

## HIV

# The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation

Melissa-Rose Abrahams<sup>1\*</sup>, Sarah B. Joseph<sup>2,3\*</sup>, Nigel Garrett<sup>4</sup>, Lynn Tyers<sup>1</sup>, Matthew Moeser<sup>3</sup>, Nancie Archin<sup>5</sup>, Olivia D. Council<sup>2</sup>, David Matten<sup>1</sup>, Shuntai Zhou<sup>3</sup>, Deelan Doolabh<sup>1</sup>, Colin Anthony<sup>1</sup>, Nilu Goonetilleke<sup>2,5</sup>, Salim Abdool Karim<sup>4,6</sup>, David M. Margolis<sup>2,5</sup>, Sergei Kosakovsky Pond<sup>7</sup>, Carolyn Williamson<sup>1,8†</sup>, Ronald Swanstrom<sup>3,9†</sup>

Copyright © 2019  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.  
Government Works

Although antiretroviral therapy (ART) is highly effective at suppressing HIV-1 replication, the virus persists as a latent reservoir in resting CD4<sup>+</sup> T cells during therapy. This reservoir forms even when ART is initiated early after infection, but the dynamics of its formation are largely unknown. The viral reservoirs of individuals who initiate ART during chronic infection are generally larger and genetically more diverse than those of individuals who initiate therapy during acute infection, consistent with the hypothesis that the reservoir is formed continuously throughout untreated infection. To determine when viruses enter the latent reservoir, we compared sequences of replication-competent viruses from resting peripheral CD4<sup>+</sup> T cells from nine HIV-positive women on therapy to viral sequences circulating in blood collected longitudinally before therapy. We found that, on average, 71% of the unique viruses induced from the post-therapy latent reservoir were most genetically similar to viruses replicating just before ART initiation. This proportion is far greater than would be expected if the reservoir formed continuously and was always long lived. We conclude that ART alters the host environment in a way that allows the formation or stabilization of most of the long-lived latent HIV-1 reservoir, which points to new strategies targeted at limiting the formation of the reservoir around the time of therapy initiation.

## INTRODUCTION

Infection with HIV-1 results in active viral replication in the face of the host immune response, eventually leading to the loss of the primary target cell for replication of the virus, CD4<sup>+</sup> T cells, and immunodeficiency. The use of multiple potent antiviral drugs stops viral replication and disease progression. However, discontinuation of antiretroviral therapy (ART) results in the rapid rebound of virus, indicating that while therapy suppresses viral replication, HIV-1 is able to persist in an infectious state for years. The best-characterized reservoir in individuals on ART is integrated viral DNA in resting memory CD4<sup>+</sup> T cells (1–3). One quantitative measure of the reservoir is the number of resting CD4<sup>+</sup> T cells that can be induced to produce replication-competent virus after stimulation of the cells in culture, called the quantitative viral outgrowth assay (QVOA). Using this assay, it has been estimated that, in people on therapy, about one in a million resting CD4<sup>+</sup> T cells in the blood can be induced to produce replication-competent virus and that this latent reservoir of infected cells has a half-life of 44 months (4, 5). Given the large number of resting CD4<sup>+</sup> T cells in the body, it is impossible to cure HIV-1 by waiting for the infected cells to decay. In addition, clonal expansion

of latently infected T cells adds another mechanism for persistence of virus in the body over time (6–13).

In people on ART, greater than 90% of the proviral genomes in resting CD4<sup>+</sup> T cells are defective (14, 15). These defective genomes may contribute to continued immune activation and exhaustion (16, 17) but are not the source of the rebound virus if ART is stopped. In contrast, most intact proviruses are theoretically capable of producing replication-competent virus, but the frequency of cells harboring intact proviruses is about 30 times higher than the frequency of cells that can be induced to produce virus in a QVOA (15). In addition, the number of resting CD4<sup>+</sup> T cells producing outgrowth viruses (OGVs) in QVOA increases with extra rounds of cell stimulation (14), indicating that the typical QVOA using a single round of stimulation underestimates the number of inducible proviruses. Together, these results imply that the reservoir of replication-competent proviruses is larger than that measured by standard QVOA. It is currently unknown whether the discrepancy arises because virus expression and outgrowth from resting CD4<sup>+</sup> T cells is a noisy stochastic process, or because latency is generated by multiple mechanisms, some of which are not readily reversed in a standard QVOA, or both.

The most widely accepted model of how the reservoir forms involves the infection of a CD4<sup>+</sup> T cell as it is transitioning to a resting state (18). However, little is known about when this happens during the course of infection. The reservoir is established even when ART is initiated early (19–21), as virus rebounds with the subsequent discontinuation of therapy even when starting ART soon (e.g., within days) after infection. This suggests that there is early and continuous formation of the reservoir during the period before therapy initiation. However, studies of viral DNA have yielded conflicting data about the timing of reservoir formation. One report claimed evidence of continuous introduction of viral DNA into the long-lived reservoir (22), whereas another report found that most of the viral DNA in the reservoir comes from virus replicating around the time of therapy initiation (23). The timing of reservoir formation is an important

<sup>1</sup>Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town 7925, South Africa. <sup>2</sup>Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>3</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>4</sup>Centre for the AIDS Programme of Research in South Africa, University of KwaZulu-Natal, Durban 4013, South Africa. <sup>5</sup>UNC HIV Cure Center and Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>6</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA. <sup>7</sup>Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122, USA. <sup>8</sup>National Health Laboratory Services of South Africa, University of Cape Town, Cape Town 7925, South Africa. <sup>9</sup>Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: risunc@med.unc.edu (R.S.); carolyn.williamson@uct.ac.za (C.W.)

question since an understanding of when the reservoir forms could provide new opportunities for limiting its size.

In this study, we perform a quantitative analysis aimed at understanding when replication-competent virus enters the long-lived latent reservoir. We compared sequences of replication-competent viruses induced from on-ART peripheral resting CD4<sup>+</sup> T cells to viral sequences in blood collected longitudinally pre-ART.

## RESULTS

### Cohort and source of viral sequences for the analysis of the latent reservoir

We investigated when replication-competent viruses enter the reservoir by examining viruses induced to replicate from resting CD4<sup>+</sup> T cells in the context of QVOA. OGVs were isolated from blood-derived cells of nine women on ART. The women were participants in the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 cohort, based in KwaZulu-Natal, South Africa, who were originally enrolled into the cohort during acute/primary HIV-1 infection (24). QVOA measures the replication-competent viral reservoir and estimates the proportion of resting (CD25<sup>low</sup>, CD69<sup>-</sup>, and HLA-DR<sup>-</sup>) CD4<sup>+</sup> T cells that can produce replication-competent virus after stimulation. This provides a representation of viruses capable of rebound upon ART interruption as opposed to assays that evaluate total viral DNA genomes, most of which cannot give rise to replicating virus (14, 15). The women were ART naïve for an average of 4.5 years

(Table 1). ART was initiated in accordance with national guidelines in place at the time, which changed over the course of sample collection from starting ART at 200 CD4<sup>+</sup> T cells/μl to starting ART at 350 CD4<sup>+</sup> T cells/μl. The participants had been on suppressive ART for an average of 5.0 years when blood samples were collected for QVOA (Fig. 1 and Table 1). Plasma-derived viral RNA genomes were sequenced at multiple time points before ART [on average every 6 months starting during acute/early infection (fig. S1)], and these evolving sequences were compared to the sequences of the replication-competent OGVs from QVOAs using samples taken while on therapy (the latent reservoir).

### Sequencing plasma virus from throughout untreated infection and OGVs

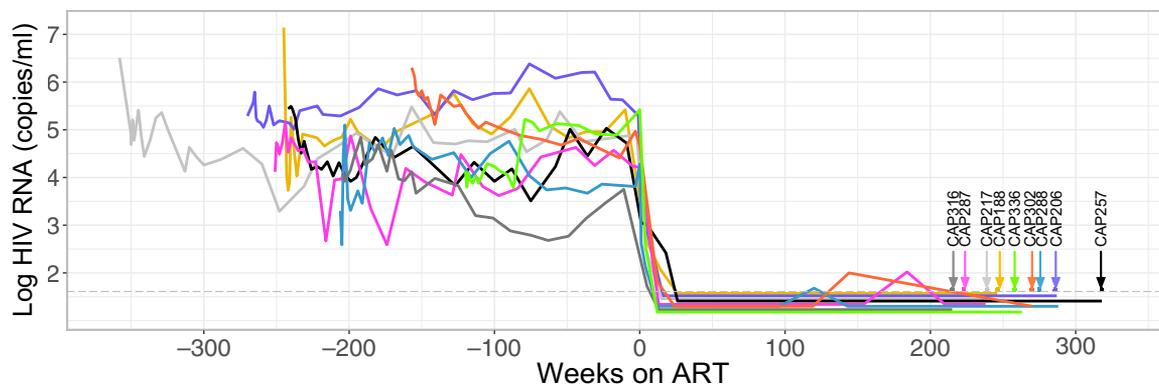
Phylogenetic trees of sequences from longitudinally sampled viruses pre-ART generally have a ladder-like structure as the viral population diverges from its homogeneous founder (25–27), thereby providing a strong phylogenetic and temporal signal for distinguishing populations over time. This allowed us to estimate when each OGV entered the long-lived reservoir by identifying the pre-ART viral sequences that it was most closely related to and therefore the pre-ART time point when it most likely entered the long-lived reservoir. We sequenced five regions (one in *gag*, one in *nef*, and three in *env*) in three viral genes known to have rapid rates of evolution driven by immune responses in the host (28, 29).

