Title

Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy

Journal

Journal of Clinical Investigation, 130(7)

ISSN

0021-9738

Authors

Antar, Annukka AR
Jenike, Katharine M
Jang, Sunyoung
et al.

Publication Date

2020-07-01

DOI

10.1172/jci135953

Peer reviewed

Longitudinal study reveals HIV-1–infected CD4+ T cell dynamics during long-term antiretroviral therapy

Annukka A.R. Antar,[1] Katharine M. Jenike,[1] Sunyoung Jang,[1] Danielle N. Rigau,[1] Daniel B. Reeves,[2] Rebecca Hoh,[3] Melissa R. Krone,[4] Jeanne C. Keruly,[1] Richard D. Moore,[1] Joshua T. Schiffer,[2,5,6] Bareng A.S. Nonyane,[7] Frederick M. Hecht,[3] Steven G. Deeks,[3] Janet D. Siliciano,[1] Ya-Chi Ho,[1] and Robert F. Siliciano[1,8]

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [2]Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [3]Department of Medicine and [4]Department of Epidemiology and Biostatistics, UCSF, San Francisco, California, USA. [5]Department of Medicine, University of Washington, Seattle, Washington, USA. [6]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [7]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [8]Howard Hughes Medical Institute, Baltimore, Maryland, USA.

Proliferation of CD4+ T cells harboring HIV-1 proviruses is a major contributor to viral persistence in people on antiretroviral therapy (ART). To determine whether differential rates of clonal proliferation or HIV-1–specific cytotoxic T lymphocyte (CTL) pressure shape the provirus landscape, we performed an intact proviral DNA assay (IPDA) and obtained 661 near–full-length provirus sequences from 8 individuals with suppressed viral loads on ART at time points 7 years apart. We observed slow decay of intact proviruses but no changes in the proportions of various types of defective proviruses. The proportion of intact proviruses in expanded clones was similar to that of defective proviruses in clones. Intact proviruses observed in clones did not have more escaped CTL epitopes than intact proviruses observed as singlets. Concordantly, total proviruses at later time points or observed in clones were not enriched in escaped or unrecognized epitopes. Three individuals with natural control of HIV-1 infection (controllers) on ART, included because controllers have strong HIV-1–specific CTL responses, had a smaller proportion of intact proviruses but a distribution of defective provirus types and escaped or unrecognized epitopes similar to that of the other individuals. This work suggests that CTL selection does not significantly check clonal proliferation of infected cells or greatly alter the provirus landscape in people on ART.

Introduction

People living with HIV-1 (PLWH) harbor a latent reservoir from which the virus typically rebounds after the cessation of antiretroviral therapy (ART), even after years of successful ART-mediated suppression of viral replication (1–6). The latent reservoir is the major barrier to cure and is primarily composed of resting CD4+ T cells whose genomes contain integrated HIV-1 proviruses that are intact and inducible (7–9). When a quantitative viral outgrowth assay is used to measure the reservoir, estimates of the frequency of latently infected cells in people on ART range between 0.1 and 10 cells per million resting CD4+ T cells (1, 10–12).

Although the latent reservoir is composed of cells harboring intact proviruses, most HIV-1–infected cells persisting in PLWH on suppressive ART harbor defective proviruses (13–18). Proviruses with overtly fatal defects outnumber intact proviruses 10–50 to 1. One major type of defect, termed hypermutation, involves numerous G→A mutations introduced by the human APOBEC3 family of cytidine deaminases. These mutations alter start codons and introduce multiple internal stop codons in open reading frames (ORFs) for viral genes (13, 14, 19–21). Another common type of defect involves deletions in the HIV-1 provirus. These deletions vary widely in length but typically encompass multiple viral genes (13–16, 22). Deletions likely arise during reverse transcription when the polymerase jumps to a homologous region on the same template during negative-strand synthesis, but other mechanisms may contribute (16, 22–24). Other less common defects include frameshift mutations and inverted segments of the HIV-1 genome inserted within a large multigenic deletion (14).

One approach to HIV-1 cure involves the shock and kill strategy (25), which depends on pharmacologic strategies to reverse latency (26) so that infected cells can be eliminated. The efficacy of such strategies depends on elimination of cells with intact proviruses, but shock strategies may induce defective proviruses, many of which have functional promoters (16, 27, 28). In fact, HIV-1 RNA is expressed from some defective HIV-1 proviruses in people on suppressive ART even in the absence of a shock intervention, perhaps due to activation of the relevant host CD4+ T cells after antigen recognition in vivo (16, 29–32). Some defective HIV-1 RNAs can be

Conflict of interest: The intact proviral DNA assay is the subject of patent application 16/078760. RFS is a consultant on cure-related HIV research for Merck and AbbVie. This research was funded in part from an unrestricted research grant to YCH and RFS from Gilead. JDS is a consultant on cure-related HIV research for Southern Research and Gilead. The following patent disclosures are unrelated to the present study: RFS is an inventor on patent 7,468,274 and also on patent applications 15/528230, 15/552452, and 15/568893. SGD is an inventor on patent application 15/867829. FMH is an inventor on patent applications 13/891970, 12/677278, and 11/880126.

Submitted: December 23, 2019; **Accepted:** March 17, 2020; **Published:** June 2, 2020.
Reference information: *J Clin Invest*. 2020;130(7):3543–3559.
https://doi.org/10.1172/JCI135953.

Table 1. Study participant and sample characteristics

Participant ID	CP or C[A]	Age at T2[B]	ART at T1[B]	ART at T2[B]	CD4 at T1[C]	CD4 at T2[C]	Months on ART before T1	Months (years) between T1 and T2	Months (years) between T1 and T3	Num. Proviruses T1	Num. Proviruses T2	Num. Proviruses T3
22	CP	44	ABC/3TC, EFV	ABC/3TC, DTG	455	687	24.0	93.6 (7.8)	N/A	48	41	N/A
1211	CP	49	ABC/3TC, ATV/r	ABC/3TC, ATV/r	598	517	24.0[D]	39.6 (3.3)	N/A	50	61	N/A
548	CP	51	ABC, AZT, TDF, 3TC, LPV/r	TDF/FTC, EFV	594	727	16.6	123.6 (10.3)	N/A	10	93	N/A
583	CP	48	TDF/3TC, EFV	TDF/FTC, EFV	673	894	22.4	110.4 (9.2)	N/A	54	58	N/A
746	CP	48	AZT/3TC, EFV	TDF/FTC, EGV/c	522	667	22.2	100.8 (8.4)	N/A	86	68	N/A
1532	C	54	TDF/FTC, RAL	TDF/FTC, RAL	309	369	7.3	38.3 (3.2)	80.5 (6.7)	3	7	13
1124	C	62	TDF/FTC, ATV/r	TAF/FTC, DTG	657	353	7.6	77.5 (6.5)	N/A	12	25	N/A
1194	C	63	TDF/FTC, RAL	ABC/3TC, DTG	283	329	6.0	82.8 (6.9)	N/A	7	25	N/A

[A]CP, chronic progressor; C, controller. [B]T1, sample time point 1; T2, sample time point 2. [C]CD4 count is either from date of sample or closest in time clinically measured absolute CD4 count and is expressed as cells/μL. [D]The exact start date of ART for participant 1211 is unknown but based on plasma ART testing and documented viral loads is ±6 months from this estimate. ABC, abacavir; 3TC, lamivudine; TDF, tenofovir disoproxil; TAF, tenofovir alafenamide; FTC, emtricitabine; AZT, zidovudine; EFV, efavirenz; /r, boosted with ritonavir; ATV, atazanavir; LPV, lopinavir; EGV/c, elvitegravir boosted with cobicistat; DTG, dolutegravir; RAL, raltegravir.

translated to yield viral proteins recognized by cytotoxic T lymphocytes (CTLs) (17, 27, 28, 33–35). If this is the case, then depletion of intact and certain defective HIV-1 proviruses should occur over long periods of time in PLWH on suppressive ART, and a cure may eventually be achieved with strategies that accelerate such depletion.

The remarkable persistence of HIV-1 proviruses in people on suppressive ART likely reflects clonal proliferation of infected cells and memory cell longevity (18, 36–47). Cells harboring intact proviruses and cells harboring defective proviruses undergo clonal proliferation in people on suppressive ART, but it is unknown whether they proliferate to the same degree (14, 15, 18, 40, 43–46, 48). If drivers of clonal proliferation such as antigen recognition by the host cell also induce HIV-1 gene expression, then cells harboring intact proviruses may be less likely to undergo extensive clonal proliferation than uninfected cells due to cytopathic effects of HIV-1 gene expression or CTL pressure. For defective proviruses, the extent of expansion may depend on the nature of the defect and which viral genes can be expressed (27, 28, 34). For some defective proviruses, particularly those with very large deletions, induction of viral gene expression is unlikely to lead to viral cytopathic effects or CTL recognition.

In light of these issues, we conducted a longitudinal study to determine whether CTL pressure or differential rates of clonal proliferation shape the proviral landscape in PLWH on suppressive ART. To magnify small effects, we obtained samples from widely separated time points (average 7 years) in people who had consistently undetectable viral loads on ART. Given that controllers have particularly strong HIV-specific CTL responses, we included controllers on ART in our study (49).

Results

To better understand HIV-1 persistence in people on suppressive ART, we obtained 661 near–full-length sequences from 8 people with HIV-1 subtype B (Table 1, Figure 1, and Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/JCI135953DS1). Two samples — in one case 3 — were obtained from each person, with an average of 7 years between samples. None of the individuals started ART during early infection, and all maintained viral suppression on ART before and in between samples, with minor exceptions as noted in the Methods. Three individuals maintained viral loads of less than 400 copies RNA/mL for more than 3 years before ART (controllers). Controllers were started on ART for reasons other than loss of virus control and hence were assumed to have had effective HIV-1–specific responses. The other 5 individuals did not control viremia in the absence of ART (chronic progressors). They had been on ART for an average of 22 months before the first sample, compared with 7 months for the controllers. See Methods and Table 1 for a detailed characterization of study participants.

The 661 near–full-length proviral sequences were obtained from purified resting $CD4^+$ T cells using a limiting-dilution single-genome analysis technique covering all ORFs and part of the LTR (13, 14). We added a seventh inner PCR with primers located just inside the outer PCR primers. This allowed us to obtain more sequences per participant and obtain data from regions that would not have been sequenced using previous techniques. An average of 57 sequences per time point (range 10–93) were obtained for the chronic progressors. Fewer sequences were obtained per time point from controllers (average 13, range 3–25), consistent with the lower frequency of latently infected cells in these individuals (50–52).

