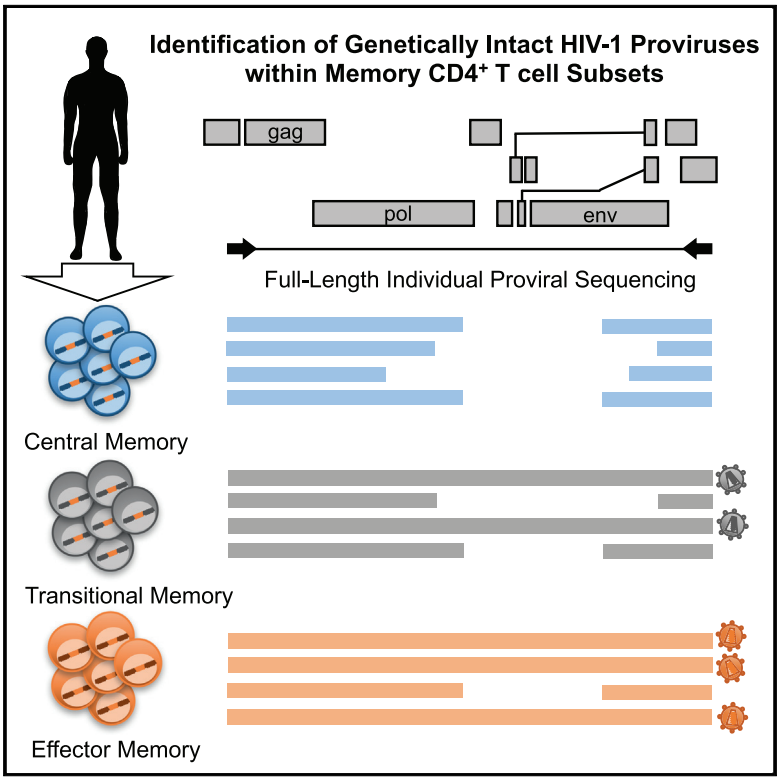
Cell Reports

Resource

Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants

Graphical Abstract



Authors

Bonnie Hiener, Bethany A. Horsburgh, John-Sebastian Eden, ..., Steven G. Deeks, Frederick M. Hecht, Sarah Palmer

Correspondence

bonnie.hiener@sydney.edu.au

In Brief

Latent, replication-competent HIV-1 proviruses pose a significant barrier to an HIV-1 cure. Hiener et al. present the Full-Length Individual Proviral Sequencing (FLIPS) assay to reveal the distribution of genetically intact and potentially replication-competent HIV-1 proviruses in different T cell subsets isolated from individuals on long-term antiretroviral therapy.

Highlights

- FLIPS utilizes NGS to sequence and genetically characterize HIV-1 proviruses
- FLIPS identifies genetically intact and likely replication-competent HIV-1 proviruses
- Identical HIV-1 proviruses suggest maintenance of reservoir by cellular proliferation
- Demonstrate the advantages of FLIPS over other common HIV-1 sequencing assays

Hiener et al., 2017, Cell Reports *21*, 813–822
October 17, 2017
https://doi.org/10.1016/j.celrep.2017.09.081

CellPress

Cell Reports
Resource

Identification of Genetically Intact HIV-1 Proviruses in Specific CD4$^+$ T Cells from Effectively Treated Participants

Bonnie Hiener,[1,9,*] Bethany A. Horsburgh,[1] John-Sebastian Eden,[1,2] Kirston Barton,[1] Timothy E. Schlub,[3] Eunok Lee,[1] Susanne von Stockenstrom,[4] Lina Odevall,[4] Jeffrey M. Milush,[5] Teri Liegler,[5] Elizabeth Sinclair,[5] Rebecca Hoh,[5] Eli A. Boritz,[6] Daniel Douek,[7] Rémi Fromentin,[8] Nicolas Chomont,[8] Steven G. Deeks,[5] Frederick M. Hecht,[5] and Sarah Palmer[1]

