDA analysis of high molecular mass DNA from sorted cells. Droplets in Q1 and Q4 largely reflect the shearing of intact proviruses (DSI = 0.46) during in DNA isolation and purification. **c**, Frequency of intact proviruses in GFP$^-$ and GFP$^+$ cells before and after correction for shearing. After correction for shearing, the frequency of intact proviruses in sorted GFP$^+$ cells is close to the expected value of 1. **d**, Map of the HIV-1 genome in GFP-expressing HIV-1 vector R7-GFP used in **a**. GFP is inserted in the *nef* ORF. Positions of outer primers in the LTR and GFP used in single genome amplifications are indicated. **e**, Sequence analysis of nine independent single genomes. Arrows indicate positions of the Ψ and *env* IPDA amplicons. Orange lines indicate intact sequence without deletions or hypermutation and identical to R7-GFP except for single base mutations (black lines).



Extended Data Fig. 7 | DSI for patient samples. The DSI was determined by ddPCR using two amplicons in a cellular gene (*RPP30*) spaced at exactly the same distance as the Ψ and *env* amplicons. It is the fraction of templates in which DNA shearing has occurred between the amplicons. Horizontal bars indicate median and interquartile range; data from $n = 62$ patient samples.



Extended Data Fig. 8 | In vivo decay rates of cells with intact and defective proviruses. The frequency of cells carrying intact proviruses, proviruses with 3′ deletion and/or hypermutation (3′ del/hyper), and proviruses with 5′ deletions (5′ del) was measured in resting $CD4^+$ T cells from patients on long-term suppressive ART. Data are plotted in terms of decay rate assuming exponential decay. Half-life values for the same decay curves are shown in Fig. 4c. Negative decay rate indicates proliferation.



Extended Data Fig. 9 | Variability in decay slopes. Shown are the mean and s.d. of the decay slopes for intact and defective proviruses in infected individuals on ART sampled longitudinally ($n = 14$). Analysis based on decay data in Fig. 4a.

nature research

Corresponding author(s): Robert Siliciano

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used in data collection.
Data analysis	Sequence data were analyzed with alignments generated manually with Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The IPDA was developed through an analysis of published near full genome HIV sequences. These sequences are described in Bruner et al. (reference 13 in the revised manuscript, Genebank accession numbers KX505390-KX505744) and Imamichi et al. (reference 14 in the revised manuscript, Genbank accession numbers KU677989-KU678196. There are no restrictions on the availability of these data.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sequence analysis was performed on a database of all available full genome HIV sequences meeting study criteria (n=431).
Data exclusions	No sequences were excluded but specific analyses were carried out on sequences from the relevant patient populations (ie. viremic or aviremic, acute or chronic) depending on the research question as described in the text.
Replication	Assay reproducibility is described in Extended Data Figures 3 and 4.
Randomization	This was not a clinical trial.
Blinding	This was not a clinical trial.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The J-Lat clone (#6.3) was obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH.
Authentication	The cells were shown the contain the expected 1 copy of HIV / cell as described in the text.
Mycoplasma contamination	The cells were negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No misidentified cells were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Blood samples were obtained from HIV-1 infected adults on suppressive ART with undetectable plasma HIV-1 RNA levels (< 50 copies per mL) for >6 months.
Recruitment	Patients were recruited through care providers at the Moore Clinic (JH) or the Scope Study (UCSF). All participants provided written informed consent.

nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-019-0898-8

In the format provided by the authors and unedited.

A quantitative approach for measuring the reservoir of latent HIV-1 proviruses

Katherine M. Bruner[1,8,10], Zheng Wang[1,10], Francesco R. Simonetti[1], Alexandra M. Bender[1], Kyungyoon J. Kwon[1], Srona Sengupta[1], Emily J. Fray[1], Subul A. Beg[1], Annukka A. R. Antar[1], Katharine M. Jenike[1], Lynn N. Bertagnolli[1], Adam A. Capoferri[1], Joshua T. Kufera[1], Andrew Timmons[1], Christopher Nobles[2], John Gregg[2], Nikolas Wada[3], Ya-Chi Ho[1,9], Hao Zhang[4], Joseph B. Margolick[4], Joel N. Blankson[1], Steven G. Deeks[5], Frederic D. Bushman[2], Janet D. Siliciano[1], Gregory M. Laird[6] & Robert F. Siliciano[1,7]*