Viral RNA from plasma was sequenced at an average of nine timepoints pre-ART per participant using the Illumina MiSeq platform

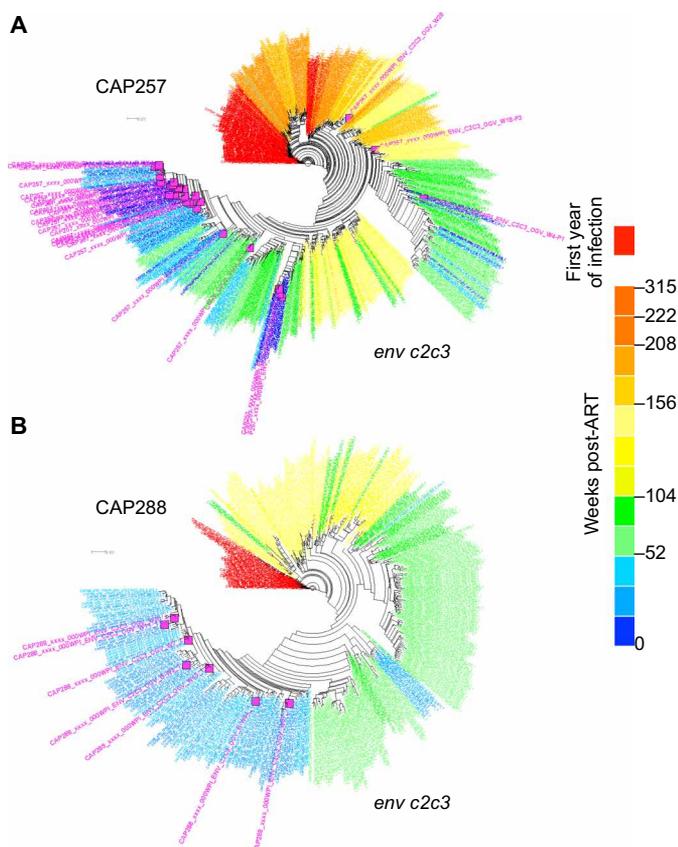
**Table 1. Clinical information for the nine women from the CAPRISA 002 cohort.** EFV, efavirenz; 3TC, lamivudine; TDF, tenofovir; NVP, nevirapine; AZT, zidovudine, LPV/r, lopinavir/ritonavir; FTC, emtricitabine.

Participant ID	Drug regimen*	Time ART naïve (weeks)	Time on ART (weeks)	CD4 count at ART initiation (cells per microliter)	CD4 count at QVOA blood draw (cells per microliter)	Number of OGVs	Percent clonal OGVs	Percent OGVs from within 1-year pre-ART	Percent OGVs from within 1 year of transmission
CAP188	EFV/3TC/TDF EFV/FTC/TDF	247	246	364	430	12	0	50	0
CAP206	EFV/3TC/TDF EFV/FTC/TDF	274	286	289	978	21	19	78	5
CAP217	EFV/3TC/TDF	360	239	287	657	17	12	44	0
CAP257	3TC/NVP/ TDF EFV/3TC/TDF EFV/FTC/TDF	249	318	170	596	48	17	89	0
CAP287	EFV/FTC/TDF	260	223	411	559	6	0	67	17
CAP288	EFV/FTC/TDF	210	275	318	860	7	0	100	0
CAP302	EFV/3TC/AZT EFV/FTC/TDF	164	270	307	541	8	38	17	17
CAP316	3TC/AZT/ LPV/r EFV/FTC/TDF	215	215	436	888	11	0	91	0
CAP336	EFV/3TC/TDF	142	258	164	558	17	47	100	0
<b>Average</b>		<b>236</b>	<b>259</b>	<b>305</b>	<b>674</b>	<b>16</b>	<b>15</b>	<b>71</b>	<b>4</b>
<b>Median</b>		<b>247</b>	<b>258</b>	<b>307</b>	<b>596</b>	<b>12</b>	<b>12</b>	<b>78</b>	<b>0</b>

\*Regimen changes indicated below original/prior regimen.



**Fig. 1. Viral load and suppression history of nine participants from the CAPRISA 002 acute infection cohort.** The graph shows the viral loads of HIV-1 RNA (copies per milliliter) in the blood before and after therapy, with the time of therapy initiation designated as  $T = 0$ . Each participant is designated by a different color, and the time of blood collection for assessing the reservoir is designated by vertical arrows. The dashed line represents the limit of detection for the viral load assay.



**Fig. 2. Most of the OGVs are most closely related to viruses replicating near the time of ART initiation.** Phylogenetic analyses of one genomic region for each of two participants are shown as representative of the pattern in which OGVs were derived from viruses replicating near the time of ART initiation. The analysis is for the c2 to c3 region of the viral *env* gene. Pre-ART sequences are colored red (within the first year after transmission) to blue (within the year before ART), and OGVs are shown in pink. Color coding to represent time is the same in Figs. 2 to 5. (A) CAP257 had 44 unique OGVs of which 89% were most closely related to viruses replicating within 1-year pre-ART. (B) CAP288 had seven unique OGVs, all (100%) of which were most closely related to viruses replicating within 1-year pre-ART. The timing of pre-ART samples are shown as negative values post-ART initiation.

with a Primer ID protocol that allowed us to deeply sample the replicating and evolving viral populations in plasma pre-ART while correcting for sequencing errors and template resampling (30, 31). Phylogenetic trees were constructed for each of the genomic regions that yielded sequences at most pre-ART time points. Only two regions were analyzed for one participant, whereas at least three regions were analyzed for the remaining eight participants. In addition, viral RNA was isolated from the culture supernatant from QVOA wells for analysis of OGV sequences from the replication-competent latent reservoir. Near full-length OGV sequences were generated using the Pacific Biosciences single molecule, real-time (SMRT) platform with barcoded primers that likewise allowed for virtually complete error correction; we estimate a residual sequencing error rate of 1 in 50,000 nucleotides sequenced on the basis of repeated sequencing of a defective virus produced from a cell clone [8E5 cells; (32)]. The sequence of each OGV was determined from a bulk polymerase chain reaction (PCR) product of the supernatant virus, given that the resting  $CD4^+$  T cells had been diluted to end point so that a single virus was present in the culture well in most cases. Within the different OGVs from a subject, an average of 15% of the sequences were clonal, defined as sequences that are identical or differ by one base (i.e., 1 difference of about 8850 nucleotides sequenced, likely due to rare sequencing errors and errors within the first round of viral replication in the well; Table 1). Because these OGVs were likely derived from cells that had undergone clonal proliferation in the host (8, 12), we assumed that clonal OGVs represent a single entry into the reservoir followed by subsequent cellular expansion in the body and therefore included only one representative sequence from each clonal population in summarizing our findings.

**Analysis of the timing of reservoir formation**

Phylogenetic analysis revealed that reservoir OGV sequences predominantly clustered in the tree with plasma viruses from the year preceding ART initiation [representative trees for two participants are shown in Fig. 2 (A and B)], although for some participants the OGVs were distributed throughout the tree [two examples are shown in Fig. 3 (A and B) and trees for all participants are shown in figs. S2 to S10]. Three methods, distance, clade support, and phylogenetic placement (33, 34), were used to estimate when each individual OGV entered the reservoir. The three methods were applied to each of the multiple regions of the genome sequenced, and the weighted median of these

Downloaded from <http://stm.sciencemag.org/> at UNIV OF CALIFORNIA - SAN DIEGO on January 15, 2020

estimates was used to summarize when each OGV entered the reservoir (Fig. 4 for the data shown in Figs. 2 and 3, and fig. S11).

The percentage of the OGVs representing viruses replicating in the last year before the initiation of therapy (i.e., “late” virus) was calculated for each of the nine participants (Fig. 5A and Table 1). The median of these values was 78%, and the average percentage was 71%. More than 90% of OGVs in four of nine women and at least 67% of OGVs in six of nine women entered the reservoir in the year preceding ART. In contrast, on average, only 4% of OGVs entered the reservoir within the first year of infection (Fig. 5A and Table 1).

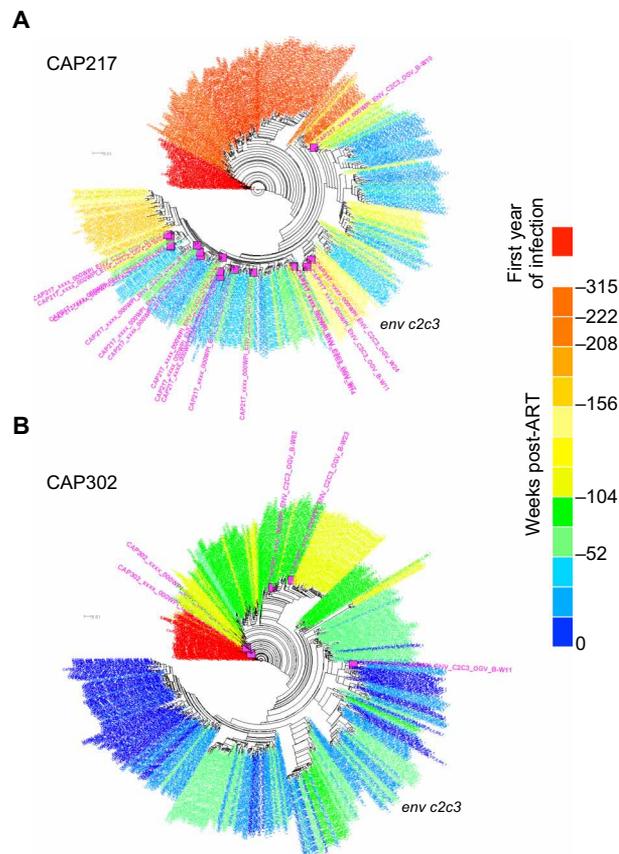
We estimate that 15% of OGVs were induced from clonally expanded cells (i.e., differed from another OGV by no more than one mutation). However, because we sampled a small number of OGVs from some participants, we were only able to identify an OGV as being from an expanded cell clone if that clone was present at a high frequency in the population. In this analysis of 25 clonal OGVs (25 total clones in 10 clone groups from five participants), 60% of the clone groups formed with a virus that was replicating in the year before ART, with the remaining 40% forming from viruses that had been replicating earlier. There was no statistical difference between the frequency of “late” viruses observed as clones in the OGVs versus their frequency in unique OGVs (Fisher’s exact test,  $P = 0.33$ ). Thus, there is no evidence (from this limited sampling) that clonal expansion differentially affected OGVs that entered the reservoir earlier in untreated infection compared to those that entered near the time of ART initiation.