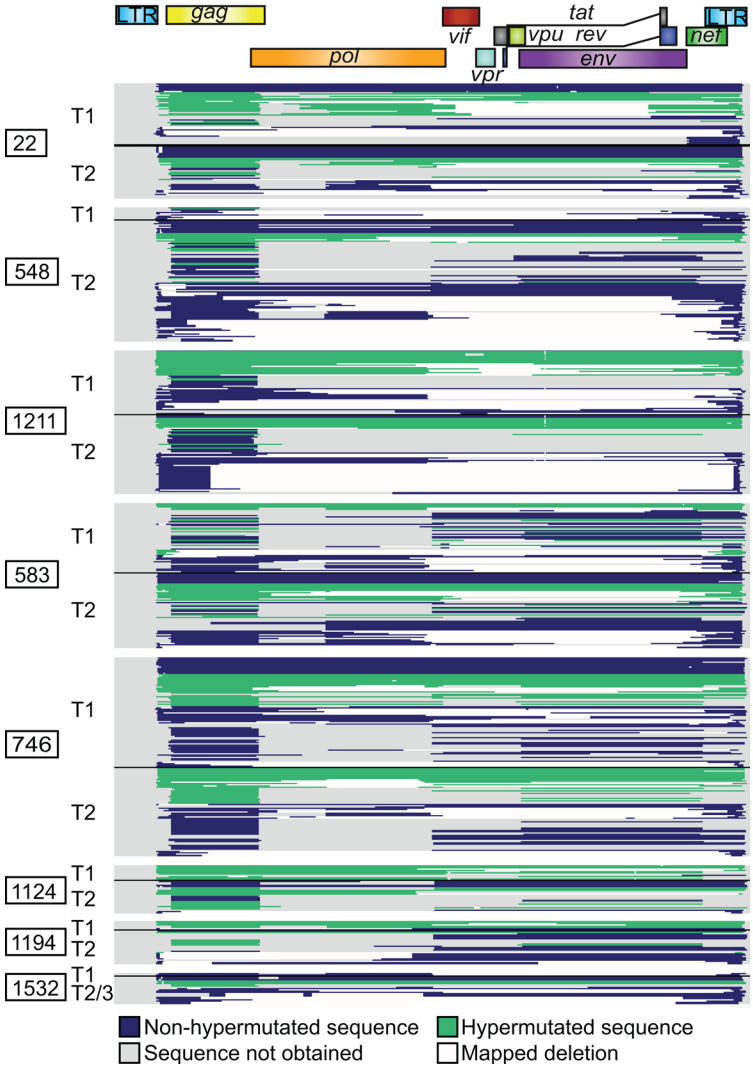

Figure 1. Landscape of HIV-1 provirus genomes at 2 widely spaced time points. Six hundred sixty-one individual provirus genomes were obtained by near-full-length sequencing from peripheral resting CD4⁺ T cells from 8 participants who maintained suppressed viral loads on ART for years. Each horizontal bar represents a single provirus genome. The inner PCR primers corresponding to Gag and Env were the most likely to amplify, being the shortest. Therefore, nucleotides in the Gag and Env regions are more likely to be sequenced using this method than nucleotides in other regions. Areas that are gray correspond to regions in the provirus for which sequence was not obtained and may or may not contain a deletion. Areas that are white correspond to mapped deletions. T1, T2, and T3: time points 1, 2, and 3.

Each proviral sequence was analyzed for defects as follows. Sequences were analyzed for APOBEC-induced hypermutation using an algorithm that compares the frequencies of G→A and A→G mutations (53). Each sequence was examined for deletions unlikely to be common sequence length polymorphisms. Deletions were categorized as lying entirely within the 5′ or 3′ half of the genome (5D, 3D), within the packaging signal or major splice donor site region (PD), spanning the central region of the genome (central deletion, CD), encompassing more than 75% of the genome (large deletion, LD), or having 2 or more deletions (2D). A fraction of the sequences were presumed to have an unmapped deletion (D) based on amplification from only a subset of inner PCRs that did not span the entire genome. Full-length, nonhypermutated sequences with a 1- to 2-nucleotide deletion resulting in a frameshift in a protein-coding HIV-1 gene were denoted FS. A few sequences had a large deletion as well as insertion of an inverted sequence (inversion sequences, INV).

Proviral sequences that did not have deletions or hypermutation were termed intact only if analysis of the amino acid sequence of the 9 major HIV-1 gene products demonstrated no internal stop codons or frameshifts. Minor exceptions are detailed in the Methods. Intact HIV-1 proviruses identified in this way show normal in vitro replication but inapparent defects affecting viral fitness cannot be ruled out (13).

The HIV-1 proviral landscape does not change dramatically over time on ART. Activated CD4⁺ T cells harboring intact proviruses and certain types of defective proviruses may be subject to negative selection by HIV-1–specific CTLs in people on suppressive ART (17, 27, 28, 34, 54, 55). If selection operates on cells comprising the latent reservoir, perhaps during transient periods of cellular activation, a decreased proportion of intact proviruses might be observed over time, while proviruses with very large deletions encompassing many ORFs might increase in proportion. Full-length hypermutated proviruses, full-length proviruses with a single frameshift mutation, and proviruses with small deletions in the packaging signal region are potentially capable of generating T cell epitopes and might decrease in proportion as well (27). Therefore, we carried out a detailed analysis of proviral sequences from individuals on ART at widely spaced time points to detect changes in the proviral landscape.

We found that the distribution of provirus types present 1 to 2 years after the start of ART was generally similar to that present after approximately 7 more years on suppressive ART (Figure 2, A and B). All major types of proviruses seen at the first time point

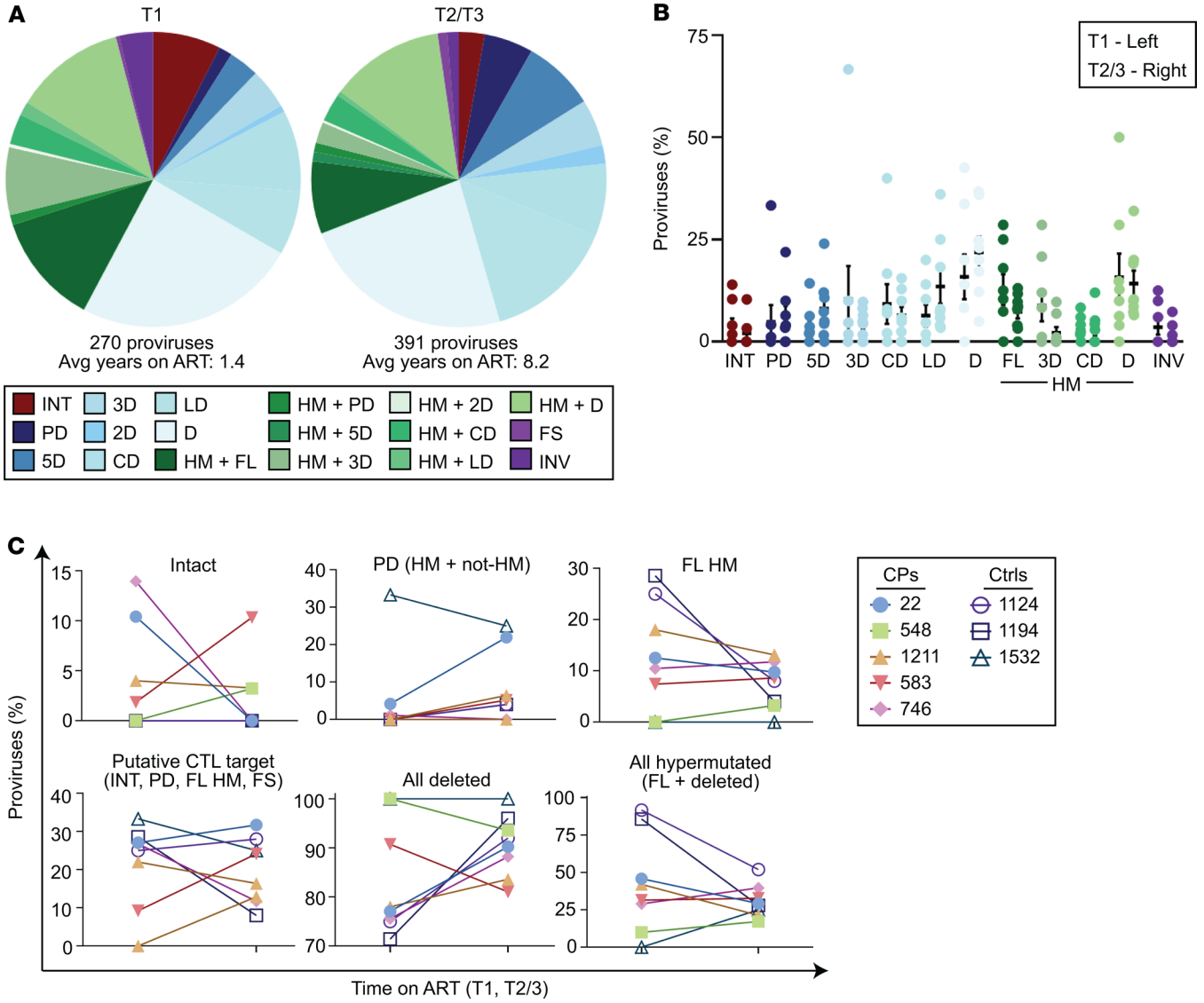

Figure 2. The distribution of provirus types appears similar over long periods of time on suppressive ART by near-full-length sequencing. (**A**) Summary of 270 proviral sequences (time point 1, T1) and 391 proviral sequences (T2/3) from 8 individuals maintained on long-term suppressive ART. Defective proviruses are categorized by type of defect. (**B**) Percentage of intact and various types of defective proviruses relative to total proviruses from each individual at the indicated time point, with mean and SEM shown in black. Left, T1; Right, T2/3. **A** and **B** share a color-coded figure legend. (**C**) Percentage of each indicated provirus type or grouping among total proviruses at the indicated time point for each individual. No intact proviruses were observed in the controllers on ART. (**B** and **C**) Paired-sample *t* tests with Bonferroni's correction were applied. No provirus type or grouping change over time was statistically significant. CPs, chronic progressors; Ctrls, controllers; INT, intact; PD, deletion within packaging signal region; 5D, deletion within 5′ half of HIV-1 genome; 3D, deletion within 3′ half of HIV-1 genome; 2D, two or more deletions; CD, deletion spans center of genome and less than 75% of genome length; LD, deletion of more than 75% of HIV-1 genome length; D, unmapped deletion; HM, hypermutated; FL, full-length; FS, intact save for 1 or more frameshifts affecting genes required for replication; INV, deleted with inverted sequence.

were still present 7 years later. Aggregate data from all participants showed a decrease in the proportions of intact proviruses (20 of 270 at time point 1 [T1], 11 of 391 at T2) and hypermutated proviruses and an increase in the proportion of proviruses with large deletions (Figure 2A). We applied a paired-samples *t* test for the hypothesis that the proportion of each provirus type did not change over time on ART. No significant longitudinal change in the proportions of any particular provirus type was observed. Similarly, we saw no significant change in the proportions of all deleted proviruses or all hypermutated proviruses (Figure 2C).

We next tested for negative selection against cells containing types of proviruses hypothesized to be capable of generating CTL epitopes following $CD4^+$ T cell activation (27). These include intact proviruses, full-length hypermutated proviruses, proviruses that are full length except for a small deletion in the packaging signal region, and full-length proviruses with a single frameshift mutation in an essential HIV gene. We found no evidence that the proportions of these proviruses changed over long periods of time on ART, either when considered separately or grouped together (Figure 2C and Supplemental Figure 2).