[1]Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, NSW 2145, Australia
[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
[3]Sydney School of Public Health, Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
[4]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Karolinska University Hospital, Stockholm 171 77, Sweden
[5]Department of Medicine, University of California, San Francisco, San Francisco, CA 94110, USA
[6]Virus Persistence and Dynamics Section, Vaccine Research Center, National Institute of Allergy and Infectious Disease, NIH, Bethesda, MD 20814, USA
[7]Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Disease, NIH, Bethesda, MD 20814, USA
[8]Centre de Recherche du CHUM and Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montréal, QC H2X 0A9, Canada
[9]Lead Contact
*Correspondence: bonnie.hiener@sydney.edu.au
https://doi.org/10.1016/j.celrep.2017.09.081

SUMMARY

Latent replication-competent HIV-1 persists in individuals on long-term antiretroviral therapy (ART). We developed the Full-Length Individual Proviral Sequencing (FLIPS) assay to determine the distribution of latent replication-competent HIV-1 within memory CD4$^+$ T cell subsets in six individuals on long-term ART. FLIPS is an efficient, high-throughput assay that amplifies and sequences near full-length (~9 kb) HIV-1 proviral genomes and determines potential replication competency through genetic characterization. FLIPS provides a genome-scale perspective that addresses the limitations of other methods that also genetically characterize the latent reservoir. Using FLIPS, we identified 5% of proviruses as intact and potentially replication competent. Intact proviruses were unequally distributed between T cell subsets, with effector memory cells containing the largest proportion of genetically intact HIV-1 proviruses. We identified multiple identical intact proviruses, suggesting a role for cellular proliferation in the maintenance of the latent HIV-1 reservoir.

INTRODUCTION

Antiretroviral therapy (ART) successfully suppresses HIV-1 replication, reduces viral load, and increases the life expectancy of infected individuals (Palella et al., 1998; Palmer et al., 2008). Despite this, ART is not curative, as HIV-1 remains latent in resting memory CD4$^+$ T cells not targeted by ART or the immune system (Finzi et al., 1997). Bruner et al. (2016) recently demonstrated that 93%–98% of latent proviruses in HIV-infected individuals on ART are defective and replication incompetent. Common mechanisms that contribute to defective proviruses include mutations from an error-prone HIV-1 reverse transcriptase (Abram et al., 2010), template switching during reverse transcription (Ho et al., 2013), and/or apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC)-induced hypermutation (Harris et al., 2003; Lecossier et al., 2003). Despite the high prevalence of defective proviruses, it is clear that replication-competent proviruses persist in individuals on long-term ART, as viral load rapidly rebounds if therapy is interrupted (Chun et al., 2010; Davey et al., 1999). Determining the source of latent replication-competent HIV-1 is vital to identifying cellular targets for future curative strategies.

Genetic characterization of the latent HIV-1 reservoir is an important tool for understanding persistent HIV-1 during long-term ART. Single-proviral sequencing (Josefsson et al., 2013a) and single-genome sequencing (SPS/SGS) (Palmer et al., 2005) are methods that genetically characterize sub-genomic regions of the HIV-1 genome. SPS/SGS have provided insight into the distribution, dynamics, and persistence of the latent HIV-1 reservoir (Josefsson et al., 2013b; Evering et al., 2012; Chomont et al., 2009; von Stockenstrom et al., 2015), yet these methods are limited because they target sub-genomic regions of the HIV-1 genome and therefore cannot capture the complete diversity and replication competency of the HIV-1 proviruses. Furthermore, the use of SPS/SGS has identified large expansions of identical HIV-1 sequences, suggesting that cellular proliferation contributes to the persistence of HIV-1 during therapy, but it remains unknown whether these HIV-1 sequences are identical or even intact throughout the entire HIV-1 genome