### Modeling formation of the long-lived HIV-1 reservoir

Our analyses suggest that cells infected with HIV-1 near the time of ART initiation (or clones of those cells) are more likely to contribute to the long-lived, replication-competent reservoir than cells infected at earlier time points. We considered two simple models to explain our findings. The first model assumes that infected resting CD4<sup>+</sup> T cells, which have been observed in untreated infection (35), have a half-life of 44 months (176 weeks), comparable to their previously estimated half-life on therapy (Fig. 5B, left) (4, 5). Under this model, the reservoir is expected to contain variants sampled throughout untreated infection with those from later in infection being only slightly more common. The second model assumes that infected resting CD4<sup>+</sup> T cells have by contrast a short half-life in untreated HIV-1 infection (here selected to be 2 weeks) that then stabilizes to the longer 44-month half-life with the onset of effective ART (Fig. 5B, middle). Similar to our observations for two participants (as examples) with predominantly late reservoir virus (CAP257 and CAP288; Fig. 5B, right), this second model predicts that most of the reservoir is seeded by viruses replicating around the time of ART initiation. Thus, the observed high similarity of viruses replicating proximal to ART initiation to those in the latent reservoir, the result seen in most of the participants, can be explained by infected resting CD4<sup>+</sup> T cells from early in infection rapidly decaying before they contribute to the long-lived reservoir. However, the second model does not fit the data for the subset of participants (e.g., CAP217 and CAP302) who had virus enter the reservoir over the entire period before the initiation of therapy (Fig. 5B, right).

### DISCUSSION

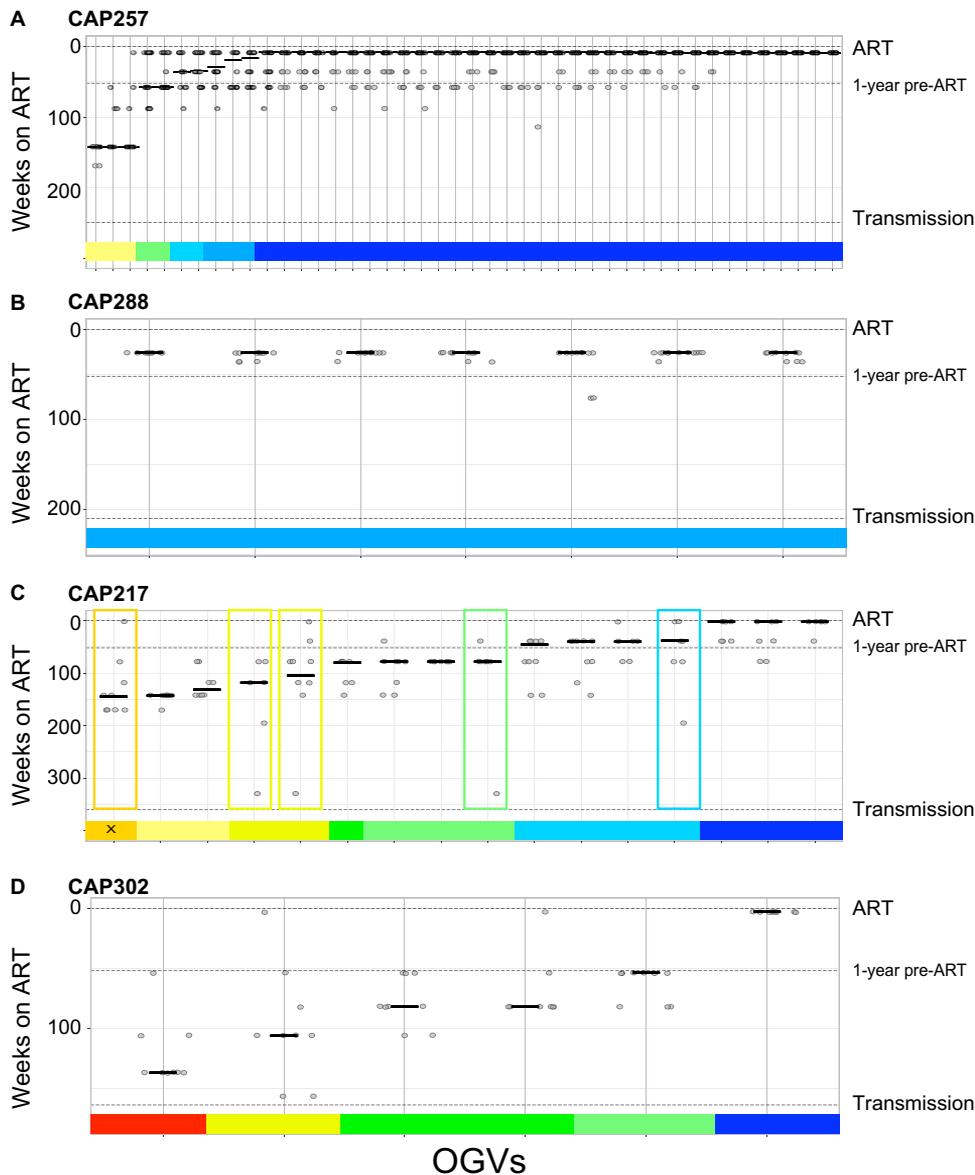
In this study, we have taken advantage of the availability of archived pre-ART viremic samples in the CAPRISA 002 cohort to establish a detailed temporal record of viral evolution in each of nine women



**Fig. 3. A subset of participants had OGVs most closely related to viruses replicating throughout untreated infection.** Phylogenetic analyses of one genomic region for each of two participants are shown. Selected participants each had OGVs with timing assignments from a wide range of times pre-ART, as illustrated by the clustering of OGV sequences (pink) with sequences from different pre-ART time points (red to blue). (A) CAP217 had 17 OGVs, two of which were clones. Of the genetically unique OGVs, 44% were most closely related to viruses replicating within 1-year pre-ART and 31% to viruses replicating more than 2 years before ART. (B) CAP302 had six unique OGVs of which one was from within the year before ART, and one was from within the first year after transmission.

enrolled from the time of primary HIV-1 infection. Using this temporal record of viral evolution, we estimated when viruses that had persisted after more than 4 years on suppressive therapy entered the long-lived latent reservoir. In most women, the majority of viruses entered the reservoir around the time of therapy initiation, suggesting that the initiation of ART indirectly changes the host environment by suppressing viral replication to favor the establishment of long-lived cells, some of which are latently infected with replication-competent HIV-1.

Brodin *et al.* (23) and Jones *et al.* (22) recently characterized the formation of the long-lived proviral DNA reservoir, which is predominantly defective (14, 15). Our study, which focused on the replication-competent component of the reservoir, is consistent with the observations of Brodin *et al.* (23), who found that the DNA reservoir in 10 individuals was skewed toward variants present in the year preceding ART, although to a lesser extent than identified here. The results of Brodin *et al.* and our study differ from those of Jones *et al.*, who inferred diversity in the viral DNA reservoir that was reflective of viral evolution over the entire course of infection in two individuals.



**Fig. 4. Timing of reservoir entry for OGVs as estimated by combining three methods.** Three methods (patristic distance, clade support, and phylogenetic placement) were used to generate estimates of OGV entry into the reservoir by comparing multiple regions of the OGV genome to that of viral RNA isolated from the plasma pre-ART. **(A and B)** Estimates are for two participants with OGVs that largely entered the reservoir near the time of ART initiation (CAP257 and CAP288) and **(C and D)** two with OGVs that entered the reservoir throughout untreated infection (CAP217 and CAP302). Each OGV is represented as a vertical line with up to 15 estimates (five regions and three methods) along that line. For each OGV, estimates of reservoir entry were summarized as a weighted median, shown as a horizontal line and represented in a colored bar along the x axis. Boxes indicate OGVs with highly variable estimates (SD of >1 year). Additional phylogenetic analyses of OGVs with variable estimates found that one (designated with an “x” at the bottom) was in a time-specific lineage that did not correspond to the weighted median. For this OGV, dating was revised to reflect the time-specific lineage where it appeared. All other OGVs with variable estimates were found in time-specific lineages that corresponded to the weighted median.