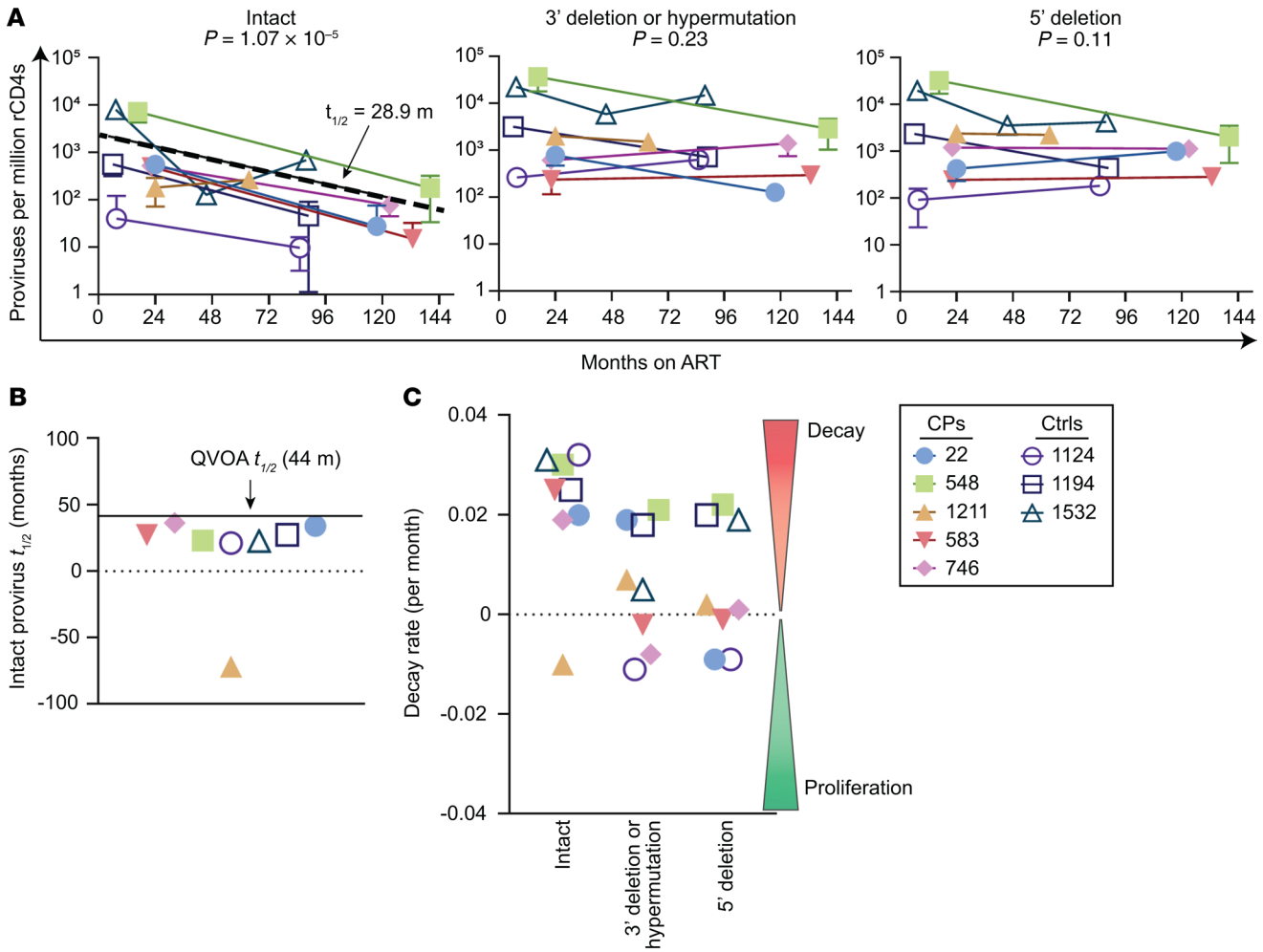

Figure 3. Longitudinal analysis of provirus frequency by IPDA reveals slow decay of intact proviruses over time on ART. IPDA analysis was performed on resting CD4⁺ T cells (rCD4s) from the same samples as those analyzed by near-full-length sequencing. (**A**) Left: Intact proviruses per million resting CD4⁺ T cells plotted as a function of months on ART. Log-linear mixed-effects modeling was used to determine the population level half-life for intact DNA (28.9 months) – plotted here as a dashed line – and generate *P* values for all types of proviruses. Proviruses with 3′ deletions and/or hypermutation (center) and proviruses with 5′ deletions (right) per million resting CD4⁺ T cells plotted as a function of months on ART. (**B**) Half-lives ($t_{1/2}$) of cells carrying intact proviruses assuming exponential decay, excluding any IPDA measurement of zero. The half-life of replication-competent HIV-1 proviruses in people on ART as measured by quantitative viral outgrowth assays (QVOA) is known to be approximately 44 months and is shown in black. (**C**) IPDA data are plotted in terms of decay rate assuming exponential decay, excluding IPDA measurements of zero. CP, chronic progressors; Ctrls, controllers.

Intact proviruses represented 5.4% of 569 total proviruses obtained from chronic progressors on ART. Notably, no intact proviruses were found among the 92 proviruses from controllers on ART. Among the chronic progressors, there was no consistent pattern of change in proportion of intact proviruses over the time interval studied (Figure 2C). However, intact proviruses are much less frequent than defective proviruses, and only 31 intact proviruses were recovered from the chronic progressors. The small number of intact proviruses recovered may have precluded detection of a subtle change in proportion.

Slow decay of intact proviruses detected by IPDA. To detect small changes in the proportion of intact proviruses over time on ART, we measured their frequencies in the same samples using the recently described intact proviral DNA assay (IPDA). The IPDA is a digital droplet PCR assay that estimates the number of intact proviruses by measuring the frequency of proviruses containing the packaging signal and a nonhypermutated conserved region of the Rev-response element (RRE) (55). It detects proviruses lacking the 2 major classes of fatal defects, large deletions and APOBEC3G-mediated hypermutation. A fraction of the proviruses detected harbor minor defects that may affect viral fitness. By near-full-length sequencing, we estimate that this fraction is approximately 30% (55). Seven of 8 participants demonstrated decay of intact proviruses (Figure 3A). The eighth participant demonstrated an increase that could reflect clonal expansion (18, 37, 42, 43, 45, 46). Using log-linear mixed-effects modeling to analyze population-level half-lives, we found that intact proviruses decayed with a half-life of 28.9 months ($P = 1 \times 10^{-5}$, 95% CI 20.0–52.1), similar to that previously reported for replication-competent proviruses as measured by quantitative viral outgrowth assay (QVOA; 30 months in people without blips, 43–44 months in people on ART generally) (Supplemental Table 1 and refs. 3, 10, 11).

Intact provirus decay by IPDA in some but not all participants on ART was previously reported (55, 56). The IPDA also measures 2 types of defective proviruses: those that contain the RRE but not the packaging signal and those that contain the packaging signal but are either missing the RRE or are hypermutated in the RRE. There was no clear longitudinal trend for these 2 types of defective proviruses, with decay in some participants and increases in others (Figure 3). However, total proviruses decayed with a population half-life of 78.2 months ($P = 0.03$, 95% CI 41.0–835.8).

Together, these results indicate that intact proviruses decay slowly over time, consistent with previous measurements by QVOA, whereas no dramatic changes are seen in the proportion of various types of defective proviruses.

No evidence for selection by CTLs over time on ART by analysis of HIV-1 epitopes in Gag, Pol, and Nef. A large proportion of immunodominant CTL epitopes in HIV-1 proviruses in people on ART show escape mutations (57). We reasoned that if HIV-1–specific CTLs can clear cells carrying proviruses that are expressed in the context of cellular activation, then the proportions of proviruses that cannot generate recognizable CTL epitopes might increase over very long periods of time on ART. These would include proviruses with escape mutations or deletions in CTL epitopes or preceding frameshifts or stop codons affecting epitope expression. To test this hypothesis, we identified the dominant CTL epitopes in HIV-1 Gag, Pol, and Nef for each participant's HLA type using published data and then examined the proportions of wild-type versus escape or unrecognized epitope sequences using our longitudinal proviral sequence database (see Methods). Gag, Pol, and Nef were chosen because of their immunogenicity and the depth of the published HLA-mapped epitope polymorphism data. We classified each epitope in each provirus as wild-type/recognized, escaped, or lying within a deleted region of the provirus. Within each provirus, any frameshift or stop codon between the start codon of the relevant gene and the epitope was noted and the epitope was categorized as FS or SC. If the amino acid sequence of a provirus at a defined epitope of interest had not been previously characterized with respect to CTL recognition, we labeled it an uncharacterized mutation and then used a previously described algorithm (NetMHC 4.0, http://www.cbs.dtu.dk/services/NetMHC) to predict whether the amino acid sequence of the epitope was a strong (SB), weak (WB), or nonbinder (NB) to the relevant MHC class I molecule. Epitopes that were in regions with no sequence information were removed from these analyses.

Using these categories, we first tested whether the proportion of escaped epitopes, deleted epitopes, and epitopes with upstream frameshifts or stop codons increased over long time intervals in people on suppressive ART, as would be expected from CTL pressure (Figure 4 and Figure 5, A and B). We found no increase. In fact, this set of nonrecognized epitope sequences had decreased by 16% by the later time point (OR 0.84, 95% CI 0.75–0.93). Concordantly, the proportion of wild-type epitope sequences actually increased over time (Figure 5, A and B). When each of the nonrecognized epitope types was analyzed individually, no specific type of epitope solely contributed to the decline in the proportion of nonrecognized epitopes in a statistically significant manner (Figure 5C). To determine if a trend toward decreasing proportions of wild-type epitopes could be detected by considering only the grouping of proviruses thought to render their host cells putative CTL targets — intact, FL hypermutated (HM), FS, and PD — we repeated the analysis with just this subset of proviruses and again observed that the proportions of wild-type epitopes increased and that of unrecognized epitopes decreased (Supplemental Figure 3). In summary, in this analysis of dominant CTL epitopes, we did not observe increasing proportions of unrecognized epitopes in HIV-1 proviruses in people on long-term ART, suggesting that HIV-1–specific CTLs do not shape the provirus landscape to a great degree in people on suppressive ART.

The proportion of intact proviruses observed in clones is no less than the proportion of defective proviruses observed in clones. It has been unclear whether cells harboring intact proviruses proliferate in people on suppressive ART to the same extent as cells harboring defective proviruses. Stimuli driving clonal expansion such as antigen-driven proliferation and homeostatic mechanisms driven by IL-7 and IL-15 may also reverse latency and allow CTL-mediated or viral cytopathic effect–mediated clearance of infected cells. Our large data set of near–full-length HIV-1 provirus sequences allowed for examination of clonal proliferation according to provirus type. Clones were identified as proviruses with identical nucleotide sequences and, if deleted, with identical deletion junctions, as has been done previously (48, 58, 59). Although definitive assignment of clonality requires integration site analysis, full genome sequence identity provides very strong evidence for clonality (60). This is especially true for defective proviruses, given the underlying sequence diversity of HIV-1, the heterogeneity of defects, and the inability of these viruses to spread. Infected $CD4^+$ T cell clones can persist in people for well over a decade, and we identify clones as any 2 or more identical sequences from the same individual regardless of the time point (37). The clonality dynamics described here apply to large clones, as any clone observed in this study is likely part of a very large clone given our small sample sizes relative to the total body population of HIV-1–infected cells. We expect that many of the singletons we observed, especially those observed at late time points, are also members of a clonal population (38).

Compared with defective proviruses, intact proviruses were just as likely, if not more so, to be observed in clones (Figure 6, A and B). This was true whether clonality rates were calculated by collapsing each clone family into one unique sequence (Figure 6A) or considering the whole population of sequenced proviruses (Figure 6B). This was also true whether all 8 participants were considered as in Figure 6 or just the 5 chronic progressors (*t* test *P* values adjusted for clustering by participant were 0.88 and 0.24 respectively). No significant differences in clonality were found whether we compared full-length versus deleted proviruses, hypermutated versus nonhypermutated proviruses, or putative CTL targets (intact, FL HM, FS, and PD) versus the rest of the proviruses with mapped deletions (Figure 6C).

Proviruses observed in clones are not enriched in escaped or unrecognized epitopes. The above results suggest that the proliferation of infected cells in people on ART occurs in a manner that is not strongly influenced by whether the provirus is intact or has various types of defects. We next wanted to understand whether proliferation of infected cells is influenced by whether the proviruses carry epitopes recognized by HIV-1–specific CTLs. We tested whether intact proviruses observed in clones are enriched in escaped