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [4]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [5]Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [6]Accelevir Diagnostics, Baltimore, MD, USA. [7]Howard Hughes Medical Institute, Baltimore, MD, USA. [8]Present address: Department of Molecular Biosciences, University of Texas, Austin, TX, USA. [9]Present address: Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, CT, USA. [10]These authors contributed equally: Katherine M. Bruner, Zheng Wang. *e-mail: rsiliciano@jhmi.edu

Supplemental Material

A novel quantitative approach for measuring the reservoir of latent HIV-1 proviruses

Katherine M. Bruner[1*†], Zheng Wang[1*], Francesco R. Simonetti[1], Alexandra M. Bender[1], Kyungyoon J. Kwon[1], Srona Sengupta[1], Emily J. Fray[1], Subul A. Beg[1], Annukka A. R. Antar[1], Katharine M. Jenike[1], Lynn N. Bertagnolli[1], Adam A. Capoferri[1], Joshua T. Kufera[1], Andrew Timmons[1], Christopher Nobles[2], John Gregg[2], Nikolas Wada[3], Ya-Chi Ho[1‡], Hao Zhang[4], Joseph B. Margolick[4], Joel N. Blankson[1], Steven G. Deeks[5], Frederic D. Bushman[2], Janet D. Siliciano[1], Gregory M. Laird[6], Robert F. Siliciano[1,7]

Table S1. Most defective HIV-1 proviruses detected by nFGS are defective in most HIV-1 genes.

Gene or element	Fraction of defective proviruses with indicated defect[a]						Fraction of defective proviruses with no defects in:	
	Deletion[b]	Deleted or mutant start codon	Internal stop codon	Splicing defect[d]	Any ORF defect[e]	Any defect[f]	ORF[g]	Gene[h]
Ψ	0.25	NA[c]	NA	NA	NA	0.25	NA	0.75
gag	0.38	0.30	0.32	NA	0.66	0.66	0.34	0.34
pol	0.79	0.30	0.30	NA	0.85	0.85	0.15	0.15
vif	0.76	0.75	0.06	0.88	0.81	0.95	0.19	0.05
vpr	0.71	0.72	0.05	0.90	0.76	0.94	0.24	0.06
tat	0.74	0.75	0.04	0.92	0.80	0.97	0.20	0.03
rev	0.74	0.73	0.11	0.92	0.81	0.96	0.19	0.04
vpu	0.73	0.73	0.04	0.91	0.77	0.94	0.23	0.06
env	0.73	0.73	0.25	0.91	0.79	0.96	0.21	0.04
nef	0.51	0.58	0.32	0.92	0.61	0.97	0.39	0.03
LTR	0.34	NA	NA	NA	NA	0.41	NA	0.59

[a]Fraction of defective proviruses with specific types of defects in the indicated gene or element. Based on bioinformatic analysis of 211 near full genome sequences from 19 patients on suppressive ART regimens who initiated ART during chronic infection (see Methods). From this set, 206 defective proviruses were identified. Note that some proviruses have more than one type of defect.

[b]Fraction of proviruses with a deletion comprising more than 5% of the gene and not attributable to normal length polymorphism.

[c]NA, not applicable.

[d]Fraction of defective proviruses with deletions or mutations affecting canonical splice donor and/or acceptor sites for the dominant mRNA species encoding the relevant viral protein as identified in Ocwieja et al[1].

[e]Fraction of defective proviruses with any defect in the indicated ORF including deletions of >5% of standard gene length, deleted or mutant start codon, or internal stop codons.

[f]Fraction of defective proviruses with any defect in the indicated gene including deletions of >5% of standard gene length, deleted or mutant start codon, internal stop codons, or missing or mutant splice donor or acceptor sites.

[g]Fraction of defective proviruses with defects within the indicated ORF (excluding splice site deletions/mutations).

[h]Fraction of defective proviruses with defects in the indicated gene including splice site deletions/mutations.

Table S2. Standard PCR assays for proviral DNA

Region[a]	Type[b]	Amplicon[c]	Reference
LTR (1)	qPCR	452-793	2
LTR (2)	qPCR	522-642	3
LTR-*gag*	AluPCR	1-1505	4
ψ	ddPCR	697-797	This work
gag	ddPCR	1359-1500	5
pol (1)	qPCR	2536-2662	6
pol (1)	ddPCR	4809-4974	5
env	ddPCR	7736-7851	This work
LTR	AluPCR	9585-9719	7

[a]Region of genome targeted. See **Fig. 1c** for map. Numbers in parentheses distinguish distinct assays targeting the same region.