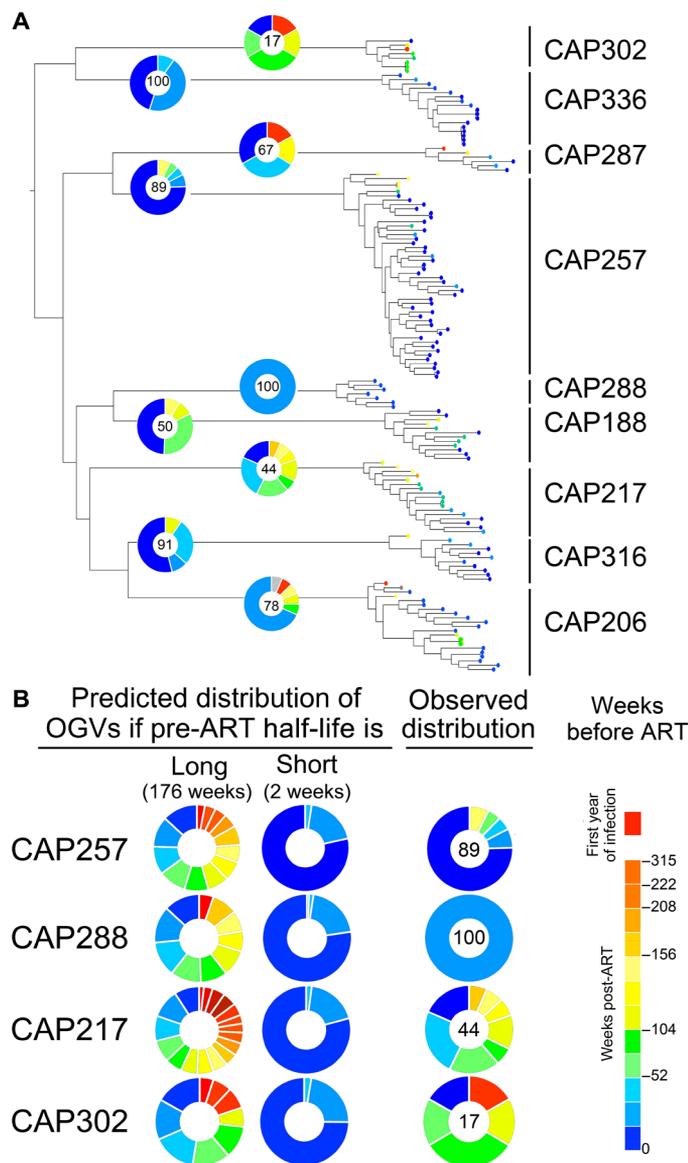
However, their analysis was based on sequences from a limited number of pre-ART time points (all during chronic infection) and thus required extrapolation based on a molecular clock model. In contrast, we were able to sequence virus from acute/early infection onward to the start of therapy and directly estimate that only a small fraction of the sampled OGV populations was seeded well before

the initiation of therapy. In seven of nine participants, there was no evidence of continuous reservoir formation. For one participant (CAP302), the replication-competent reservoir was seeded primarily with virus from early to mid-stages of the untreated infection, whereas another (CAP217) also appeared continuous but with a bias toward the last year before ART. Overall, our data and modeling analysis suggest that much of the long-lived HIV-1 reservoir forms after, or is stabilized during, the initiation of ART, which also coincides with rapid declines in HIV-1 viral load, increases in CD4<sup>+</sup> T cell counts, and reductions in markers of immune activation. One limitation is that in those subjects where the number of OGVs is small (Table 1), we are able to glean only a coarse picture of reservoir timing. However, even such a broad-stroke picture makes it clear that, on average, there is preferential entry into the reservoir by viruses replicating around the time of therapy initiation. Another limitation is that the observations were made in only nine women, all of whom had a relatively uniform response to suppressive ART.

It is not clear why only a small fraction of the full-length proviruses from resting CD4<sup>+</sup> T cells produce detectable virus in a QVOA (14). It is possible that most proviruses are induced in QVOA, but that the assay is very inefficient at amplifying the small amounts of virus produced by a single resting cell. Alternatively, proviruses may exist in different states of latency with only a small fraction of them inducible in cell culture. Consistent with these possibilities, Ho *et al.* (14) have shown that within a QVOA repeated rounds of stimulation can give rise to an additional 24% of induced viruses, either pointing to a stochastic nature of the assay or a distinct set of proviruses per cells that are inducible only with multiple rounds of stimulation.

If proviruses exist in different states of latency, then OGVs recovered by QVOA may represent a skewed sample of the viral reservoir, missing part of the reservoir that may have been seeded earlier in untreated infection. Although it is likely

that proviruses exist in different degrees of latency, perhaps representing different host mechanisms of gene silencing, the work of Brodin *et al.* (23) finding that the viral DNA reservoir is also skewed toward variants circulating late in untreated infection indicates that most of both the induced and uninduced reservoir is seeded near the time of ART initiation. Although we cannot rule out the possibility that



**Fig. 5. Skewing of reservoir virus to virus replicating near the time of ART initiation may be explained by a change in the half-life of latently infected cells in the presence of ART.** (A) All OGV sequences are shown in an approximate maximum-likelihood tree. Branch tips are colored according to the estimated time when each OGV entered the reservoir, with OGVs from the year after transmission in red and OGVs from the year before ART in blue. For each participant, the OGV timing distribution is illustrated in a pie chart with the percentage of OGVs from the year pre-ART listed. Overall, 71% of OGVs were produced by cells infected near the time of ART initiation. (B) Two models were considered to explain this pattern. One model assumes that latently infected CD4<sup>+</sup> T cells have a long half-life of 176 weeks (44 months) both before and during therapy and predicts that the reservoir will contain variants from throughout untreated infection. The other model assumes that latently infected CD4<sup>+</sup> T cells have a short half-life in untreated infection (here chosen as 2 weeks) that then stabilizes to a long half-life during therapy. Four participants are shown here to show the range in outcomes as seen in (A).

a small fraction of the reservoir is seeded early in untreated infection and persists in a state of “deep latency,” evidence from our study and Brodin *et al.* (23) indicates that most of the reservoir (both inducible and uninducible) is seeded near the time of ART initiation.

In this study, we focused on virus induced to grow from resting CD4<sup>+</sup> T cells isolated from the blood of virologically suppressed women. We have not examined the reservoir in activated cells (CD25<sup>High</sup>, CD69<sup>+</sup>, and HLA-DR<sup>+</sup>), cells in tissues, or cells in HIV-positive men. Also, several studies have found that viral DNA, including intact viral genomes and replication-competent genomes, are found in a wide variety of different CD4<sup>+</sup> T cell subsets (13, 36–38). Although there is a need to confirm our observations more broadly, they are likely generalizable. For example, Boritz *et al.* (39) observed that in elite controllers, HIV-infected cells migrate between lymphoid tissue and the blood, suggesting that viral populations at these sites are genetically linked. In addition, analysis of gut tissue biopsies and lymph node biopsies compared to cells from the blood suggest that viral populations in latently infected cells are largely equilibrated between blood and tissue (40, 41), making it less likely that there is a genetically distinct population of latent virus residing within tissue. As mentioned above, the results of Brodin *et al.* (23) using total peripheral blood mononuclear cells (PBMCs), which would have contained other cell types including activated CD4<sup>+</sup> T cells, showed that most viral DNA came from late replicating virus, again consistent with there not being a distinct reservoir consisting of viruses from earlier time points. In addition, whereas our study of biologically active, inducible virus in the reservoir was done using women, the analogous results of Brodin *et al.* (23) were obtained in a cohort largely composed of men, suggesting that sex is not a major determinant of when viral sequences enter the long-lived reservoir.

There is ongoing interest as to whether HIV-1 replicates during suppressive antiviral therapy. We were able to account for all OGVs as being phylogenetically close to, and at times identical to, viruses that were replicating before the initiation of therapy, i.e., the OGVs did not represent an outlier group of sequences that had undergone further sequence evolution due to ongoing replication during the average of 5 years on therapy. These results are consistent with those of Brodin *et al.* (23), McManus *et al.* (41), and Van Zyl *et al.* (42) who have failed to find any evidence for viral sequence evolution during suppressive ART, an observation most compatible with the latent reservoir not being sustained by ongoing viral replication.

Our results point to the long-lived reservoir forming under two very distinct conditions. Most of the reservoir (on average) forms around the time of therapy initiation and is likely driven by the rapidly changing immune environment within the host that occurs in response to lowering viral loads and decreased antigen drive. These immunologic changes include absolute increases in peripheral CD4<sup>+</sup> T cells, reductions in CD4<sup>+</sup> T cell turnover (43, 44), and increased frequencies of long-lived memory CD4<sup>+</sup> T cells (45). Initiation of ART may therefore promote the increased transition of both infected and uninfected CD4<sup>+</sup> T cells from being short-lived effectors to being long-lived memory T cells (46). Similarly, ART-mediated reduction of HIV-1 antigen restores the proliferative ability of HIV-specific CD4<sup>+</sup> T cells (47). These changes may directly affect HIV-specific CD4<sup>+</sup> T cells by allowing them to become long-lived memory cells capable of proliferation. As this model would predict, HIV-infected people on ART have been shown to have HIV-specific CD4<sup>+</sup> T cells that are infected at rates greater than cytomegalovirus-specific T cells or total memory CD4<sup>+</sup> T cells, although HIV-specific CD4<sup>+</sup> T cells represent only a small fraction of the total memory CD4<sup>+</sup> T cells and contain only a small fraction of the total viral DNA (48). Together, these findings suggest that profound immunologic changes at the time of ART may allow

HIV-infected cells to become long-lived memory cells and form most of the stable reservoir.

An additional mechanism is needed to explain how a variable fraction of the reservoir forms in the presence of active viral replication (see examples in Fig. 3). The fraction of OGVs that entered the reservoir at earlier time points before therapy initiation was highly variable, with most of the OGVs from two participants (CAP302 and CAP217) seeded earlier in untreated infection. It is not clear what factors contribute to early viral entry into the reservoir. For the more extreme case (CAP302), the infection course was only moderately different from the cohort overall: 3 years in study before therapy, set point viral load about two times higher than the mean, and nadir CD4<sup>+</sup> T cell count only slightly lower than the mean. An important question about the phenomenon of early entry into the reservoir is whether it is continuous or episodic where it might be linked to some other biological events that occur in the person, such as transient exposure to other infectious agents.