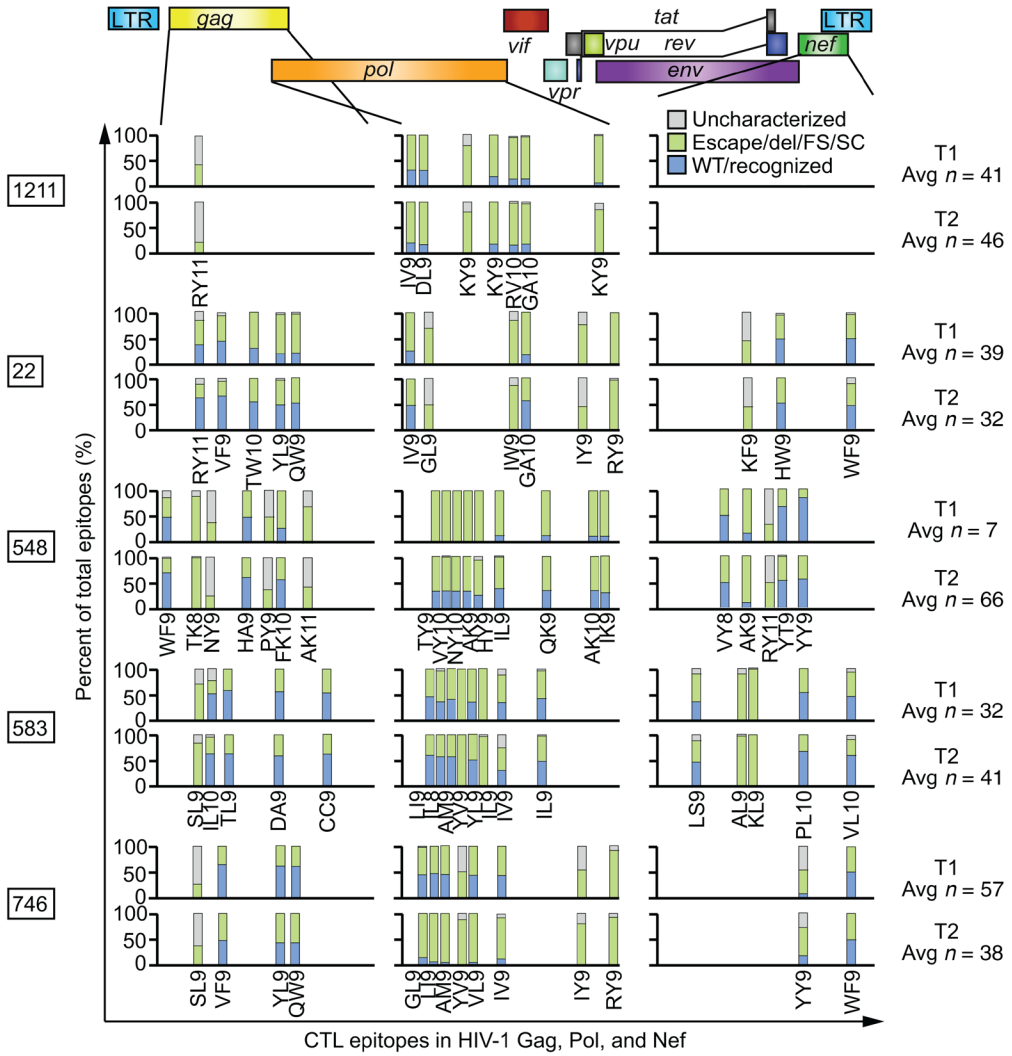

Figure 4. Percentages of unrecognized CTL epitopes in Gag, Pol, and Nef from HIV-1 proviruses do not change dramatically over long periods of time on suppressive ART. Percentages of wild-type, unrecognized (deleted, escaped, frameshifted, or preceding stop codon), and uncharacterized epitope sequences at each dominant epitope in 5 chronic progressors on suppressive ART at 2 widely spaced time points. Provirus epitope sequences that lay in an area with unavailable sequence information were discarded from the analysis. The average number of total available epitope sequences per participant time point is shown at right. Predicted dominant epitopes in Gag, Pol, and Nef given each participant's HLA type are shown and labeled. Epitope information can be found in Supplemental Table 2. Del, epitope lay within a mapped deletion; FS, a frameshift mutation precedes the epitope; SC, a stop codon precedes the epitope.

or unrecognized epitopes and found no significant enrichment (Figure 6D). Similarly, when we extended the analysis to include defective proviruses, we again found no significant enrichment of escaped or unrecognized epitopes in clones (Figure 6, E and F, and Supplemental Table 3). In fact, accounting for clustering by participant, unrecognized epitopes were 35% less likely to be found in clones (OR 0.65, 95% CI 0.55–0.77). These data suggest that clonal expansion of infected cells is not hampered by proviruses carrying epitopes that can be recognized by CTLs.

The observed richness of proviral species decreases with time on ART. We found that HIV-1 proviruses were increasingly enriched in large clones over time in both chronic progressors and controllers and that many clones detected at the first time point were present at the second time point, consistent with other studies (Figure 7, A–D, and refs. 17, 40, 61). However, it was unknown whether the sampling depth of prior studies was sufficient to detect a difference in clonality over time on ART with confidence. To do so, we adapted methods from the field of ecology developed to make observations about species diversity using small samples. For the present study, observed richness represents the number of unique provirus sequences from infected cells in a sample. Rarefaction is an ecological technique used to compare observed richness values given different sampling depths. We plotted rarefaction curves for the 5 chronic progressors — chosen because of the larger number of proviruses sequenced — in which random resampling of N (x axis) proviruses from the total number of proviruses observed per participant time point was performed and the average number of unique sequences Y was plotted (y axis), with 95% CIs shown. We found that for participants 1211 and 746, provirus sampling was deep enough to detect a statistically significant decrease in observed proviral richness over time on ART (Figure 8A). We also observed a decrease in proviral richness in participants 22 and 583. Undersampling at the first time point precluded rarefaction analysis for participant 548 (Supplemental Figure 4). The decrease in observed proviral richness normalized for sample size across all 8 participants was significant (Pearson P = 0.005, Figure 8C). No individual exhibited evidence of increased proviral richness over time.

A recent study used HIV-1 integration-site data sets and QVOA sequences to demonstrate that the clonality of HIV-1–infected cells approximates a power-law distribution in which the HIV-1 provirus landscape is composed of a small number of large clones and a large number of small clones (38). To determine how

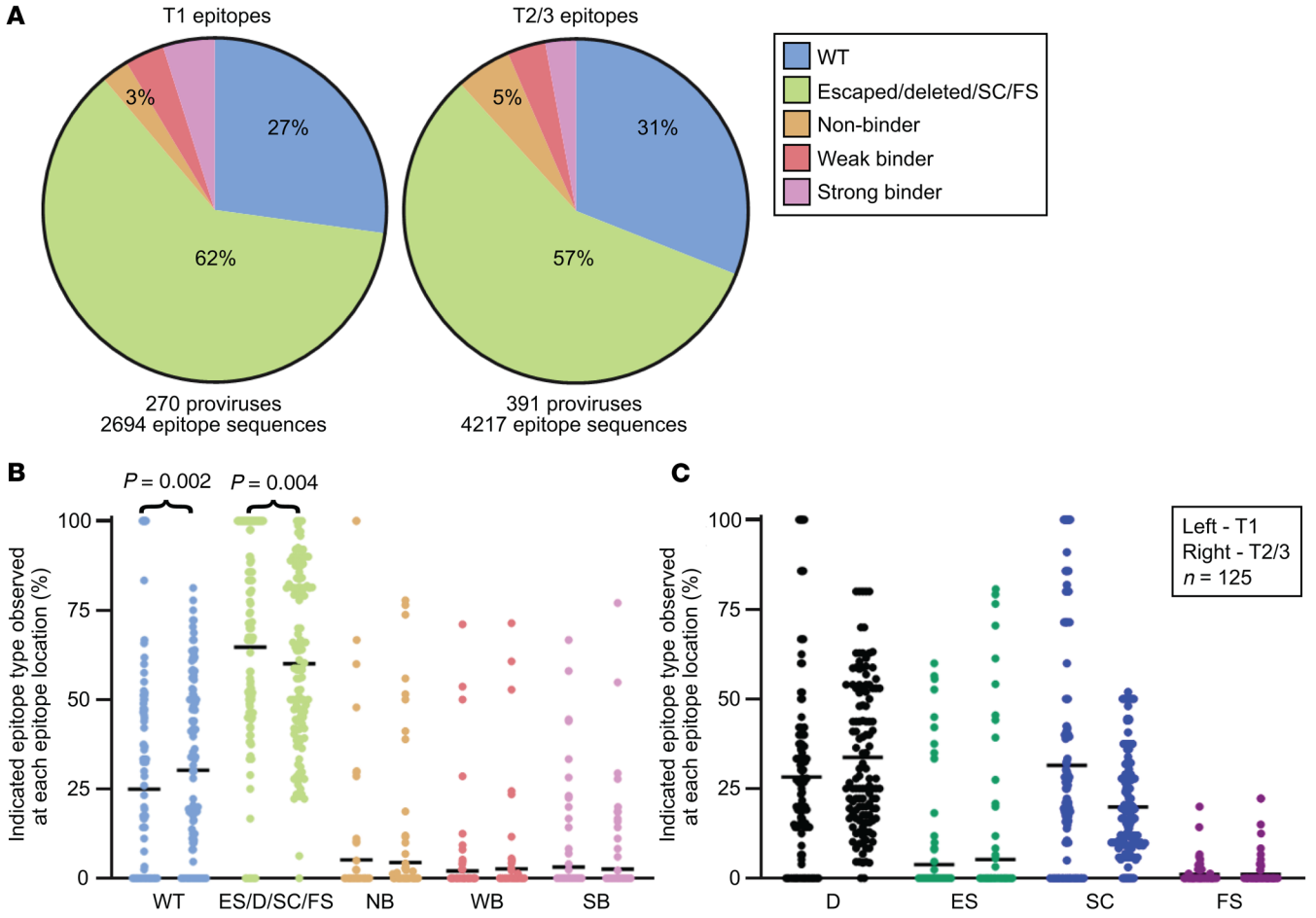

Figure 5. Percentages of unrecognized epitope sequences in Gag, Pol, and Nef from HIV-1 proviruses decrease over time in people on suppressive ART. (**A**) Summary pie charts of Gag, Pol, and Nef epitope sequences from 125 predicted dominant epitopes from all 8 participants from widely spaced time points after the initiation of suppressive ART. Unrecognized epitopes include escaped epitopes (ES), epitopes that lie within a deleted region of the provirus (D), and epitopes with a preceding frameshift (FS) mutation or stop codon (SC). Epitope sequences that were uncharacterized in the literature are categorized as nonbinder (NB), weak binder (WB), or strong binder (SB) based on predictive software (NetMHC 4.0) and relevant HLA allele. (**B** and **C**) For each of the 125 predicted dominant epitopes, the percentage of the indicated epitope type among the proviruses with epitope sequence available at that site is plotted as 1 dot, with the mean shown in black. In each pair, left denotes time point 1 (T1) and right, T2 and T3. *P* values were generated to test the hypothesis that the percentage of each type of epitope did not change over time on ART using the nonparametric Wilcoxon's signed-rank test. Significant results after application of Bonferonni's correction are shown. (**B**) Epitope sequences were divided into 5 categories shown here, with all unrecognized epitope sequences grouped together. (**C**) Epitope sequences were divided into 8 categories, with each type of unrecognized epitope counted individually. Shown here are the 4 unrecognized epitope types. WT, wild-type.

the provirus clonality distribution changes over time on ART, we used the observed rank abundance of clones (Figure 8B) at each participant time point to extrapolate the clonal abundance in the total-body infected resting $CD4^+$ T cell pool. Modeled reservoir extrapolations illustrate that the proportion of HIV-1 proviruses in the largest clones increases over time on ART (Figure 8D and Supplemental Figure 5). Together with the IPDA and richness observations above, this work suggests that over time on suppressive ART, the total frequency of HIV-1 proviruses decays extremely slowly while proviruses are increasingly concentrated among a relatively small number of predominant large clones.

The provirus landscape in controllers on ART. To the best of our knowledge, the dynamics of HIV-1 proviruses in controllers on ART have not been previously described. All major types of defective HIV-1 proviruses were found in controllers on ART (Figure 9A). We found no significant differences between controllers on ART and chronic progressors on ART in the proportions of proviruses with packaging signal deletions, very large deletions encompassing over 75% of the genome (LD), all deleted proviruses, or all hypermutated proviruses (Figure 9B). The provirus grouping hypothesized to render their host cells HIV-1–specific CTL targets upon expression (intact, FL HM, FS, and PD) was no less likely to be present in controllers on ART than chronic progressors (Supplemental Figure 6). The other types of proviruses were not present in great enough numbers in our study to analyze separately.