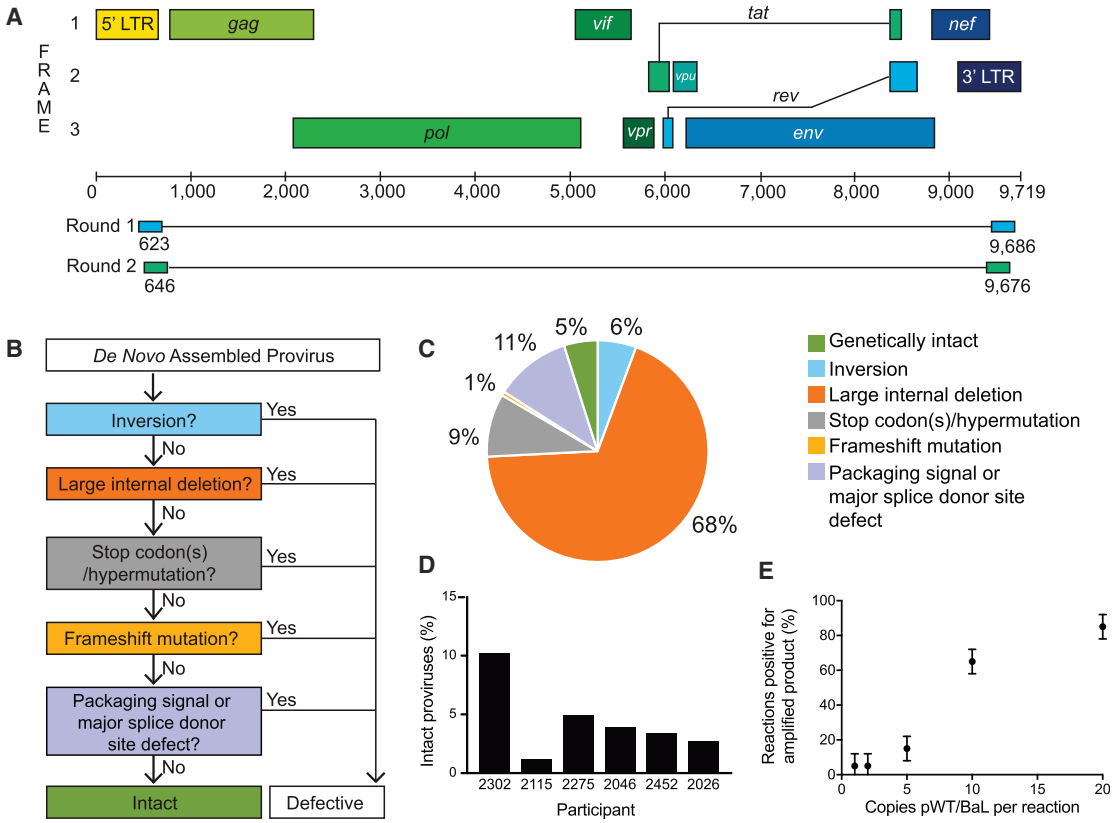


Figure 1. Full-Length Individual Proviral Sequencing Assay

(A) Primer binding sites in HIV-1 5′ and 3′ LTR regions used by full-length individual proviral sequencing (FLIPS) to amplify full-length (defective and intact) HIV-1 proviruses via nested PCR.

(B) Process used to identify genetically intact and defective *de novo* assembled proviruses.

(C) Percentage of defective and intact proviruses within the total pool of sequenced proviruses.

(D) Percentage of intact proviruses sequenced for each participant.

(E) Percentage of PCR reactions positive for amplified product with increasing copies of pWT/BaL.

Data are represented as mean ± SD (n = 2).

(Laskey et al., 2016). Full-length HIV-1 proviral sequencing methods, which sequence ~9 kb of the HIV-1 genome, overcome the limitations of SPS. Previously available full-length HIV-1 proviral sequencing methods have provided insight into the prevalence and development of defective proviruses (Bruner et al., 2016; Ho et al., 2013). These assays require multiple internal sequencing primers that carry the risk of erroneously identifying defective proviruses and make resolving the entire proviral sequence technically challenging. Additionally, it is possible that these methods may not capture the entire population of proviruses present in an individual as, due to the number and complexity of primers used, these methods may be influenced by primer mismatches. In response to these limitations, we and others have developed next-generation sequencing (NGS)-based assays to sequence near-full-length HIV-1 proviruses (Lee et al., 2017; Imamichi et al., 2016). Here, we present the Full-Length Individual Proviral Sequencing (FLIPS) assay: a high-throughput assay utilizing NGS to sequence single, full-length HIV-1 proviruses and predict their potential replication competency by comparative genomics. We apply FLIPS to determine the distribution of intact and potentially replication-competent proviruses within memory CD4$^+$ T cell subsets isolated from six individuals on long-term ART and demonstrate the advantages of FLIPS over existing sequencing methods.