Untreated HIV-1 infection is characterized by increased immune activation and dysfunction, which are largely reversed after ART initiation. Most relevant to this study is the loss of CD127[interleukin-7 (IL-7) receptor]<sup>+</sup> memory CD4<sup>+</sup> T cells during untreated infection (49–51). IL-7 signaling through CD127 is essential for the transition of effector to memory CD4<sup>+</sup> T cells and the subsequent maintenance of memory cells (52). With ART initiation, the number of CD127<sup>+</sup> memory CD4<sup>+</sup> T cells increases (43, 49, 50), most likely due to reduced inflammation (53). These results suggest a model in which the increase in CD127 expression during ART allows latently infected cells to become long-lived as they now become responsive to IL-7 [fig. S12 (54)]. Consistent with this model in which much of the reservoir should be in cells sensitive to IL-7, administering IL-7 to HIV-infected individuals on ART increases the size of the HIV-1 reservoir (55), and the rate of reservoir decay is slower when IL-7 concentrations are higher (13).

Therapies that limit CD4<sup>+</sup> T cell memory generation, perhaps through dampening IL-7/IL-7R signaling in the window from the time of ART initiation to when all viral replication is fully suppressed, could limit reservoir formation (54). Such an approach would not eliminate all of the reservoir, as it is clear that latency can be seeded at different points during untreated infection. However, blocking the formation of most of the reservoir at the time of therapy initiation would greatly reduce the size of the reservoir, lowering the barrier to eradicating HIV-1 from an infected person.

## MATERIALS AND METHODS

### Study design

Viral RNA from pre-ART time points ranging from acute/early HIV-1 infection to the time of ART initiation was sequenced across several regions of the viral genome using a deep sequencing protocol. After the women had been on ART for several years, they provided a large blood sample that was used to grow out latent HIV-1 from the long-lived viral reservoir. The OGV sequences were compared to the sequences from pre-ART longitudinal samples of replicating virus to determine when the OGV entered the long-lived reservoir before the initiation of ART.

### Study participants

The CAPRISA 002 cohort is composed of women from rural and urban KwaZulu-Natal, South Africa, who were identified during the

period of acute/primary infection with subtype C HIV-1 and followed longitudinally (24). Participants in the cohort gave blood samples every 3 to 6 months and were provided ART on the basis of the prevailing in-country guidelines. The women were retained in the cohort for up to 5 years after therapy initiation. We identified nine women in the cohort who had been ART naïve for at least 3 years before initiating therapy and who had been on ART for a minimum of 4 years. The women provided a 200 ml blood draw while on therapy from which PBMCs were purified using a Ficoll gradient and cryopreserved in liquid nitrogen. Women were excluded on the basis of pregnancy or if they had hemoglobin concentrations lower than 10 g/dl. The use of stored samples and the collection and analysis of new samples were done under a protocol approved by the Biomedical Research Ethics Committee of the University of KwaZulu Natal (BE178/150) and the Human Research Ethics Committee of the University of Cape Town (588/2015) in South Africa and at the University of North Carolina at Chapel Hill in the United States.

### Quantitative viral outgrowth assay

Resting CD4<sup>+</sup> T cells were isolated from PBMCs by negative selection with CD25 depletion to isolate cells that were CD69<sup>-</sup>, HLA-DR<sup>-</sup>, and CD25<sup>low</sup> (Custom Kit, STEMCELL). Purified cells were then examined by flow cytometry (LSRII Fortessa) for expression of CD69 (phycoerythrin, BD), CD25 (anaphase-promoting complex, BD), CD8 (fluorescein isothiocyanate, BD), CD4 (peridinin chlorophyll protein complex-Cy5.5, BD), and viability (Aqua LIVE/DEAD, Thermo Fisher Scientific) to confirm purity.

Resting cells were cultured at 100,000 cells per well with highly purified phytohemagglutinin (PHA) (1.5 µg/ml; Remel-PHA, Thermo Fisher Scientific), IL-2 (60 U/ml), and irradiated PBMCs from a seronegative donor, and QVOA was performed as previously described (4). On days 15 and 19, cultures were tested for HIV-1 p24 capsid protein production using an enzyme-linked immunosorbent assay. Culture supernatants positive for p24 were stored at -80°C.

### Sequencing of viral populations in blood plasma pre-ART

Viral RNA copies were extracted from blood plasma using the QIAamp Viral RNA Mini Kit (Qiagen). The purified RNA was reverse transcribed to complementary DNA (cDNA) using Super-Script III/IV Reverse Transcriptase (Invitrogen). We used the Primer ID method (30, 31), which tags each RNA template through its cDNA primer with a unique 12-nucleotide-long identifier. This allows amplified sequences to be grouped according to their Primer ID tags, from which a consensus sequence can then be derived for each individual template (template consensus sequence). Multiple cDNA primers were used during the cDNA synthesis step for each sample to generate cDNAs corresponding to multiple regions of the genome and allow multiplexing of the PCR amplicons followed by sequencing. The gene-specific sequence of the cDNA primer was used as an index for the specific region of the genome. Two separate cDNA reactions were performed for each sample to allow efficient use of the available RNA to sequence the different regions of the viral genome. The cDNA products were purified twice using Agencourt RNAClean XP magnetic beads (Beckman Coulter). Multiplexing during PCR was accomplished by including a common PCR primer binding sequence at the 5' end of each cDNA primer and a gene-specific forward primer for each amplicon/region. PCR amplification was performed using the KAPA2G Fast Multiplex Mix (Kapa Biosystems) with an equal molar amount of each forward primer

and an excess of the universal reverse primer. A second PCR step using the Expand High Fidelity PCR System (Roche) allowed the incorporation of the Illumina MiSeq version 2 indexes and adapters. PCR products were purified using SPRIselect beads (Beckman Coulter), with a 0.6:1 volume ratio of beads to PCR product (to allow for size exclusion of residual primers) following both PCR steps. The final purified amplicon libraries were quantified using the Qubit double-stranded DNA high sensitivity (HS) assay pooled at equimolar ratios, and purified again using SPRIselect beads. Illumina MiSeq 2 × 300–base paired-end sequencing was performed on the multiplexed amplicons.

### Sequencing of OGV populations

Viral RNA was isolated from the QVOA culture supernatant of p24-positive wells and converted to cDNA using SuperScript III Reverse Transcriptase and an oligo(dT) primer. The 5′ and 3′ half genomes were amplified in separate PCR reactions using barcoded primers, and the PCR products were gel purified. For CAP257, about 35% of the OGVs had only a 3′ half genome amplicon available for analysis. The SMRTbell Template Prep Kit (PacBio) was used to add adaptors to pooled, barcoded amplicons, and the pools of amplicons were then submitted for PacBio sequencing (movie time of 10 hours for data collection). Sequences were grouped by barcode, and high quality sequences were analyzed using the PacBio Long Amplicon Analysis package. The 3′ and 5′ amplicons for the same virus were joined and visually screened to confirm that open reading frames were intact. The near full-length genome sequences are deposited in GenBank (accession nos. MN09755 to MN097697). The MiSeq sequences are deposited in the Sequencing Read Archive (accession nos. SAMN12126374 to SAMN12126459).

### Illumina MiSeq data processing

Raw reads were processed using a custom pipeline written in Python and R programming languages. The MotifBinner2.R program (<https://github.com/HIVDiversity/MotifBinner2>; DOI: 10.5281/zenodo.3372204) carries out quality filtering of sequence data, merging of overlapping paired-end reads and implements the Primer ID processing method described by Zhou *et al.* (31). The resulting Primer ID template consensus sequences were processed with an in-house pipeline ([https://github.com/HIVDiversity/NGS\\_processing\\_pipeline](https://github.com/HIVDiversity/NGS_processing_pipeline); DOI: 10.5281/zenodo.3372202), which removes sequences with degenerate bases or deletions >50 nucleotides long and filters out any contaminant nontarget gene sequences.

### Phylogenetic and statistical analyses

Hypervariable loop regions of *env* were manually removed, where accurate alignment was not possible. In-frame codon alignments were generated for each genomic region using a custom pipeline (<https://github.com/veg/ogv-dating>; DOI: 10.5281/zenodo.3370141). Briefly, translated protein sequences were used to identify RNA sequences that were in-frame. Sequences that contained frameshifts were aligned to each in-frame sequence using the Smith-Waterman algorithm implemented in the BioExt package (<https://github.com/veg/BioExt>), and the pairs with the highest score were used to correct frameshifts. Multiple sequence comparison by log-expectation (MUSCLE) (56) was then used to align in-frame amino acid sequences and then generate in-frame codon-aligned RNA sequences lacking frameshifts. For each pre-ART time point analyzed, identical RNA sequences within that time point were replaced with a single

representative, and approximately-maximum-likelihood trees were constructed using FastTree2 with more rigorous search options (57).

Three methods were used to analyze sequence data from each genomic region and estimate when the OGV entered the reservoir (pipeline and additional description of the methods are available at <https://github.com/veg/ogv-dating>; DOI: 10.5281/zenodo.3370141). Initially, the tree was rooted to maximize root-to-tip distance versus its time linear regression coefficient. The first method was “patristic distance,” which computes the path lengths between an OGV sequence and every other pre-ART sequence in the tree, finds the minimum distance (*d*), and assigns the OGV sequence to the time point from which the simple majority of pre-ART sequences within distance 2*d* come from. The second method was “clade support,” which starts at the OGV sequence and traverses the tree toward the root until it encounters a well-supported (bootstrap ≥ 90%) internal node that has at least one pre-ART sequence. Each OGV sequence was assigned to the time point from which most of the pre-ART sequences in the well-supported node came. The third method was phylogenetic placement (33, 34), which was used to analyze phylogenetic trees containing only pre-ART sequences and identify the position in these reference trees where each OGV was most likely to be found and use that new position to identify when the OGV entered the reservoir. Patristic distance and clade support estimates were excluded from subsequent analyses if their tree support value was less than 0.4. Tree support values were estimated by applying the same dating procedure to each pre-ART sequence and seeing how often the correct sampling time point was recovered. A support of 0.4 means that the correct time point was recovered for 40% of pre-ART sequences. Because generating tree support estimates for phylogenetic placement is computationally slow, we did not exclude any estimates generated by phylogenetic placement. Figure 4 and fig. S11 illustrate all of the estimates analyzed for each OGV.