No intact proviruses were found among the 92 proviral sequences from controllers on ART (Figure 9, A and C). Therefore, we used a more sensitive measure of intact proviruses, the IPDA, to measure the frequency of intact proviruses at each time point (55). No difference between controllers and chronic progressors was seen (Figure 9D). However, at least 1 clone of proviruses in participant 1124 that was nonhypermutated and full-length save

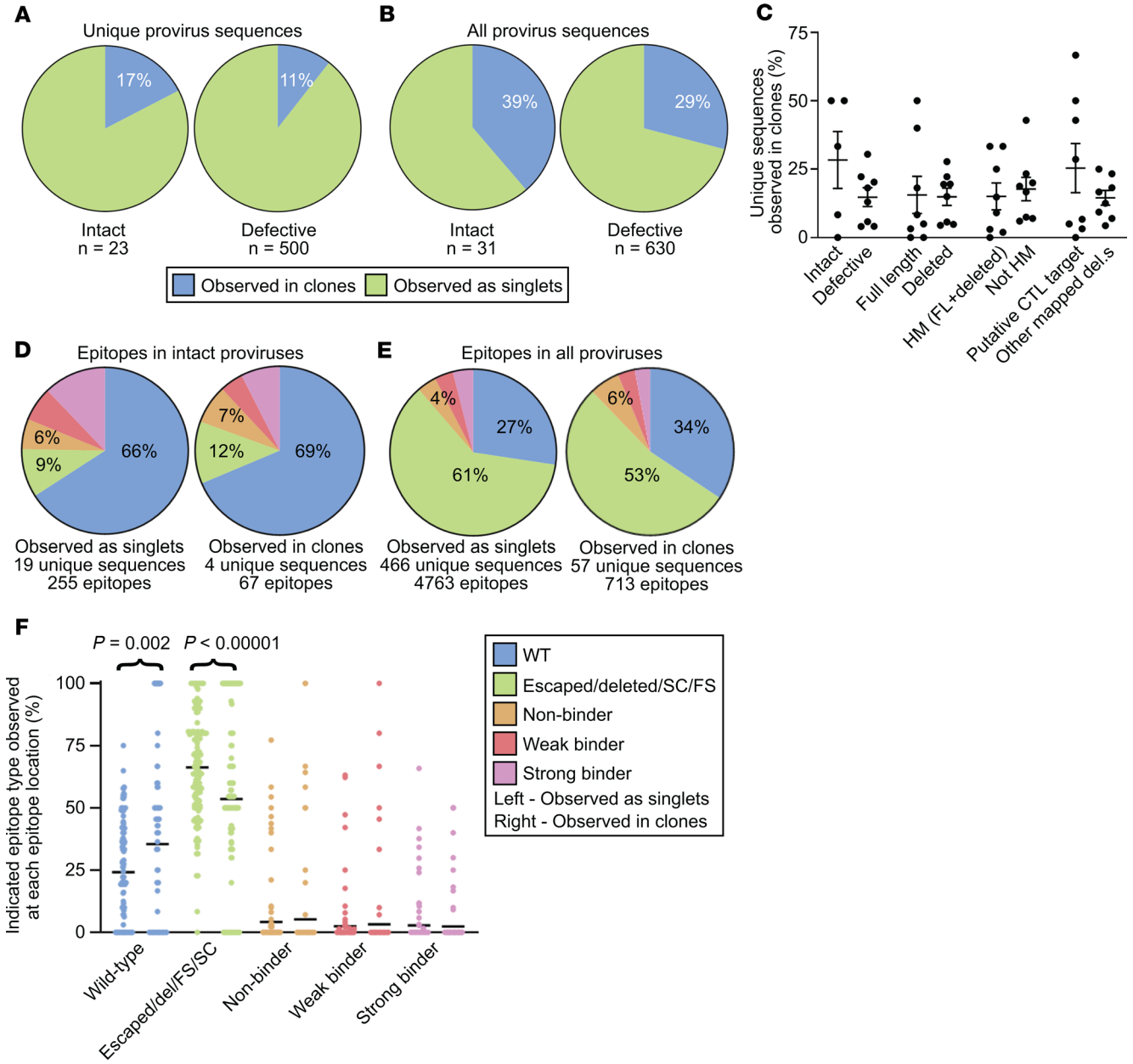

Figure 6. The proportion of intact proviruses observed in large clones is similar to the proportion of defective proviruses observed in large clones. (**A** and **B**) Summary pie charts of all intact (left) and defective (right) proviruses from all participant time points in this study showing proportions observed in clones (blue) versus those observed as singlets (green). In **A** and **C**–**F**, each clone family was collapsed into one unique sequence for counting purposes. In **B**, clone families were not collapsed. (**C**) Percentage of proviruses of the indicated type observed in clones relative to total proviruses of the indicated type from each individual, with mean and SEM shown. The putative CTL target grouping includes intact, FL HM, PD, and FS. There was no statistically significant difference in each pair when the nonparametric Kruskal-Wallis test was applied. (**D**) Summary pie charts of Gag, Pol, and Nef epitope sequences from predicted dominant epitopes from intact proviruses that were observed as singlets (left) or observed in clones (right). (**E**) Summary pie charts of Gag, Pol, and Nef epitope sequences from predicted dominant epitopes from all proviruses observed as singlets (left) or observed in clones (right). (**F**) Percentage of indicated epitope type relative to total epitope sequences at each of 125 predicted dominant epitopes from 8 individuals, with mean shown in black. Left: Epitope sequences from proviruses observed as singlets. Right: Epitope sequences from proviruses observed in clones. The nonparametric Wilcoxon's test was applied and significant *P* values are shown; see also Supplemental Table 3. **D**, **E**, and **F** share a figure legend. Del, epitope within mapped deletion in provirus; SC or FS, stop codon or frameshift observed between relevant gene's start codon and epitope; HM, hypermutated; FL, full-length.

for a single frameshift in Pol would have been counted as defective by near-full-length sequencing and intact by the IPDA.

If CTL selection shapes the proviral landscape in people on ART, we hypothesized that controllers on ART would have a higher proportion of escape or unrecognized epitopes than chronic progressors on ART because controllers have strong and polyfunctional HIV-1–specific CTL responses (49). However, we found no statistically significant difference between the proportions of wild-type and unrecognized epitopes in HIV-1 proviruses when we compared chronic progressors on ART to controllers on ART (Figure 9, E and F).

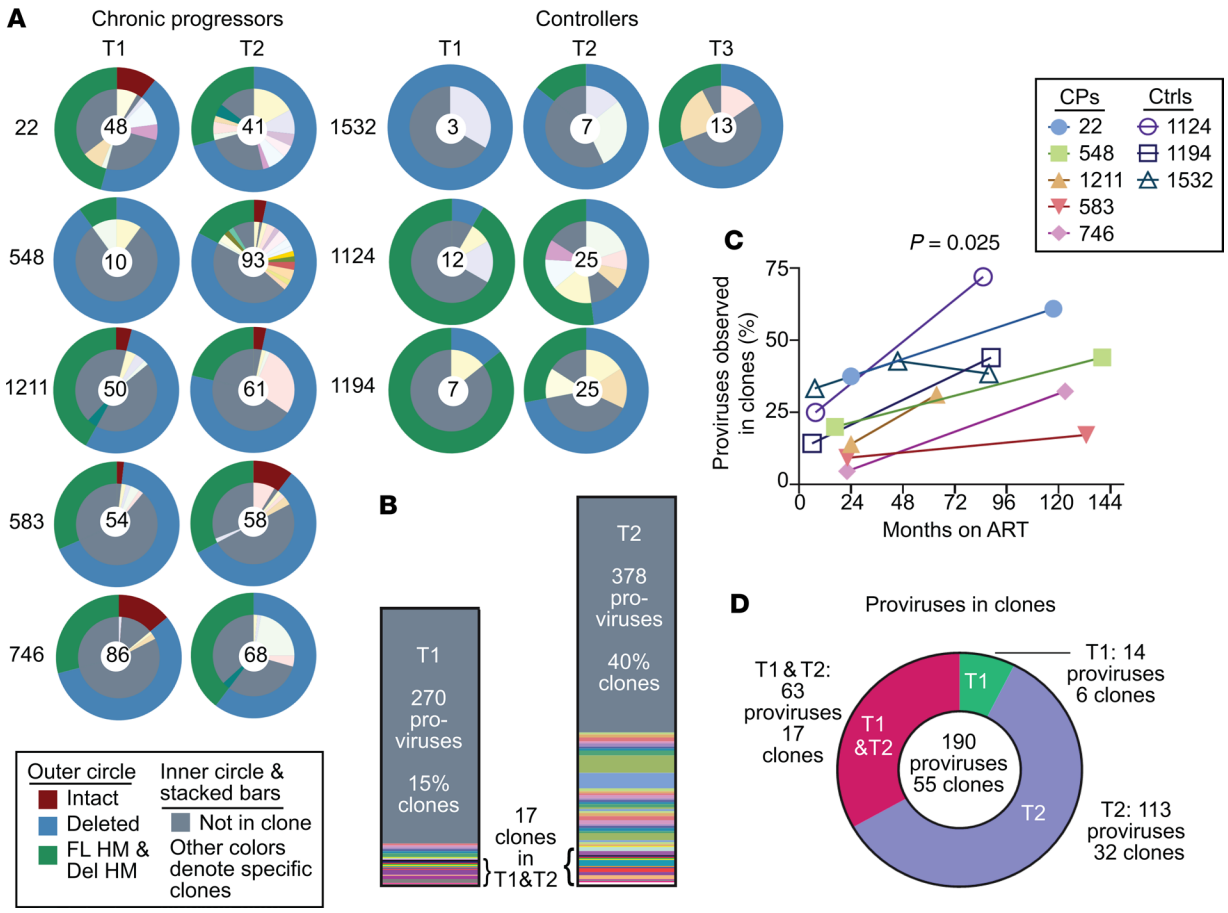

Figure 7. The proportion of HIV-1 proviruses observed in clones increases with time on suppressive ART. (**A**) Summary of provirus type and clonality for all 8 individuals. The outer circles' colors denote whether the provirus is intact, deleted but not hypermutated, or hypermutated with or without deletions. The inner circle denotes unique provirus sequences (gray) and those observed in clones (other colors). In the center is the number of proviruses sampled per participant time point. (**B**) Stacked bar chart summary of clonality for all 8 participants. Time point 3 for participant 1532 is not included. Gray color indicates unique provirus sequences. Each color and its height indicate a clone family and the number of proviruses observed in it. (**C**) Percentage of proviruses observed in clones by months on suppressive ART. Pearson's $r = 0.54$. Pearson's P value shown. (**D**) Of proviruses observed in clones, pie chart of the percentage found in time point 1 only (green), time point 2 only (purple), or found in both time points (red). T1, time point 1; T2, time point 2; HM, hypermutated; FL, full-length; Del, deleted; CPs, chronic progressors; Ctrls, controllers.

We asked whether the percentage of escaped epitopes in controllers increased over long periods of time on ART, implying activation and immune clearance during ART. We found the opposite — that the percentage of wild-type epitopes that could be recognized by CTLs increased in controllers on ART at later time points (Figure 9G). Together, these results indicate that persistence of defective proviruses in controllers on ART does not depend on escape from HIV-1–specific CTLs.

We also examined clonal expansion in controllers. Controllers on ART have clonally expanded cells harboring HIV-1 proviruses, as has been shown in controllers who are ART naive (Figure 7, A–C, and refs. 52, 62). Furthermore, the proportions of proviruses that are observed in clones in controllers on ART are similar to the proportions seen in chronic progressors on ART and, as with chronic progressors, the proportions observed in clones increases over time on ART (Figure 7C), suggesting that the mechanisms driving clonal proliferation of infected cells in people on ART are not greatly altered by strong HIV-1–specific CTL responses.