[b]Type of PCR assay. qPCR, quantitative PCR; ddPCR, digital droplet PCR.

[c]Amplicon position in HXB2 coordinates.

Table S3. Characteristics of study participants

Partici-pant ID	Age	Sex	Race[a]	ART Regimen[b]	CD4 count[c] (cells/μl)	Plasma HIV-1 RNA[c] (copies/ml)	Time on ART[d] (mo)	Time on ART with HIV RNA <50[e] (mo)	CP or EC[f]
AP01	44	M	AA	ABC/3TC/DTG	687	<20	121	115	CP
AP02	36	F	AA	EVG/TAF/FTC/COBI	840	<20	78	77	CP
AP04	52	M	W	EVG/TDF/FTC/COBI	NA	<20	79.0	70	CP
AP05	44	F	AS	ABC/3TC/DTG	NA	<20	41.3	40	CP
AP07	47	M	W	FTC/TDF, DTG	NA	<20	15.2	8	CP
AP08	40	M	W	FTC/TDF, RAL	NA	<20	32.1	28	CP
AP09	33	M	W	ABC/3TC/DTG	NA	<20	24.3	20	CP
AP10	31	M	AS	ABC/3TC/DTG	NA	<20	14.8	13	CP
CP01	63	M	AA	ETR, FTC, RAL	907	<20	101	100	CP
CP02	61	M	AA	ABC/3TC, DRV/r, RAL	423	<20	46	44	CP
CP03	76	M	W	ABC/3TC/ZDV, ATZ/r	635	<20	192	177	CP
CP04	49	F	AA	DRV/r, EFV/FTC/TDF	890	<20	210	55	CP
CP04	51	F	AA	EVG/TAF/FTC/COBI	582	<20	229	74	CP
CP05	55	F	AA	ABC, FTC, ETR, RAL	441	<20	199	104	CP
CP06	63	M	W	ABC, 3TC, RAL, ETR	997	<20	253	50	CP
CP07	36	M	AA	EFV/FTC/TDF	NA	<20	61	55	CP
CP07	40	M	AA	ABC/3TC/DTG	976	<20	108	102	CP
CP08	47	M	AA	ETR, MVC, RAL	829	<20	163	78	CP
CP09	50	M	AA	3TC, DRV/r, RAL	1635	<20	143	101	CP
CP10	58	M	AA	ETR, DRV/r, RAL	570	<20	>208	120	CP
CP11	53	M	W	ABC/3TC/DTG	912, 1214	<20	43	30	CP
CP12	54	M	AA	3TC/ZDV, EFV, ATV/r	586	<20	264	135	CP
CP13	70	M	W	EFV/FTC/TDF	665	<20	224	161	CP
CP14	56	M	AA	EFV, DRv/r, RAL	1011	<20	187	179	CP
CP15	55	M	AA	EFV, FTC/TAF	1550	<20	130	112	CP
CP16	53	M	AA	FTC/TAF, RAL	788	<20	108	62	CP
CP17	57	F	AA	ABC/3TC/DTG	802	<20	107	92	CP
CP18	49	F	AA	EVG/TAF/FTC/COBI	1086	<20	52	45	CP
CP19	53	F	AA	None	535	2130	N/A[g]	N/A[g]	CP
CP20	49	F	AA	ABC/3TC/DTG	623	<20	107	98	CP
CP21	38	F	AA	ABC/3TC/DTG	1098	<20	64	60	CP
CP22	57	F	AA	DRV/r, MVC, RAL	976	<20	92	90	CP
CP23	56	M	AA	FTC/TAF, DTG	670	<20	96	31	CP