For each OGV, timing of reservoir entry was summarized as the weighted median of estimates generated by the three different methods for each genomic region analyzed and the support value for each estimate. As a way of validating this approach, we identified OGVs whose individual estimates had an SD of greater than 1 year. We then visually inspected their phylogenetic trees to identify whether the weighted median accurately characterized the position of each OGV in the phylogenetic trees. A total of 25 OGVs (25 of 132 unique OGVs) were identified as having an SD of greater than 1 year. We then inspected the phylogenetic trees to identify whether each OGV was found in a time-specific lineage and whether that lineage dated to the same time as the weighted median. This approach revealed that the weighted median accurately characterized 18 of 25 OGVs examined. However, six OGVs could be found in time-specific lineages that did not correspond to the weighted median. We therefore revised the estimate of timing of reservoir entry of these six OGVs to reflect the age of the time-specific lineage in which they were found. One additional OGV was not located in a time-specific lineage, thus we revised its entry into the reservoir to be “ambiguous.” Most of the seven OGVs that were mischaracterized by our automated methods had a nearby group of identical clonal sequences observed across multiple pre-ART time points, which may have been problematic for the methods. Furthermore, six of these seven OGVs were from a single participant (CAP206). Further examination of sequence data for this participant found no evidence of contamination or poor sequence

alignment that could explain why our methods failed for this participant.

## SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/11/513/eaaw5589/DC1

Fig. S1. Longitudinal MiSeq sampling depth for nine women from the CAPRISA 002 cohort.

Fig. S2. Timing of reservoir OGVs for participant CAP188.

Fig. S3. Timing of reservoir OGVs for participant CAP206.

Fig. S4. Timing of reservoir OGVs for participant CAP217.

Fig. S5. Timing of reservoir OGVs for participant CAP257.

Fig. S6. Timing of reservoir OGVs for participant CAP287.

Fig. S7. Timing of reservoir OGVs for participant CAP288.

Fig. S8. Timing of reservoir OGVs for participant CAP302.

Fig. S9. Timing of reservoir OGVs for participant CAP316.

Fig. S10. Timing of reservoir OGVs for participant CAP336.

Fig. S11. Estimates of OGV entry into the long-lived reservoir.

Fig. S12. Model of the relationship between viral suppression and formation of the long-lived HIV-1 reservoir.

Table S1. Data of participant viral loads.

Table S2. Data for timing of OGVs.

[View/request a protocol for this paper from Bio-protocol.](#)

## REFERENCES AND NOTES

1. T.-W. Chun, L. Stuyver, S. B. Mizell, L. A. Ehler, J. A. M. Mican, M. Baseler, A. L. Lloyd, M. A. Nowak, A. S. Fauci, Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13193–13197 (1997).
2. D. Finzi, M. Hermankova, T. Pierson, L. M. Carruth, C. Buck, R. E. Chaisson, T. C. Quinn, K. Chadwick, J. Margolick, R. Brookmeyer, J. Gallant, M. Markowitz, D. D. Ho, D. D. Richman, R. F. Siliciano, Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
3. J. K. Wong, M. Hezareh, H. F. Günthard, D. V. Havlir, C. C. Ignacio, C. A. Spina, D. D. Richman, Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295 (1997).
4. A. M. Crooks, R. Bateson, A. B. Cope, N. P. Dahl, M. K. Griggs, J. D. Kuruc, C. L. Gay, J. J. Eron, D. M. Margolis, R. J. Bosch, N. M. Archin, Precise quantitation of the latent HIV-1 reservoir: Implications for eradication strategies. *J. Infect. Dis.* **212**, 1361–1365 (2015).
5. J. D. Siliciano, J. Kajdas, D. Finzi, T. C. Quinn, K. Chadwick, J. B. Margolick, C. Kovacs, S. J. Gange, R. F. Siliciano, Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4<sup>+</sup> T cells. *Nat. Med.* **9**, 727–728 (2003).
6. Z. Wang, E. E. Gurule, T. P. Brennan, J. M. Gerold, K. J. Kwon, N. N. Hosmane, M. R. Kumar, S. A. Beg, A. A. Capoferri, S. C. Ray, Y.-C. Ho, A. L. Hill, J. D. Siliciano, R. F. Siliciano, Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2575–E2584 (2018).
7. N. N. Hosmane, K. J. Kwon, K. M. Bruner, A. A. Capoferri, S. Beg, D. I. Rosenbloom, B. F. Keele, Y. C. Ho, J. D. Siliciano, R. F. Siliciano, Proliferation of latently infected CD4<sup>+</sup> T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J. Exp. Med.* **214**, 959–972 (2017).
8. T. A. Wagner, S. McLaughlin, K. Garg, C. Y. K. Cheung, B. B. Larsen, S. Styrchak, H. C. Huang, P. T. Edlefsen, J. I. Mullins, L. M. Frenkel, Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).
9. J. K. Bui, M. D. Sobolewski, B. F. Keele, J. Spindler, A. Musick, A. Wiegand, B. T. Luke, W. Shao, S. H. Hughes, J. M. Coffin, M. F. Kearney, J. W. Mellors, Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog.* **13**, e1006283 (2017).
10. M. R. Pinzone, D. J. VanBelzen, S. Weissman, M. P. Bertuccio, L. Cannon, E. Venanzi-Rullo, S. Migueles, R. B. Jones, T. Mota, S. B. Joseph, K. Groen, A. O. Pasternak, W.-T. Hwang, B. Sherman, A. Vourekas, G. Nunnari, U. O'Doherty, Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat. Commun.* **10**, 728 (2019).
11. F. R. Simonetti, M. D. Sobolewski, E. Fyne, W. Shao, J. Spindler, J. Hattori, E. M. Anderson, S. A. Watters, S. Hill, X. Wu, D. Wells, L. Su, B. T. Luke, E. K. Halvas, G. Besson, K. J. Penrose, Z. Yang, R. W. Kwan, C. Van Waes, T. Uldrick, D. E. Citrin, J. Kovacs, M. A. Polis, C. A. Rehm, R. Gorelick, M. Piatak, B. F. Keele, M. F. Kearney, J. M. Coffin, S. H. Hughes, J. W. Mellors, F. Maldarelli, Clonally expanded CD4<sup>+</sup> T cells can produce infectious HIV-1 in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1883–1888 (2016).
12. F. Maldarelli, X. Wu, L. Su, F. R. Simonetti, W. Shao, S. Hill, J. Spindler, A. L. Ferris, J. W. Mellors, M. F. Kearney, J. M. Coffin, S. H. Hughes, Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
13. N. Chomont, M. El-Far, P. Ancuta, L. Trautmann, F. A. Procopio, B. Yassine-Diab, G. Boucher, M. R. Boulassel, G. Ghattas, J. M. Brenchley, T. W. Schacker, B. J. Hill, D. C. Douek, J.-P. Routy, E. K. Haddad, R. P. Sékaly, HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).
14. Y.-C. Ho, L. Shan, N. N. Hosmane, J. Wang, S. B. Laskey, D. I. S. Rosenbloom, J. Lai, J. N. Blankson, J. D. Siliciano, R. F. Siliciano, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
15. K. M. Bruner, A. J. Murray, R. A. Pollack, M. G. Soliman, S. B. Laskey, A. A. Capoferri, J. Lai, M. C. Strain, S. M. Lada, R. Hoh, Y.-C. Ho, D. D. Richman, S. G. Deeks, J. D. Siliciano, R. F. Siliciano, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat. Med.* **22**, 1043–1049 (2016).
16. H. Imamichi, R. L. Dewar, J. W. Adelsberger, C. A. Rehm, U. O'Doherty, E. E. Paxinos, A. S. Fauci, H. C. Lane, 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.* **113**, 8783–8788 (2016).
17. R. A. Pollack, R. B. Jones, M. Perete, K. M. Bruner, A. R. Martin, A. S. Thomas, A. A. Capoferri, S. A. Beg, S.-H. Huang, S. Karandish, H. P. Hao, E. Halper-Stromberg, P. C. Yong, C. Kovacs, E. Benko, R. F. Siliciano, Y.-C. Ho, 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.e4 (2017).
18. L. Shan, K. Deng, H. B. Gao, S. F. Xing, A. A. Capoferri, C. M. Durand, S. A. Rabi, G. M. Laird, M. Kim, N. N. Hosmane, H. C. Yang, H. Zhang, J. B. Margolick, L. H. Li, W. P. Cai, R. A. Ke, R. A. Flavell, J. D. Siliciano, R. F. Siliciano, Transcriptional reprogramming during effector-to-memory transition renders CD4<sup>+</sup> T cells permissive for latent HIV-1 infection. *Immunity* **47**, 766–775.e3 (2017).
19. D. J. Colby, L. Trautmann, S. Pinyakorn, L. Leyre, A. Pagliuzza, E. Kroon, M. Rolland, H. Takata, S. Buranapraditkun, J. Intasan, N. Chomchey, R. Muir, E. K. Haddad, S. Tovanabutra, S. Ubolyam, D. L. Bolton, B. A. Fullmer, R. J. Gorelick, L. Fox, T. A. Crowell, R. Trichavaroj, R. O'Connell, N. Chomont, J. H. Kim, N. L. Michael, M. L. Robb, N. Phanuphak, J. Ananworanich, RV411 Study Group, Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nat. Med.* **24**, 923–926 (2018).
20. T. J. Henrich, H. Hatano, O. Bacon, L. E. Hogan, R. Rutishauser, A. Hill, M. F. Kearney, E. M. Anderson, S. P. Buchbinder, S. E. Cohen, M. Abdel-Mohsen, C. W. Pohlmeier, R. Fromentin, R. Hoh, A. Y. Liu, J. M. McCune, J. Spindler, K. Metcalf-Pate, K. S. Hobbs, C. Thanh, E. A. Gibson, D. R. Kuritzkes, R. F. Siliciano, R. W. Price, D. D. Richman, N. Chomont, J. D. Siliciano, J. W. Mellors, S. A. Yuki, J. N. Blankson, T. Liegler, S. G. Deeks, HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: An observational study. *PLoS Med.* **14**, e1002417 (2017).
21. J. B. Whitney, A. L. Hill, S. Sanisetty, P. Penalzo-MacMaster, J. Y. Liu, M. Shetty, L. Parenteau, C. Cabral, J. Shields, S. Blackmore, J. Y. Smith, A. L. Brinkman, L. E. Peter, S. I. Mathew, K. M. Smith, E. N. Bordini, D. I. S. Rosenbloom, M. G. Lewis, J. Hattersley, B. Li, J. Hesselgesser, R. Gelezunias, M. L. Robb, J. H. Kim, N. L. Michael, D. H. Barouch, Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–77 (2014).
22. B. R. Jones, N. N. Kinloch, J. Horacek, B. Ganase, M. Harris, P. R. Harrigan, R. B. Jones, M. A. Brockman, J. B. Joy, A. F. Y. Poon, Z. L. Brumme, Phylogenetic approach to recover integration dates of latent HIV sequences. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8958–E8967 (2018).
23. J. Brodin, F. Zanini, L. Thebo, C. Lanz, G. Bratt, R. A. Neher, J. Albert, Establishment and stability of the latent HIV-1 DNA reservoir. *eLife* **5**, e18889 (2016).
24. F. van Loggerenberg, K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. A. Karim, A. Grobler, N. Barnabas, I. Iriogbe, S. S. A. Karim; CAPRISA 002 Acute Infection Study Team, Establishing a cohort at high risk of HIV infection in South Africa: Challenges and experiences of the CAPRISA 002 Acute Infection Study. *PLoS ONE* **3**, e1954 (2008).
25. M.-R. Abrahams, J. A. Anderson, E. E. Giorgi, C. Seoighe, K. Mlisana, L.-H. Ping, G. S. Athreya, F. K. Treurnicht, B. F. Keele, N. Wood, J. F. Salazar-Gonzalez, T. Bhattacharya, H. Chu, I. Hoffman, S. Galvin, C. Mapanje, P. Kazembe, R. Thebus, S. Fiscus, W. Hide, M. S. Cohen, S. A. Karim, B. F. Haynes, G. M. Shaw, B. H. Hahn, B. T. Korber, R. Swanstrom, C. Williamson; CAPRISA Acute Infection Study Team; Center for HIV-AIDS Vaccine Immunology Consortium, Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* **83**, 3556–3567 (2009).
26. B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. X. Sun, T. Grayson, S. Y. Wang, H. Li, X. P. Wei, C. L. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L.-H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, G. M. Shaw, Identification and characterisation of transmitted