Discussion

Our intensive study of 8 individuals on prolonged suppressive ART provides insights into HIV-1 persistence. First, cells harboring an intact proviral sequence are just as likely — if not more — to be observed in clones as cells harboring defective proviruses. This is important because only intact proviruses can contribute to viral rebound, and understanding the proliferative processes that preserve them is critical to finding a cure. Intact and replication-competent proviruses have been reported in clones previously by integration site and near-full-length sequence analysis, but our study provides the first direct comparison to our knowledge of the clonality of intact and defective proviruses in vivo (17, 18, 43–46, 48, 63). Furthermore, we found that intact proviruses observed in clones do not have a dramatically different proportion of epitopes that can be recognized by CTLs compared with intact proviruses observed as singlets. These results suggest that cells harboring intact proviruses proliferate in vivo without experiencing great negative selective pressure by viral cytopathic effects or virus-specific CTLs. These findings were unex-

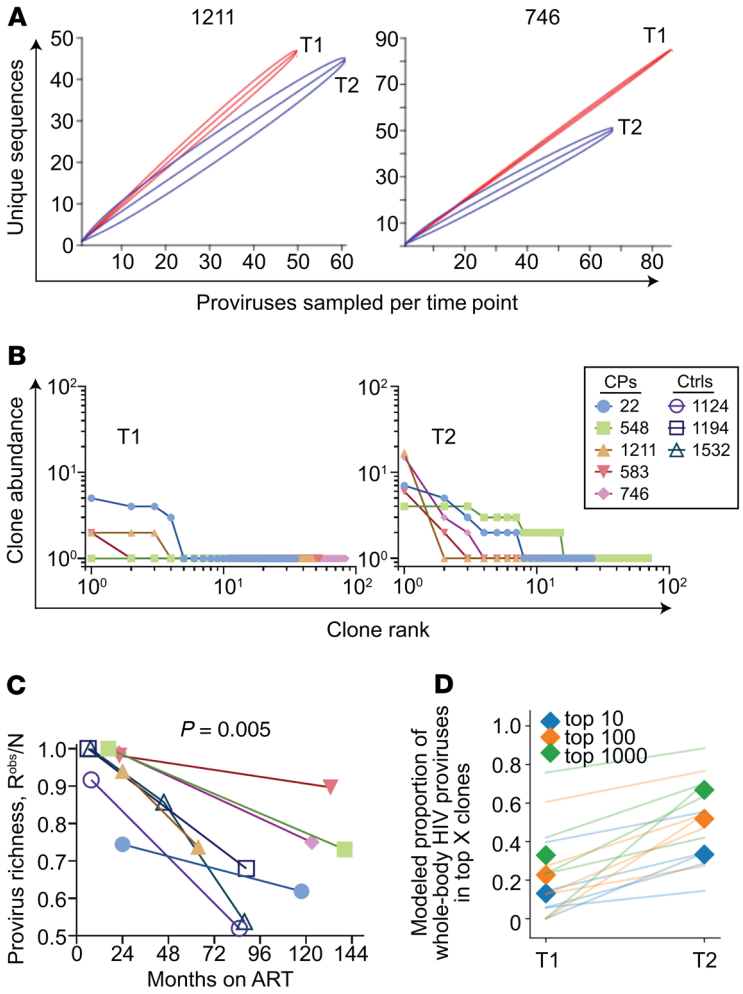

Figure 8. HIV-1 proviruses are concentrated into large clones during ART. (**A**) Rarefaction curves of observed provirus richness in 2 chronic progressor participants. See also Supplemental Figure 4. (**B**) Rank-abundance curves of provirus clones at time point 1 (T1) and T2. (**C**) Provirus richness divided by the number of sampled proviruses by months on suppressive ART. Pearson's r = –0.67. Pearson's P value shown. (**D**) The data from **B** were fit into a power-law model, where clone abundance is directly proportional to 1/rank$^{\alpha}$. See Supplemental Figure 5 for details. Modeled reservoir extrapolations illustrate the proportion of whole-body HIV-1 proviruses contained in the top 10, 100, and 1000 clones, respectively (colors). Proviruses are increasingly concentrated among dominant large clones over time for all CPs (lines) and when averaged across individuals (shaded diamonds). In **B** and **D**, analyses exclude controllers given low sampling depth. CPs, chronic progressors; Ctrls, controllers.

pected, given that in vitro proliferation of cells harboring intact proviruses appears to be less robust than proliferation of cells harboring defective proviruses, presumably due to viral cytopathic effects following induction of viral gene expression (55). However, such in vitro proliferation is stimulated via broad T cell activation using a combination of anti-CD3 and anti-CD28 antibodies and is known to lead to HIV-1 gene expression. In contrast, the stimuli driving proliferation of HIV-1–infected cells in vivo have not been clearly defined, and it is likely that there are several different mechanisms at play, including homeostatic proliferation, antigen-driven proliferation, and integration site–driven proliferation. Regardless of which stimuli are operative, we did not find a smaller ratio of intact to defective proviruses in large clones, nor a smaller fraction of proviruses capable of generating recognizable CTL epitopes in clones. Taken together, these results suggest that HIV-1 gene expression and subsequent immune clearance do not greatly affect clonal expansion of HIV-1–infected cells in vivo.

We also demonstrate that the proportion of proviruses observed in clones increases with time on ART, consistent with previous work (37, 40). We used rarefaction curves to demonstrate that our sample sizes are large enough to detect a significant decrease in provirus richness with time. Using modeled extrapolations of rank-abundance curves from our data set, we report that our data set, like others, appears to fit a power-law model (38). We report the observation of a trend toward increasing proportions of proviruses among the largest clones, illustrated by a higher power-law exponent at later time points. This suggests that, despite the extremely slow decay of total HIV-1 proviruses over time on ART, persistent HIV-1 is increasingly concentrated in a smaller number of extremely large clones during ART. This finding awaits confirmation with larger data sets and prompts fundamental questions about HIV-1 persistence. Are proliferation and decay dynamics of HIV-1–infected cells in people on ART a natural consequence of T cell biology and aging with HIV-1 simply a bystander in these processes, or is the concentration of HIV-1–infected cells into smaller numbers of large clone families driven by other factors? This question could be addressed by examining whether all resting memory CD4$^+$ T cells, and not just infected cells, concentrate over long intervals into extremely large clones.

We uncovered interesting, if subtle, differences between the persistence of intact and defective proviruses in people on long-term suppressive ART. By near-full-length sequencing, we found no significant changes over time in the distribution of defective proviruses, the proportion of intact proviruses, or the proportion of proviruses capable of gene expression (intact, PD, FS, FL HM). However, the number of intact proviruses recovered using near-full-length sequencing was low (n = 31). Therefore, we used the IPDA to examine a larger number of intact proviruses from the same samples. Our

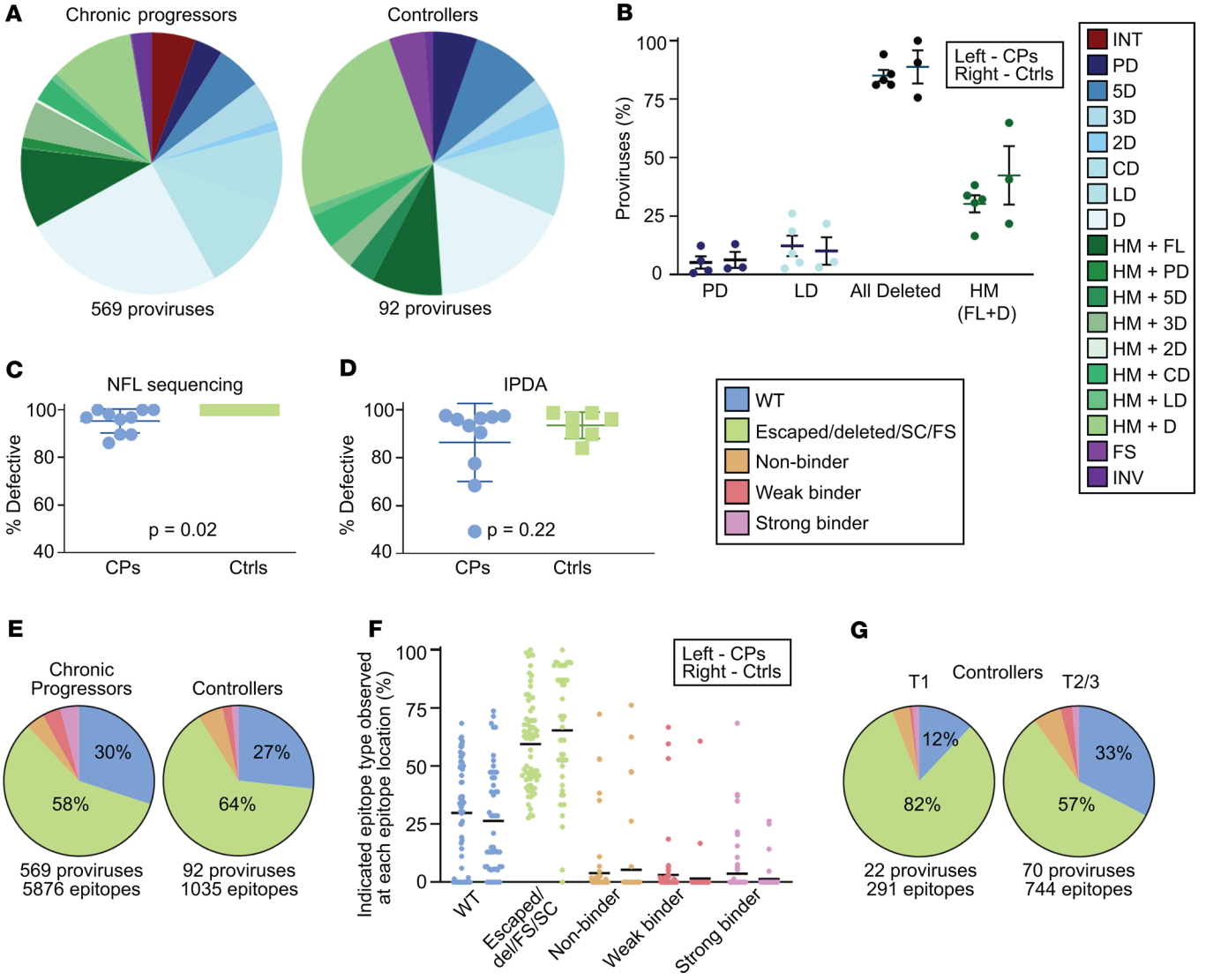

Figure 9. Persistence of defective HIV-1 proviruses in controllers on ART is similar to that of chronic progressors on ART. (**A**) Summary of provirus types obtained from all time points of the chronic progressors (left) and controllers (right) maintained on long-term suppressive ART. (**B**) Percentage of various types of defective proviruses relative to total sampled proviruses from 5 chronic progressors (left) and 3 controllers (right), with mean and SEM shown in black. Only 4 categories are shown because only these types are present in all 3 controllers. (**C** and **D**) Plots of the percentage of defective proviruses at each participant time point for chronic progressors (left) and controllers (right) on ART by near-full-length sequencing (**C**) and by IPDA (**D**). Two-tailed *t* test *P* values shown. (**E**) Summary pie charts of Gag, Pol, and Nef epitope sequences from predicted dominant epitopes from proviruses isolated from chronic progressors and controllers. (**F**) Percentage of indicated epitope type relative to total epitope sequences at each of 75 (CP, left) or 50 (Ctrl, right) predicted dominant epitopes from 5 CPs and 3 Ctrls, with the mean shown in black. (**B** and **F**) No significant differences were observed when the nonparametric Kruskal-Wallis test was applied. (**G**) Summary pie charts of Gag, Pol, and Nef epitope sequences from Ctrls at time point 1 (T1) and T2/3. **E**–**G** share a figure legend. CPs, chronic progressors; Ctrls, controllers; INT, intact; PD, deletion in packaging signal or major splice donor site; LD, deletion of more than 75% of HIV-1 genome length; HM, hypermutated; FL, full-length; FS, intact provirus save for 1 frameshift (**A**) or preceding frameshift mutation (**E**–**G**); INV, deleted with inverted sequence; Del, deleted; SC, preceding stop codon; NFL, near full-length; IPDA, intact proviral DNA assay.

results, combined with IPDA data from other longitudinal samples from people on ART, provide evidence for the slow decay of intact proviruses in most individuals, consistent with previous studies using the QVOA (3, 10, 11, 55, 56). That QVOA and IPDA yield similar decay rates in people on long-term suppressive ART was not necessarily expected given that the IPDA measures all intact proviruses and the QVOA measures only those that release replication-competent virus after a single round of T cell activation in vitro. These observations suggest that the ratio of intact-inducible proviruses to total intact proviruses remains constant over time. This implies that cells with intact-inducible proviruses do not experience a greater negative selective pressure than cells with intact-noninducible proviruses, whether by HIV-1–specific CTLs or viral cytopathic effects. By contrast, the average decay rate of defective and total proviruses is slower, with a larger proportion of individuals on ART experiencing expansion of defective proviruses (55, 56). Because total HIV-1 DNA includes mostly defective proviruses, our work is consistent with previous studies demonstrating that the decay rate of total HIV-1 DNA is on average very slow and that in many people there are slight increases over time on ART (64, 65).