Partici-pant ID	Age	Sex	Race[a]	ART Regimen[b]	CD4 count[c] (cells/μl)	Plasma HIV-1 RNA[c] (copies/ ml)	Time on ART[d] (mo)	Time on ART with HIV RNA <50[e] (mo)	CP or EC[f]
CP24	58	F	AA	ETR, DRV/r, RAL	752, 1146	<20	92	38	CP
CP25	67	F	AA	FTC/RPV/ TAF	482	<20	153	142	CP
CP26	44	M	AA	EFV/FTC/TDF	495 454	<20	97	94	CP
CP27	63	M	AA	EFV/FTC/TDF	502	<20	87	73	CP
CP28	50	M	AA	DRV/COBI, DTG	553	<20	121	111	CP
CP29	60	M	AA	EFV, ABC/3TC	452	<20	189	151	CP
CP30	64	M	W	ETR, 3TC, TDF	NA[h]	<20	NA	127	CP
CP31	64	M	W	NNRTI regimen	1750, 1748	<50, <20	NA	72, 170	CP
CP33	67	M	W	NNRTI regimen	659, 721	<50, <20	NA	45, 114	CP
CP35	69	M	AA	PI regimen	371, 635	<50, <20	NA	8, 74	CP
CP37	74	M	W	PI regimen	618, 771	<50, <50	NA	19, 85	CP
CP39	71	M	W	NNRTI+PI regimen	529, 435	<50, <20	NA	91, 151	CP
CP41	74	M	W	NNRTI regimen	NA, 739	<50, <50	NA	65, 130	CP
CP43	61	M	AA	PI regimen	760, 484	<50, <20	NA	74, 140	CP
CP45	49	M	AA	PI regimen	296, 655	<50, <20	NA	105, 171	CP
CP47	40	M	W	NNRTI regimen	715, 768	<20, <20	NA	47, 108	CP
CP49	49	M	W	NNRTI regimen	599, 931	<20, <20	NA	58, 117	CP
ES24	64	M	AA	None	1500	< 20	N/A	N/A	EC
ES3	65	F	AA	None	1149	<20	N/A	N/A	EC
ES31	64	F	AA	None	744	<20	N/A	N/A	EC
ES46	50	M	AA	None	612	<20	N/A	N/A	EC
ES5	65	F	AA	None	735	<20	N/A	N/A	EC

[a]AA, African American; AS, Asian; W, white.

[b]Standard drug abbreviations are used. "/" indicates drugs combined into a single pill. For some cryopreserved samples from MACS cohort participants, regimen details are not available an information on whether the regimen contained a non-nucleoside reverse transcriptast inhibitor (NNRTI) or a protease inhibitor (PI).

[c]At time of sampling.

[d]Total time on ART.

[e]Time on ART with plasma HIV-1 RNA <50 copies/ml.

fChronic progressor (CP) or elite controller (EC, infected individual who maintains HIV-1 RNA levels below 50 copies/ml without ART).

[g]N/A, not applicable.

[h]NA, not available.

Table S4. Integration site analysis from microcultures.

Pt ID	U3 loci[a]	U5 loci[b]	Gene[c]	Provirus type[d]	Copy number[e]	Expan-ded clone[f]	In trans-criptional unit[g]	CA association[h]
CP04	chr5-80876827	chr5+80876823	MSH3	3' del/h	32	+		+
CP04	chr5-80876827	chr5+80876823	MSH3	5' del	12	+		+
CP04	chr19+34451344	chr19-34451348	UBA2	5' del	28	+	+	
CP04	chr19+34451344	chr19-34451348	UBA2	3' del/h	32	+	+	
CP04	chr1-35495247	chr1+35495243	KIAA0319L	3' del/h	64		+	
CP04	chr7-106102652	chr7+106102648	SYPL1	5' del	24		+	
CP04	chr19+35190297	chr19-35190301	FXYD5	3' del/h	20			+
CP07	chr17-32199245	chr17-32199249	RHOT1	5' del	16		+	
CP07	chr3-98519105	chr3-98519101	CLDND1	3' del/h	308		+	
CP07	chr1+113782457	chr1-113782461	RSBN1	3' del/h	292		+	
CP07	chr17-44145548	chr17-44145552	C17orf53	3' del/h	96		+	
CP07	chr6+136733815	chr6+136733811	MAP3K5	3' del/h	8		+	
CP07	chr16+27412884	chr16-27412888	LOC105369891	3' del/h	8			
CP08	chr16+67046878	chr16-67046882	CBFB	3' del/h	16		+	+
CP08	chr17+42262488	chr17-42262492	STAT5B	3' del/h	20		+	+
CP08	chr17+42240249	chr17-42240253	STAT5B	3' del/h	24		+	+
CP08	chr2+9594060	chr2-9594064	YWHAQ	3' del/h	56		+	
CP08	chr17+60468915	chr17-60468919	APPBP2	5' del	12		+	
CP08	chr19-47092528	chr19+47092524	ZC3H4	3' del/h	4		+	
CP08	chr8+60557609	chr8-60557613	RAB2A	3' del/h	4		+	
CP08	chr2+61079679	chr2-61079683	KIAA1841	3' del/h	8		+	
CP08	chr7-75478021	chr7+75478017	POM121C	3' del/h	4		+	
CP08	chr17+47475167	chr17-47475171	MRPL45P2	3' del/h	4		+	
CP08	chr15-78491538	chr15+78491534	IREB2	3' del/h	4		+	
CP08	chrX+79360929	chrX-79360933	ITM2A	3' del/h	164		+	
CP08	chr1-160517973	chr1+160517969	SLAMF6	3' del/h	8		+	+
CP08	chr6-13719937	chr6+13719933	RANBP9	3' del/h	28			
CP08	chr17+42263312	chr17-42263316	STAT5B	3' del/h	4		+	+
CP08	chr6+90030225	chr6-90030229	BACH2	3' del/h	8		+	+
CP08	chr16-24750574	chr16+24750570	TNRC6A	3' del/h	8		+	
CP08	chr7+128564161	chr7-128564165	METTL2B	5' del	12			
CP08	chr2+190667403	chr2-190667407	NAB1	3' del/h	28		+	
CP08	chr7+72887790	chr7-72887794	POM121	3' del/h	4		+	
CP08	chr19+57845452	chr19-57845456	ZNF587B	3' del/h	68		+	
CP08	chr17+55289949	chr17-55289953	HLF	5' del	4		+	+
CP08	chr11+123370535	chr11-123370539	MIR4493	3' del/h	4			