- and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7552–7557 (2008).
27. R. Shankarappa, J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X.-L. Huang, J. I. Mullins, Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* **73**, 10489–10502 (1999).
  28. Z. L. Brumme, M. John, J. M. Carlson, C. J. Brumme, D. Chan, M. A. Brockman, L. C. Swenson, I. Tao, S. Szeto, P. Rosato, J. Sela, C. M. Kadie, N. Frahm, C. Brander, D. W. Haas, S. A. Riddler, R. Haubrich, B. D. Walker, P. R. Harrigan, D. Heckerman, S. Mallal, HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS ONE* **4**, e6687 (2009).
  29. S. D. W. Frost, T. Wrin, D. M. Smith, S. L. K. Pond, Y. Liu, E. Paxinos, C. Chappey, J. Galovich, J. Beauchaine, C. J. Petropoulos, S. J. Little, D. D. Richman, Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18514–18519 (2005).
  30. C. B. Jabara, C. D. Jones, J. Roach, J. A. Anderson, R. Swanstrom, Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20166–20171 (2011).
  31. S. Zhou, C. Jones, P. Mieczkowski, R. Swanstrom, Primer ID validates template sampling depth and greatly reduces the error rate of next-generation sequencing of HIV-1 genomic RNA populations. *J. Virol.* **89**, 8540–8555 (2015).
  32. T. M. Folks, D. Powell, M. Lightfoote, S. Koenig, A. S. Fauci, S. B. Bunn, A. Rabson, D. Daugherty, H. E. Gendelman, M. D. Hoggan, Biological and biochemical characterization of a cloned Leu-3-cell surviving infection with the acquired immune deficiency syndrome retrovirus. *J. Exp. Med.* **164**, 280–290 (1986).
  33. S. D. W. Frost, B. Murrell, A. S. M. M. Hossain, G. J. Silverman, S. L. K. Pond, Assigning and visualizing germline genes in antibody repertoires. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140240 (2015).
  34. F. A. Matsen, R. B. Kodner, E. V. Armbrust, pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).
  35. T.-W. Chun, L. Carruth, D. Finzi, X. F. Shen, J. A. DiGiuseppe, H. Taylor, M. Hermankova, K. Chadwick, J. Margolick, T. C. Quinn, Y.-H. Kuo, R. Brookmeyer, M. A. Zeiger, P. Barditch-Crovo, R. F. Siliciano, Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
  36. N. Soriano-Sarabia, R. E. Bateson, N. P. Dahl, A. M. Crooks, J. D. Kuruc, D. M. Margolis, N. M. Archin, Quantitation of replication-competent HIV-1 in populations of resting CD4<sup>+</sup> T cells. *J. Virol.* **88**, 14070–14077 (2014).
  37. M. J. Buzon, E. Martin-Gayo, F. Pereyra, Z. Y. Ouyang, H. Sun, J. Z. Li, M. Piovoso, A. Shaw, J. Dalmau, N. Zangger, J. Martinez-Picado, R. Zurakowski, X. G. Yu, A. Telenti, B. D. Walker, E. S. Rosenberg, M. Lichterfeld, Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. *J. Virol.* **88**, 10056–10065 (2014).
  38. B. Hiener, B. A. Horsburgh, J.-S. Eden, K. Barton, T. E. Schlub, E. Lee, S. von Stockenström, L. Odevall, J. M. Milush, T. Liegler, E. Sinclair, R. Hoh, E. A. Boritz, D. Douek, R. Fromentin, N. Chomont, S. G. Deeks, F. M. Hecht, S. Palmer, Identification of genetically intact HIV-1 proviruses in specific CD4<sup>+</sup> T cells from effectively treated participants. *Cell Rep.* **21**, 813–822 (2017).
  39. E. A. Boritz, S. Darko, L. Swaszek, G. Wolf, D. Wells, X. L. Wu, A. R. Henry, F. Laboune, J. F. Hu, D. Ambrozak, M. S. Hughes, R. Hoh, J. P. Casazza, A. Vostal, D. Bunis, K. Nganou-Makamdop, J. S. Lee, S. A. Migueles, R. A. Koup, M. Connors, S. Moir, T. Schacker, F. Maldarelli, S. H. Hughes, S. G. Deeks, D. C. Douek, Multiple origins of virus persistence during natural control of HIV infection. *Cell* **166**, 1004–1015 (2016).
  40. T. H. Evering, S. Mehandru, P. Racz, K. Tenner-Racz, M. A. Poles, A. Figueroa, H. Mohri, M. Markowitz, Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiretroviral therapy initiated during primary infection. *PLoS Pathog.* **8**, e1002506 (2012).
  41. W. R. McManus, M. J. Bale, J. Spindler, A. Wiegand, A. Musick, S. C. Patro, M. D. Sobolewski, V. K. Musick, E. M. Anderson, J. C. Cyktor, E. K. Halvas, W. Shao, D. Wells, X. Wu, B. F. Keele, J. M. Milush, R. Hoh, J. W. Mellors, S. H. Hughes, S. G. Deeks, J. M. Coffin, M. F. Kearney, HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *J. Clin. Invest.* **130**, 126714 (2019).
  42. G. U. Van Zyl, M. G. Katusiime, A. Wiegand, W. R. McManus, M. J. Bale, E. K. Halvas, B. Luke, V. F. Boltz, J. Spindler, B. Laughton, S. Engelbrecht, J. M. Coffin, M. F. Cotton, W. Shao, J. W. Mellors, M. F. Kearney, No evidence of HIV replication in children on antiretroviral therapy. *J. Clin. Invest.* **127**, 3827–3834 (2017).
  43. M. K. Hellerstein, R. A. Hoh, M. B. Hanley, D. Cesar, D. Lee, R. A. Neese, J. M. McCune, Subpopulations of long-lived and short-lived T cells in advanced HIV-1 infection. *J. Clin. Invest.* **112**, 956–966 (2003).
  44. H. Mohri, A. S. Perelson, K. Tung, R. M. Ribeiro, B. Ramratnam, M. Markowitz, R. Kost, A. Hurley, L. Weinberger, D. Cesar, M. K. Hellerstein, D. D. Ho, Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. *J. Exp. Med.* **194**, 1277–1287 (2001).
  45. W. Xu, J. Li, Y. Wu, J. K. Zhou, J. Zhong, Q. Lv, H. Shao, H. Rao, CD127 expression in naive and memory T cells in HIV patients who have undergone long-term HAART. *Lab. Med.* **48**, 57–64 (2017).
  46. T. J. Scriba, M. Tameris, E. Smit, L. van der Merwe, E. J. Hughes, B. Kadira, K. Mauff, S. Moyo, N. Brittain, A. Lawrie, H. Mulenga, M. de Kock, L. Makhethe, E. J. van Rensburg, S. Gelderbloem, A. Veldsman, M. Hatherill, H. Geldenhuys, A. V. S. Hill, A. Hawkridge, G. D. Hussey, W. A. Hanekom, H. McShane, H. Mahomed, A phase IIa trial of the new tuberculosis vaccine, MVA85A, in HIV- and/or *Mycobacterium tuberculosis*-infected adults. *Am. J. Respir. Crit. Care Med.* **185**, 769–778 (2012).
  47. A. C. McNeil, W. L. Shupert, C. A. Iyasere, C. W. Hallahan, J. Mican, R. T. Davey Jr., M. Connors, High-level HIV-1 viremia suppresses viral antigen-specific CD4<sup>+</sup> T cell proliferation. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13878–13883 (2001).
  48. D. C. Douek, J. M. Brenchley, M. R. Betts, D. R. Ambrozak, B. J. Hill, Y. Okamoto, J. P. Casazza, J. Kuruppu, K. Kunstman, S. Wolinsky, Z. Grossman, M. Dybul, A. Oxenius, D. A. Price, M. Connors, R. A. Koup, HIV preferentially infects HIV-specific CD4<sup>+</sup> T cells. *Nature* **417**, 95–98 (2002).
  49. M. Paiardini, B. Cervasi, H. Albrecht, A. Muthukumar, R. Dunham, S. Gordon, H. Radziejewicz, G. Piedimonte, M. Magnani, M. Montroni, S. M. Kaech, A. Weintrob, J. D. Altman, D. L. Sodora, M. B. Feinberg, G. Silvestri, Loss of CD127 expression defines an expansion of effector CD8<sup>+</sup> T cells in HIV-infected individuals. *J. Immunol.* **174**, 2900–2909 (2005).
  50. S. A. Koesters, J. B. Alimonti, C. Wachihi, L. Matu, O. Anzala, J. Kimani, J. E. Embree, F. A. Plummer, K. R. Fowke, IL-7R $\alpha$  expression on CD4<sup>+</sup> T lymphocytes decreases with HIV disease progression and inversely correlates with immune activation. *Eur. J. Immunol.* **36**, 336–344 (2006).
  51. R. M. Dunham, B. Cervasi, J. M. Brenchley, H. Albrecht, A. Weintrob, B. Sumpter, J. Engram, S. Gordon, N. R. Klatt, I. Frank, D. L. Sodora, D. C. Douek, M. Paiardini, G. Silvestri, CD127 and CD25 expression defines CD4<sup>+</sup> T cell subsets that are differentially depleted during HIV infection. *J. Immunol.* **180**, 5582–5592 (2008).
  52. R. M. Kondrack, J. Harbertson, J. T. Tan, M. E. McBreen, C. D. Surh, L. M. Bradley, Interleukin 7 regulates the survival and generation of memory CD4 cells. *J. Exp. Med.* **198**, 1797–1806 (2003).
  53. S. A. K. Kiazky, K. R. Fowke, Loss of CD127 expression links immune activation and CD4<sup>+</sup> T cell loss in HIV infection. *Trends Microbiol.* **16**, 567–573 (2008).
  54. N. Goonetilleke, G. Clutton, R. Swanstrom, S. B. Joseph, Blocking formation of the stable HIV reservoir: A new perspective for HIV-1 cure. *Front. Immunol.* **10**, 1966 (2019).
  55. C. Vanderveeten, R. Fromentin, S. DaFonseca, M. B. Lawani, I. Sereti, M. M. Lederman, M. Ramgopal, J.-P. Routy, R. P. Sekaly, N. Chomont, Interleukin-7 promotes HIV persistence during antiretroviral therapy. *Blood* **121**, 4321–4329 (2013).
  56. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
  57. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).