Although slow decay of intact proviruses could be detected with a sensitive method (IPDA), our analysis of predicted dominant CTL epitopes in proviruses uncovered no increase in escaped or unrecognized epitopes in HIV-1 proviruses that persist over long periods of time on ART. Whether due to low HIV-1 expression during ART, the majority of CTL selection occurring before sampling, or CTL exhaustion, this work suggests that HIV-1–specific CTLs do not have a major negative selective effect on infected cells that persist for years in treated individuals. An important caveat is that the great majority of proviruses analyzed were defective, and so it is possible that a negative selective effect on intact proviruses went undetected using these methods. An alternative possibility is that CTLs play a role in maintaining latency, as studies of $CD8^+$ T cell depletion in nonhuman primate models of treated HIV-1 infection indicate (66, 67).

Little is known about the proviral landscape in controllers on ART. We studied these individuals to determine whether strong and polyfunctional HIV-1–specific CTLs shape the proviral landscape in people on ART. We found no intact proviruses among 92 sequenced proviruses from controllers on ART. Also, no significant differences between controllers on ART and chronic progressors on ART were observed in the landscape of defective proviruses. Further, there was no increase in the proportion of escaped or unrecognized epitopes seen over time in the defective proviruses of controllers on ART. Large proviral clones were observed in controllers on ART, as reported in treatment-naive controllers (62). The proportion of proviruses found in large clones was similar to that of chronic progressors. Taken together, these data indicate no significant role for HIV-1–specific CTLs in checking clonal expansion or significantly shaping the defective provirus landscape in controllers or chronic progressors on ART. It remains unclear whether CTLs play a role in the decay of intact proviruses during suppressive ART in controllers and chronic progressors.

This study has limitations. The immunodominant epitopes chosen for analysis were based on published data and are predicted rather than validated as immunodominant. It is likely that a subset of each participant's predicted epitopes were immunodominant. Regardless, Figure 4 demonstrates no dramatic shift in the CTL escape status from the first time point to the second. Another limitation is the number of proviruses that can be obtained and analyzed using the near-full-length sequencing technique. Also, infected cells in our study were isolated from peripheral blood rather than lymphoid tissue. However, recent work demonstrated that clonal populations of infected cells and viral RNA expression in people on long-term ART in peripheral blood and lymph nodes are not significantly different, nor are HIV-1 proviruses in these 2 compartments genetically divergent (68). A final limitation is that this study did not examine proviruses from naive or memory $CD4^+$ T cell subsets. Important studies have noted differences in the distribution of HIV-1 proviruses in different memory cell subsets (15, 18, 47, 69, 70). We have recently analyzed the distribution and inducibility of replication-competent HIV-1 proviruses in resting $CD4^+$ T cells from the naive, central, transitional, and effector memory subsets and did not observe substantial and consistent differences in the frequency or inducibility of intact HIV-1 proviruses (71). In this study, multiple rounds of T cell activation were used to reverse latency and allow viral outgrowth from purified cell populations representing different subsets. We observed substantial interperson variability and no consistent enrichment of intact or inducible proviruses in any T cell subset. Nevertheless, further examination of proviral dynamics in T cell subsets may be warranted.

In summary, we found over long time intervals that the proviral landscape is shaped by proliferative processes but not clearly by negative selective forces resulting from viral gene expression, including viral cytopathic effects and lysis by CTLs. We show that $CD4^+$ T cells harboring intact HIV-1 proviruses clonally expand to a similar degree as cells harboring defective proviruses despite their greater potential susceptibility to these negative factors. Further, we found that persistent HIV-1 is increasingly concentrated in a smaller number of extremely large clones over long periods of time on ART. Although we observed a slow decay of intact proviruses over time, no discernable longitudinal patterns were observed among specific types or groupings of defective proviruses. We also found that the dynamics of provirus persistence and clonal proliferation are largely similar in controllers on ART and chronic progressors on ART, with the exception that fewer intact proviruses were found in controllers on ART by near-full-length sequencing. This work suggests that CTL selection does not appear to significantly alter the defective provirus landscape or clonal proliferation of infected cells, even over long periods of time on ART.

Methods

Study participants

Eight HIV-1 subtype B^+ individuals were included. Seven are from the SCOPE and OPTIONS cohorts at UCSF. One is from the Johns Hopkins HIV Clinical Cohort. Table 1 contains characteristics of participants and samples.

Chronic progressors. The 5 chronic progressors were viremic for 26 months before ART (range 19–37 months). Time point 1 (T1) was 22 months after ART initiation (range 16.6–24.0 months). Each maintained undetectable viral loads — checked at least every 6 months but typically more often — beginning shortly after starting ART through T2, with a few exceptions as follows. Participant 746: between T1 and T2 (8.4 years), 1 viral load was 42.4. Participant 548: between T1 and T2 (10.3 years), 2 viral loads were 151 and 93. Also, there were 5 instances in which viral loads were measured more than 6 months apart for 548: once at 12 months, 3 times at 9 months, and once at 7 months. Participant 1211 had 1 viral load of 50 before T1, a viral load of 127 at T1, and another of 134 between T1 and T2. Also, for 1211 no viral loads were collected for 17 months before T2.

Controllers. Three participants with natural control of HIV-1 had undetectable or very low levels of viremia over 4, 6, and 14 years of observation before ART initiation. Participant 1124 had 37 undetectable viral loads and 1 detectable viral load of less than 200 copies/mL before ART. Participant 1194 had 3 undetectable viral loads and 10 viral loads of less than 200 before ART. Participant 1532 had 49 undetectable viral loads, 13 detectable viral loads of less than 200, and 2 viral loads between 200 and 400 before ART. The medical providers for the controllers initiated ART for clinical reasons unrelated to this study. For all 3, each viral load after starting ART was

undetectable. Samples used here were from the time period on ART. The controllers were on ART for 7 months before T1 (range 6.0–7.6 years). Two controllers have protective MHC class I alleles. One has B57:03 and another has B14:02 and Cw08:02 (72).

Isolation of resting CD4+ T lymphocytes

Peripheral blood mononuclear cells (PBMCs) from participant 22 at T2, participant 1124 at T2, participant 1194 at T2, and participant 1532 at T3 were isolated using density centrifugation on a Ficoll-Hypaque gradient. For other samples, viably frozen PBMCs were quick-thawed at 37°C, rinsed, centrifuged (5 minutes, 500 *g*), resuspended, and rested overnight in an incubator. $CD4^+$ T cells were isolated 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 (biotin-conjugated anti-CD25, anti-biotin MicroBeads, CD69 MicroBead Kit II, and anti–HLA-DR MicroBeads, Miltenyi Biotec).

DNA extraction and limiting-dilution PCR

Resting $CD4^+$ T cell DNA was extracted in 100- to 200-kb fragments to minimize disruption of HIV-1 provirus genomes (Qiagen Gentra Puregene Cell Kit). DNA was subjected to a nested limiting-dilution PCR protocol modified from Bruner et al. using Platinum Taq HiFi Polymerase (Life Technologies) and a seventh inner PCR reaction spanning HXB2 coordinates 651 to 9632 (14). PCR products were visualized in 1% agarose gels. Any well that appeared to contain more than 1 provirus was discarded.

Provirus sequencing

PCR products (QIAquick Gel Extraction Kit, Qiagen) were directly sequenced to minimize PCR-induced error by Sanger sequencing (Genewiz) and/or Illumina MiSeq. Sanger sequencing was performed on PCR products of less than 2000 bp. Most PCR products of more than 2000 bp were sequenced by Illumina MiSeq (paired-end 300-bp reads) after library preparation (Nextera XT, Illumina), and a subset of these were also sequenced by the Sanger method. Redundant sequencing was performed to ensure fidelity of the 2 methods. Two pipelines — a resequencing pipeline using HXB2 and a de novo pipeline — were developed in-house (CLC Genomics Workbench, Qiagen). The resequencing pipeline was used first. If the provirus contained a deletion or had a mapped read percentage below 80%, the de novo pipeline was used for confirmation. Each called nucleotide in each provirus was of high quality; Sanger sequencing was repeated as needed to obtain high quality nucleotide calls. Any uncertain nucleotide was marked as unknown. Then the inner PCR sequences were aligned and compared with HXB2 (CodonCode Aligner) to construct near-full-length sequences for each provirus.

Provirus characterization

Hypermutation by APOBEC3G, APOBEC3F, or both was determined using the near-full-length provirus sequence and the Los Alamos Hypermut algorithm (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html) (53). Most deletions were mapped but many wells experienced amplification of a subset of the inner PCR bands that did not span the outer PCR region. The sites of deletion or mutation of these proviruses were unmappable and these were presumed to be deleted.

For the purposes of our study, intact proviruses were those without deletions or hypermutation. Additionally, each intact provirus had no stop codons or frameshifts in the amino acid sequence of the 9 major HIV-1 gene products. The few exceptions include participant 548, in whom all Gag sequences contained a stop codon 3 amino acids before HXB2's stop codon. Participant 746 had 1 intact provirus with a stop codon in Vif one-third of the way into the protein and a frameshift in Nef 5 amino acids into the protein. Another intact provirus from 746 had a stop codon in Nef 23 amino acids from the end of the protein. These proviruses were considered intact here.

Provirus sequence alignment

The near-full-length provirus sequences from each participant were aligned to the reference sequence HXB2 in 3 steps. First, sequences were aligned to HXB2 (Clustal W, http://www.ebi.ac.uk/Tools/msa/clustalw2). This algorithm encounters difficulty with the deletions common in defective proviruses, the Clustal W output nucleotide alignment was adjusted by hand, referencing each sequence's reconstruction in CodonCode Aligner as necessary. Finally, for each of the 9 HIV-1 gene products, the nucleotide alignment was translated to amino acids and the amino acid alignment was adjusted by hand.

Phylogenetic analysis

Gag and *env* regions were selected for phylogenetic analysis. Neighbor-joining trees based on *P* distance were constructed in MEGA 7.0 (http://megasoftware.net) using the subset of proviruses that had known sequence and no large deletions in the region of interest (*gag* region, HXB2 888-2143; *env* region, HXB2 6324-8190). Reference subtype B sequences were used as outgroups to root the tree (HXB2, NL, TH, and US: GenBank accession K03455, AY423387, AY173951, and DQ853437, respectively). One hundred and thirty of 661 proviruses were not included in these phylogenetic trees because they had deletions within the specified regions. Each of these 130 proviruses was individually compared with the proviruses from the same participant that were included in the phylogenetic analyses and found to have a similar pattern of nucleotide and amino acid polymorphisms.

IPDA

The IPDA was performed as described previously (55). Participant 583 required modifications as follows: ACTGGTGAGTACGCTGAAA-5′ probe and AGTGGTGCAAAGAGAAAAAAGAGC-3′ forward primer. Half-lives of decay were calculated as described previously (55).

Epitope analysis

Each participant's cells were HLA typed by the Johns Hopkins University Immunogenetics Laboratory. The best-defined CTL epitopes in Gag, Pol, and Nef for each participant based on their HLA type were identified using the Los Alamos National Lab's "Best-defined CTL/$CD8^+$ Epitope Summary," a regularly updated list of well-described, fine-mapped HLA class I–restricted epitopes (Supplemental Table 2). For each provirus, we identified the amino acid sequence of each relevant epitope, or we noted if the epitope lay in deleted/unsequenced regions or if there was a frameshift or stop codon between the start codon of the relevant gene and the epi-

tope. We used the Los Alamos National Lab's "CTL/CD8+ Epitope Variants and Escape Mutations" table (http://www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html), which summarizes published data on epitope variants by HIV subtype and HLA type, to classify each epitope as wild-type/recognized or escaped.