Pt ID	U3 loci[a]	U5 loci[b]	Gene[c]	Provirus type[d]	Copy number[e]	Expan-ded clone[f]	In trans-criptional unit[g]	CA association[h]
CP08	chr7-5692004	chr7+5692000	RNF216	3' del/h	8		+	+
CP08	chr16+30296748	chr16-30296752	SMG1P5	3' del/h	16		+	
CP08	chr11-75347724	chr11+75347720	ARRB1	3' del/h	28		+	
CP08	chr5+176497230	chr5-176497234	FAF2	3' del/h	160		+	
CP08	chr6-42778282	chr6+42778278	GLTSCR1L	5' del	12		+	
CP08	chr6-26915454	chr6+26915450	GUSBP2	3' del/h	4		+	
CP08	chr6-135036395	chr6+135036391	HBS1L	3' del/h	4		+	
CP08	chr12-65250700	chr12+65250696	LEMD3	3' del/h	8			
CP08	chr3-196305363	chr3+196305359	TCTEX1D2	3' del/h	4		+	
CP08	chr19+12328836	chr19-12328840	ZNF563	5' del	4		+	
CP08	chr7-110998623	chr7+110998619	IMMP2L	3' del/h	4		+	
CP08	chr13+74007800	chr13-74007804	KLF12	3' del/h	4		+	
CP08	chr11-73752097	chr11+73752093	RAB6A	3' del/h	8		+	
CP08	chr9+3295869	chr9-3295873	RFX3	3' del/h	24		+	
CP08	chr9+132308637	chr9-132308641	SETX	3' del/h	16		+	
CP08	chr11-121540206	chr11+121540202	SORL1	3' del/h	4		+	+
CP08	chr17-61074679	chr17+61074675	BCAS3	3' del/h	1280		+	+
CP08	chr19-15290881	chr19+15290877	BRD4	5' del	24		+	+
CP08	chr11-108718192	chr11+108718188	DDX10	3' del/h	4		+	+
CP08	chr9-3522329	chr9+3522325	RFX3	5' del	16		+	
CP08	chr12-56044087	chr12+56044083	RPS26	5' del	8		+	
CP08	chr16+29557050	chr16-29557054	SMG1P2	3' del/h	16		+	
CP08	chr6+149336745	chr6-149336749	TAB2	3' del/h	4		+	
CP08	chr16-2493872	chr16+2493868	TBC1D24	3' del/h	4		+	
CP08	chr17+62510253	chr17-62510257	TLK2	5' del	12		+	
CP08	chr17-77986110	chr17+77986106	TNRC6C	3' del/h	8			

[a]Integration sites were determined using a previously described[8-10] linker ligation method with HIV-1 primers in the U3 region of the 5' LTR and the U5 region of the 3' LTR. Only sites detected with both U3 and U5 primers are reported. This column gives the U3 site.

[b]Location of integration site detected with U5 primer.

[c]Name of gene containing the integration site. In cases where the integration site is intergenic, the name of the nearest transcriptional unit is given.