**Acknowledgments:** We would like to acknowledge all participants of the CAPRISA 002 acute infection cohort and the staff at the Vulindlela and eThekweni Clinical Research Sites, KwaZulu-Natal, South Africa. We would also like to thank J. Ambler and the Bioinformatics Support Team at UCT for assistance with pipeline development. **Funding:** This work was supported by the National Institutes of Health (NIH)—South African Medical Research Council (MRC) U.S.-South Africa Program for Collaborative Biomedical Research grants R01 AI115981 to C.W. and R.S. and an NIH award to the Collaboratory of AIDS Researchers for Eradication (UM1 AI126619). The CAPRISA 002 acute infection cohort study has been funded by the South African Department of Science and Technology and the National Research Foundation's Centre of Excellence in HIV Prevention (grant no. UID: 96354), the South African Department of Health and the South African Medical Research Council Special Initiative on HIV Prevention (grant no. 96151), the NIH (U19 AI51794), USAID and CONRAD (USAID cooperative grant no. GP00-08-00005-00, subproject agreement no. PPA-09-046), the South African National Research Foundation (grant nos. 67385 and 96354), the South African Technology Innovation Agency, and the Fogarty International Center, NIH (D43 TW00231). The Centre for Infectious Diseases Research in Africa (CIDRI-Africa) is supported by core funding from the Wellcome Trust (203135/Z/16/Z). The work was also supported by the UNC Center for AIDS Research (NIH award P30 AI50410) and the UNC Lineberger Comprehensive Cancer Center (NIH award P30 CA16068). **Author contributions:** C.W. and R.S. proposed, designed, and supervised this study. M.-R.A., S.B.J., and N. Garrett directed all of the data collection, experiments, and data analyses. M.M., N.A., O.D.C., L.T., S.Z., and D.D. performed all of the experiments. Phylogenetic analyses were performed by S.K.P., C.A., M.M., and D.M. The manuscript was written by M.-R.A., S.B.J.,

N. Goonetilleke, C.W., and R.S. with editorial help from N. Garrett, S.A.K., D.M.M., and S.K.P.

**Competing interests:** UNC is pursuing IP protection for Primer ID, and R.S. is listed as a coinventor and has received nominal royalties. All other authors declare that they have no competing interests. **Data and materials availability:** All data and computer code associated with this study are present in the paper and the Supplementary Materials or accessible at a publicly available websites (DOIs: 10.5281/zenodo.3370141, 10.5281/zenodo.3372204, and 10.5281/zenodo.3372202). The near full-length genome sequences are deposited in GenBank (accession nos. MN09755 to MN097697). The MiSeq sequences are deposited in the Sequencing Read Archive (accession nos. SAMN12126374 to SAMN12126459).

Submitted 4 January 2019

Accepted 29 August 2019

Published 9 October 2019

10.1126/scitranslmed.aaw5589

**Citation:** M.-R. Abrahams, S. B. Joseph, N. Garrett, L. Tyers, M. Moeser, N. Archin, O. D. Council, D. Matten, S. Zhou, D. Doolabh, C. Anthony, N. Goonetilleke, S. A. Karim, D. M. Margolis, S. K. Pond, C. Williamson, R. Swanstrom, The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. *Sci. Transl. Med.* **11**, eaaw5589 (2019).

## The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation

Melissa-Rose Abrahams, Sarah B. Joseph, Nigel Garrett, Lynn Tyers, Matthew Moeser, Nancie Archin, Olivia D. Council, David Matten, Shuntai Zhou, Deelan Doolabh, Colin Anthony, Nilu Goonetilleke, Salim Abdool Karim, David M. Margolis, Sergei Kosakovsky Pond, Carolyn Williamson and Ronald Swanstrom

*Sci Transl Med* 11, eaaw5589.  
DOI: 10.1126/scitranslmed.aaw5589

### Au revoir HIV reservoir?

Curing HIV requires eliminating the latent viral reservoir. To better understand how this reservoir is formed, Abrahams *et al.* examined longitudinal samples from nine South African women with HIV. HIV was grown from resting CD4<sup>+</sup> T cells isolated after several years of antiretroviral therapy (ART) and compared to viral sequences from longitudinal samples collected before ART. They found that in most subjects, reservoir viral sequences were related to viruses present at the time of ART initiation. These results suggest that the majority of the latent reservoir is seeded in response to therapy and likely is not continually formed prior to ART. HIV treatment and cure efforts could potentially exploit this information to limit the size of the reservoir when treatment is initiated.

#### ARTICLE TOOLS

<http://stm.sciencemag.org/content/11/513/eaaw5589>

#### SUPPLEMENTARY MATERIALS

<http://stm.sciencemag.org/content/suppl/2019/10/07/11.513.eaaw5589.DC1>

#### RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/10/430/eaap9927.full>  
<http://stm.sciencemag.org/content/scitransmed/10/437/eaar6759.full>  
<http://stm.sciencemag.org/content/scitransmed/11/520/eaax7350.full>  
<http://science.sciencemag.org/content/sci/366/6470/1197.full>  
<http://science.sciencemag.org/content/sci/366/6470/eaax4380.full>  
<http://science.sciencemag.org/content/sci/366/6470/eaay7199.full>  
<http://science.sciencemag.org/content/sci/366/6470/1199.full>  
<http://science.sciencemag.org/content/sci/366/6470/1255.full>

#### REFERENCES

This article cites 57 articles, 25 of which you can access for free  
<http://stm.sciencemag.org/content/11/513/eaaw5589#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Translational Medicine* (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works