RESULTS

FLIPS

In conducting the FLIPS assay, NGS is used to sequence single (intact and defective) HIV-1 proviruses. Two rounds of nested PCR run at limiting dilution using primers that target the highly conserved 5′ and 3′ U5 long terminal repeat (LTR) regions are used to amplify ~9 kb of the HIV-1 genome (Figure 1A). High-throughput NGS of amplified proviruses is undertaken on the Illumina MiSeq platform, and individual proviral contigs are generated by *de novo* assembly. To identify defective proviruses, a stringent process of elimination is used (Figure 1B) whereby proviruses containing inversion sequences, large internal deletions (>100 bp), hypermutation/deleterious stop codons, frameshifts, or mutations in the packaging signal or major splice

Table 1. Participant Characteristics

Participant (SCOPE ID)	Sex	Age	Viral Load (Copies/mL)[a]	$CD4^+$ T Cell Count (Cells/μL)[a]	Time of Infection before Initiation of Therapy (Months)	ART Duration (Years)[a]	Therapeutic Regimen
2302	male	27	<40	696	3.4	4.6	FPV, RTV, TDF/FTC
2115	male	51	<40	601	0.6	17.3	FTC/TDF, NVP
2275	male	47	<40	1,842	0.8	15.3	FTC/TDF, NVP
2452	male	66	<40	604	24.4	3.2	MVC, RAL, ETR
2026	male	59	<40	476	117	17.7	TDF, ABC/3TC, RTV, DRV
2046	male	50	<40	1,099	70	16.3	EFV/TDF/FTC

FPV, fosamprenavir; RTV, ritonavir; TDF, tenofovir disoproxil fumarate; FTC, emtricitabine; NVP, nevirapine; MVC, maraviroc; RAL, raltegravir; ETR, etravirine; ABC, abacavir; 3TC, lamivudine; DRV, darunavir; EFV, efavirenz.
[a]At time of sampling.

donor (MSD) site are defined as defective and replication incompetent. Proviruses lacking these defects are considered genetically intact and potentially replication competent. In optimizing the FLIPS assay, we first determined the extent to which FLIPS amplifies the HIV-1 DNA present in a sample. To do this, we diluted the HIV-1 plasmid pWT/BaL (obtained through the AIDS Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases [NIAID], NIH: HIV-1 HXB3/BaL Infectious Molecular Clone (pWT/BaL) from Dr. Bryan R. Cullen) from 20 copies down to 1 copy and performed multiple full-length PCR reactions at each dilution, recording the number of reactions with amplified product (Figure 1E). The number of positive reactions decreased with the number of templates per reaction, with FLIPS amplifying 5% of the reactions when pWT/BaL was diluted to 1 and 2 copies per reaction. In calculating the Poisson distribution for this experiment, the expected number of amplified templates per reaction closely matched the actual number of copies per reaction, supporting that, at dilutions of 1 and 2 copies per well, the amplified products originate from a single template.

The classification of a sequenced provirus as intact relies on the absence of deletions, insertions, and/or mutations. Next, it was important to determine the error rate of the FLIPS assay, as polymerase errors could artificially introduce defects within the final proviral contig. We amplified the intact pWT/BaL plasmid at limiting dilution and sequenced 10 amplicons that we compared to the available pWT/BaL reference sequence. None of the sequences contained deletions or insertions; however, one sequence contained a single nucleotide change (C|T) at position 1519 (relative to HXB2). This allowed determination of the error rate for the FLIPS assay as 1 change in every 89,330 nt sequenced (or 1.12×10^{-5} nt^{-1}) (confidence interval, 1 in 16,000 to 1 in 3,530,000).

Latent HIV-1 Reservoirs of Individuals on Long-Term ART Contain Few Genetically Intact Proviruses

Replication-competent proviruses persist in a latent state in individuals on suppressive ART, and viral load rapidly rebounds upon the cessation of therapy (Chun et al., 2010; Davey et al., 1999). Previous studies have identified latent replication-competent HIV-1 in a number of $CD4^+$ T cell subsets including naive (T_N), central (T_{CM}), transitional (T_{TM}), and effector (T_{EM}) memory (Chomont et al., 2009; Soriano-Sarabia et al., 2014). To determine whether the distribution of potentially replication-competent HIV-1 proviruses was different between cell subsets, we applied the FLIPS assay to genetically characterize the latent HIV-1 reservoir in different $CD4^+$ T cell subsets isolated from the peripheral blood of participants on long-term suppressive ART. We obtained T_N, T_{CM}, T_{TM}, and T_{EM} $CD4^+$ T cell subsets from six participants who were on suppressive ART for 3–17 years (3 initiated ART during acute/early infection (2115, 2275, and 2302), and 3 initiated ART during chronic infection (2026, 2046, 2452) (Figure S2; Table 1).