Accession codes

The provirus sequences reported in this study were deposited in the GenBank database (787 nucleotide sequences from 661 proviruses, with accession numbers MT033120 to MT033906). Illumina sequencing data were deposited in the Sequence Read Archive under BioProject accession PRJNA613107.

Statistics

In all analyses in this study, a P value less than 0.05 was considered significant.

IPDA analyses. To estimate the population decay parameters for total, intact, 3′del, and 5′del sequences, we took the natural logarithm of those values and performed linear regression in a mixed-effects modeling approach. This assumes decay follows a single exponential, which is reasonable given that the first time point was always more than 6 months after ART initiation. Code was implemented in Python using the statsmodels mixedlm package (github.com/dbrvs).

Provirus type analyses. For each patient time point, we generated a provirus-type prevalence rate, the number of sequences of that provirus type over the total number sequenced. Two-tailed paired t tests with Bonferroni's correction were applied for longitudinal analysis, and the nonparametric Kruskal-Wallis test was applied for analyses where samples were not paired.

Epitope analyses. There were 125 predicted dominant epitopes from the 8 participants. For each dominant epitope at each time point, a percentage was generated for each category of epitope sequence (e.g., wild-type, escape, deleted, nonbinder, etc.) over the total number of epitope sequences from that participant with sequence information available. If the particular dominant epitope of a certain provirus lay in a region for which we did not have sequencing information, we excluded that provirus from that dominant epitope's analysis. This analysis included zeroes for the categories not observed within each epitope. We used a nonparametric Wilcoxon's signed-rank test because of the skewed distribution with a large number of zeroes.

Correlation analyses. Percentages of proviruses observed in clones and provirus richness divided by number of sampled proviruses by months on suppressive ART were analyzed using Pearson's correlation.

Rarefaction. Observed provirus sequence richness from T1 and T2/T3 were compared using rarefaction analysis (Past 3, http://folk.uio.no/ohammer/past).

Study approval

The studies from which samples were obtained were approved by the UCSF Committee on Human Research and the Johns Hopkins Institutional Review Board. All participants provided written informed consent before enrollment.

Author contributions

YCH, AARA, JDS, and RFS designed the study. FMH, SGD, YCH, AARA, RH, MRK, JCK, RDM, and RFS identified and provided participant samples. AARA, KMJ, SJ, DNR, and YCH acquired the data. AARA, KMJ, SJ, DNR, DBR, JTS, BASN, and RFS analyzed the data. AARA and RFS wrote the manuscript.

Acknowledgments

We thank the study participants for their valuable contribution. We thank Katherine Bruner, Francesco Simonetti, Joel Blankson, Stuart Ray, Karen Beemon, Srona Sengupta, Alexandra Bender, and Andrew Timmons for advice and discussion. This work was supported by the NIH Martin Delaney I4C (UM1 AI126603), BeatHIV (UM1 AI126620), and DARE (UM1 AI12661) collaboratories; the Johns Hopkins Center for AIDS Research (P30AI094189); NIH grants T32 AI007291-27 and K08 AI143391-01; the Howard Hughes Medical Institute; the Bill and Melinda Gates Foundation (OPP1115715); an unrestricted research grant from Gilead (125304); and the Pearl M. Stetler Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Address correspondence to: Robert F. Siliciano, Room 879, Edward D. Miller Research Building, 733 North Broadway, Baltimore, Maryland 21205, USA. Phone: 410.955.2958; Email: rsiliciano@jhmi.edu.

YCH's present affiliation is: Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, Connecticut, USA.

1. Finzi D, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997;278(5341):1295–1300.
2. Chun TW, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A*. 1997;94(24):13193–13197.
3. 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*. 1999;5(5):512–517.
4. Wong JK, et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science*. 1997;278(5341):1291–1295.
5. Davey RT, et al. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc Natl Acad Sci U S A*. 1999;96(26):15109–15114.
6. Rothenberger MK, et al. Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proc Natl Acad Sci U S A*. 2015;112(10):E1126–E1134.
7. Sengupta S, Siliciano RF. Targeting the latent reservoir for HIV-1. *Immunity*. 2018;48(5):872–895.
8. Eisele E, Siliciano RF. Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity*. 2012;37(3):377–388.
9. Zerbato JM, McMahon DK, Sobolewski MD, Mellors JW, Sluis-Cremer N. Naive CD4+ T cells harbor a large inducible reservoir of latent, replication-competent human immunodeficiency virus type 1. *Clin Infect Dis*. 2019;69(11):1919–1925.
10. Siliciano JD, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med*. 2003;9(6):727–728.
11. Crooks AM, et al. Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J Infect Dis*. 2015;212(9):1361–1365.
12. Wang Z, Simonetti FR, Siliciano RF, Laird GM. Measuring replication competent HIV-1: advances and challenges in defining the latent reservoir. *Retrovirology*. 2018;15(1):21.
13. Ho YC, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier

to HIV-1 cure. *Cell*. 2013;155(3):540–551.
14. Bruner KM, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med*. 2016;22(9):1043–1049.
15. Hiener B, et al. Identification of genetically intact HIV-1 Proviruses in specific CD4+ T cells from effectively treated participants. *Cell Rep*. 2017;21(3):813–822.
16. 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;113(31):8783–8788.
17. Pinzone MR, et al. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun*. 2019;10(1):728.
18. Lee GQ, et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J Clin Invest*. 2017;127(7):2689–2696.
19. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*. 2003;300(5622):1112.
20. Desimmie BA, Delviks-Frankenberrry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol*. 2014;426(6):1220–1245.
21. Harris RS, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell*. 2003;113(6):803–809.
22. Sanchez G, Xu X, Chermann JC, Hirsch I. Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. *J Virol*. 1997;71(3):2233–2240.
23. Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci U S A*. 1993;90(15):6900–6903.
24. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*. 2012;2(10):a006882.
25. Archin NM, et al. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*. 2012;487(7408):482–485.
26. Mbonye U, Karn J. The molecular basis for human immunodeficiency virus latency. *Annu Rev Virol*. 2017;4(1):261–285.
27. Pollack RA, et al. Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. *Cell Host Microbe*. 2017;21(4):494–506.e4.
28. Huang SH, et al. Latent HIV reservoirs exhibit inherent resistance to elimination by CD8+ T cells. *J Clin Invest*. 2018;128(2):876–889.
29. Wiegand A, et al. Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proc Natl Acad Sci U S A*. 2017;114(18):E3659–E3668.
30. Barton K, et al. Broad activation of latent HIV-1 in vivo. *Nat Commun*. 2016;7:12731.
31. Kearney MF, et al. Origin of rebound plasma HIV includes cells with identical proviruses that are transcriptionally active before stopping of antiretroviral therapy. *J Virol*. 2016;90(3):1369–1376.
32. Imamichi H, et al. Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. *AIDS*. 2014;28(8):1091–1099.
33. Casartelli N, Guivel-Benhassine F, Bouziat R, Brandler S, Schwartz O, Moris A. The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells. *J Exp Med*. 2010;207(1):39–49.
34. Thomas AS, et al. T-cell responses targeting HIV Nef uniquely correlate with infected cell frequencies after long-term antiretroviral therapy. *PLoS Pathog*. 2017;13(9):e1006629.
35. Sharaf R, et al. HIV-1 proviral landscapes distinguish posttreatment controllers from noncontrollers. *J Clin Invest*. 2018;128(9):4074–4085.
36. Bale MJ, Kearney MF. Review: HIV-1 phylogeny during suppressive antiretroviral therapy. *Curr Opin HIV AIDS*. 2019;14(3):188–193.
37. Wagner TA, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014;345(6196):570–573.
38. Reeves DB, Duke ER, Wagner TA, Palmer SE, Spivak AM, Schiffer JT. A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nat Commun*. 2018;9(1):4811.
39. Rasmussen TA, et al. The effect of antiretroviral intensification with dolutegravir on residual virus replication in HIV-infected individuals: a randomised, placebo-controlled, double-blind trial. *Lancet HIV*. 2018;5(5):e221–e230.
40. Cohn LB, et al. HIV-1 integration landscape during latent and active infection. *Cell*. 2015;160(3):420–432.
41. Bailey JR, et al. Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *J Virol*. 2006;80(13):6441–6457.
42. Maldarelli F, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*. 2014;345(6193):179–183.
43. Simonetti FR, et al. Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci U S A*. 2016;113(7):1883–1888.
44. Lorenzi JC, et al. Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci U S A*. 2016;113(49):E7908–E7916.
45. Bui JK, et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog*. 2017;13(3):e1006283.
46. Hosmane NN, et al. Proliferation of latently infected CD4+ T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J Exp Med*. 2017;214(4):959–972.
47. Chomont N, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med*. 2009;15(8):893–900.
48. Wang Z, et al. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc Natl Acad Sci U S A*. 2018;115(11):E2575–E2584.
49. O'Connell KA, Bailey JR, Blankson JN. Elucidating the elite: mechanisms of control in HIV-1 infection. *Trends Pharmacol Sci*. 2009;30(12):631–637.
50. Julg B, et al. Infrequent recovery of HIV from but robust exogenous infection of activated CD4(+) T cells in HIV elite controllers. *Clin Infect Dis*. 2010;51(2):233–238.
51. Blankson JN, et al. Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. *J Virol*. 2007;81(5):2508–2518.
52. Boritz EA, et al. Multiple origins of virus persistence during natural control of HIV infection. *Cell*. 2016;166(4):1004–1015.
53. Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. *Bioinformatics*. 2000;16(4):400–401.
54. Sevilya Z, et al. Killing of latently HIV-infected CD4 T cells by autologous CD8 T cells is modulated by Nef. *Front Immunol*. 2018;9:2068.
55. Bruner KM, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature*. 2019;566(7742):120–125.
56. Peluso MJ, et al. Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy. *JCI Insight*. 2020;5(4):132997.
57. Deng K, et al. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature*. 2015;517(7534):381–385.
58. Bender AM, et al. The landscape of persistent viral genomes in ART-treated SIV, SHIV, and HIV-2 infections. *Cell Host Microbe*. 2019;26(1):73–85.e4.
59. Salantes DB, et al. HIV-1 latent reservoir size and diversity are stable following brief treatment interruption. *J Clin Invest*. 2018;128(7):3102–3115.
60. Laskey SB, Pohlmeyer CW, Bruner KM, Siliciano RF. Evaluating clonal expansion of HIV-infected cells: optimization of PCR strategies to predict clonality. *PLoS Pathog*. 2016;12(8):e1005689.
61. von Stockenstrom S, et al. Longitudinal genetic characterization reveals that cell proliferation maintains a persistent HIV type 1 DNA pool during effective HIV therapy. *J Infect Dis*. 2015;212(4):596–607.
62. Veenhuis RT, et al. Long-term remission despite clonal expansion of replication-competent HIV-1 isolates. *JCI Insight*. 2018;3(18):122795.
63. Einkauf KB, et al. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest*. 2019;129(3):988–998.
64. Besson GJ, et al. HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. *Clin Infect Dis*. 2014;59(9):1312–1321.
65. Bachmann N, et al. Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. *Nat Commun*. 2019;10(1):3193.
66. Cartwright EK, et al. CD8(+) lymphocytes are required for maintaining viral suppression in SIV-infected macaques treated with

short-term antiretroviral therapy. *Immunity*. 2016;45(3):656–668.
67. McBrien JB, et al. Robust and persistent reactivation of SIV and HIV by N-803 and depletion of CD8$^+$ cells. *Nature*. 2020;578(7793):154–159.
68. McManus WR, et al. HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *J Clin Invest*. 2019;130:4629–4642.
69. Kulpa DA, et al. Differentiation into an effector memory phenotype potentiates HIV-1 latency reversal in CD4$^+$ T cells. *J Virol*. 2019;93(24):e00969-19.
70. Soriano-Sarabia N, et al. Quantitation of replication-competent HIV-1 in populations of resting CD4$^+$ T cells. *J Virol*. 2014;88(24):14070–14077.
71. Kwon KJ, et al. Different human resting memory CD4$^+$ T cell subsets show similar low inducibility of latent HIV-1 proviruses. *Sci Transl Med*. 2020;12(528):eaax6795.
72. International HIV Controllers Study, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. 2010;330(6010):1551–1557.