[d]Nature of provirus detected by IPDA in cells from individual microculture wells. 3' del/h indicated provirus with a 3' deletion and/or hypermutation. 5' del indicated provirus with a 5' deletion. Cells with intact proviruses did not expand sufficiently for integration site analysis.

[e]Number of proviruses present in microculture as determined by IPDA droplet count adjusted for fraction of DNA used in IPDA.

[f]Identical integration site detected in >1 independent microculture from a single donor.

[g]Integration site located within a transcriptional unit.

[h]Integration site within or near a gene which has been associated with some form of cancer in previous studies as defined in reference 11.

Table S5. Primers and probes for IPDA.

Primer Name	HXB2 coordinates	Fluorophore, quencher	Sequence (5→3')
Ψ F	692-711		CAGGACTCGGCTTGCTGAAG
Ψ R	797-775[a]		GCACCCATCTCTCTCCTTCTAGC
Ψ Probe	758-740[a]	FAM, MGB[b]	TTTTGGCGTACTCACCAGT
Env F	7736-7759		AGTGGTGCAGAGAGAAAAAAGAGC
Env R	7851-7832[a]		GTCTGGCCTGTACCGTCAGC
Env intact probe	7781-7796	VIC, MGB	CCTTGGGTTCTTGGGA
Env hypermut[c]	7781-7798	Unlabeled, MGB	CCTTAGGTTCTTAGGAGC

[a]Reverse compliment.

[b]MGB, minor groove binder.

[c]Unlabeled competitor probe.

Table S6. PCR conditions for IPDA.

Cycling Step	Temperature	Time	Number of Cycles
Enzyme Activation	95°C	10 mins	1
Denaturation	94°C	30 secs	45
Annealing / Extension	59°C	1 min	
Enzyme Deactivation	98°C	10 mins	1
Hold	4-12°C	Infinite	1

References

1. Ocwieja, K. E. *et al*. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* **40**, 10345-10355 (2012).

2. Vandergeeten, C. *et al*. Cross-clade ultrasensitive PCR-based assays to measure HIV persistence in large-cohort studies. *J. Virol.* **88**, 12385-12396 (2014).

3. Avettand-Fenoel, V. *et al*. LTR real-time PCR for HIV-1 DNA quantitation in blood cells for early diagnosis in infants born to seropositive mothers treated in HAART area (ANRS CO 01). *J. Med. Virol.* **81**, 217-223 (2009).

4. Yu, J. J. *et al*. A more precise HIV integration assay designed to detect small differences finds lower levels of integrated DNA in HAART treated patients. *Virology* **379**, 78-86 (2008).

5. Massanella, M., Gianella, S., Lada, S. M., Richman, D. D. & Strain, M. C. Quantification of Total and 2-LTR (Long terminal repeat) HIV DNA, HIV RNA and Herpesvirus DNA in PBMCs. *Bio Protoc.* **5** (2015).

6. Schmid, A. *et al*. Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection. *PLoS One* **5**, e13310 (2010).

7. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration in vivo. *Nat. Med.* **7**, 631-634 (2001).

8. Lewinski, M. K. *et al*. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**, 6610-6619 (2005).

9. Berry, C. C. *et al*. Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* **28**, 755-762 (2012).

10. Sherman, E. *et al*. INSPIIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes. *Mol. Ther. Methods Clin. Dev.* **4**, 39-49 (2016).

11. Sherman, E. *et al*. INSPIIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes. *Mol. Ther. Methods Clin. Dev.* **4**, 39-49 (2016).

12. Detels, R. *et al*. The multicenter AIDS Cohort Study, 1983 to ... *Public Health* **126**, 196-198 (2012).

13. Bruner, K. M. *et al*. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat. Med.* **22**, 1043-1049 (2016).

14. Imamichi, H. *et al*. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* (2016).

15. Rose, P. P. & Korber, B. T. Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics* **16**, 400-401 (2000).

16. Finzi, D. *et al*. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295-1300 (1997).

17. Laird, G. M., Rosenbloom, D. I., Lai, J., Siliciano, R. F. & Siliciano, J. D. Measuring the Frequency of Latent HIV-1 in Resting CD4(+) T Cells Using a Limiting Dilution Coculture Assay. *Methods Mol. Biol.* **1354**, 239-253 (2016).

18. Laird, G. M. *et al*. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398 (2013).

19. Rosenbloom, D. I. *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).

20. Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708-712 (1999).

21. Chomont, N. *et al*. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893-900 (2009).

22. Durand, C. M. *et al*. 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 (2012).