Applying FLIPS, we obtained a total of 531 sequences from the six participants. HIV-1 can also exist in the form of 2-LTR circles within infected cells, albeit at very low levels. However, no 2-LTR circles were identified within total $CD4^+$ T cells from four of the participants with cells available for 2-LTR quantification (Table S1), indicating that the sequences we obtained were most likely derived from proviral DNA (Vandergeeten et al., 2014). Of the 531 sequences, 26 (5%) were characterized as genetically intact and potentially replication competent (Figure 1C). The total frequency of intact proviruses ranged from 1.3% to 10.3% in the individual participants (Figure 1D). Defective proviruses were identified following our stringent process of elimination (Figure 1B): 6% of proviruses were rendered defective due to an inversion sequence, 68% due to a large internal deletion, 9% by the presence of deleterious stop codons, and 1% by frameshift mutations. We identified 11% of proviruses that were intact in the coding region but contained deletions in the packaging signal and/or mutations in the MSD site, rendering them defective (Bruner et al., 2016; Ho et al., 2013) (Figure 1C). The proportion of intact proviruses in participants who initiated therapy during acute/early infection (6%) and chronic infection (3%) was similar to those obtained by Bruner et al. (2016) (acute 7% and chronic 2%). Additionally, the proportions of the different defective proviruses were similar across the two studies (Table S2).

Genetically Intact Proviruses Are Unequally Distributed in $CD4^+$ T Cell Subsets

The FLIPS assay was used to assess the frequency of genetically intact HIV-1 genomes within different $CD4^+$ T cell subsets in a two-step process that, first, measures the number of cells positive for HIV-1 and, second, measures the proportion of these positive cells that have genetically intact HIV-1 genomes. We found strong evidence for a difference in the number of

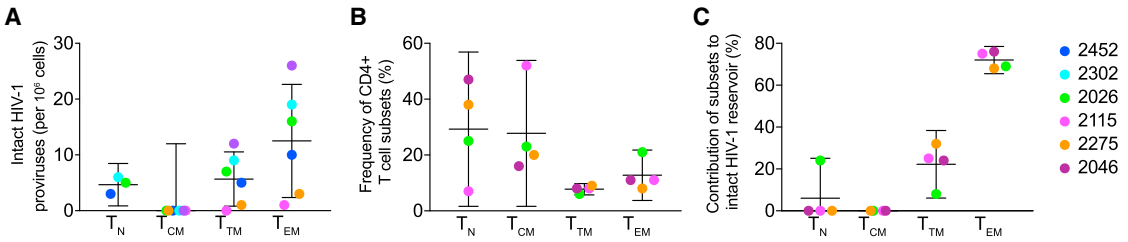


Figure 2. Comparison of Intact Proviruses between Cell Subsets

(A) Number of intact proviruses per million cells in each cell subset.

(B) Proportion of each cell subset to the total pool of $CD4^+$ T cells.

(C) Contribution of each $CD4^+$ T cell subset to the overall intact reservoir. Data are represented as mean ± confidence interval (CI) for all graphs. Confidence intervals are calculated from one-sample t tests, except for T_{CM} in (A), whose confidence interval upper bound is calculated from the product of the mean proportion of HIV+ cells in the T_{CM} subset, and the upper bound of the confidence interval for the proportion of proviruses that are intact for the T_{CM} subset. See also Figure S1.

HIV-1-positive cells across the cell subsets ($p = 0.003$, $\chi^2 = 13.70$, df = 3; Figure S1A). We observed T_N to contain the least HIV-1-positive cells, followed by T_{CM}, T_{EM}, and, finally, T_{TM}. We also found strong evidence that the proportion of intact proviruses was different across the cell subsets ($p < 0.001$, $\chi^2 = 20.02$, df = 3; Figure S1B). We observed the order from lowest to highest intact proportion to be $T_{CM} < T_{TM} < T_N < T_{EM}$. As the frequency of intact HIV-1 proviruses is a product of the proportion of cells that are HIV-1 positive, and the proportion that are intact, these results provide strong evidence that the frequency of intact HIV-1 proviruses differed across the cell subsets. We observed the order of this frequency from lowest to highest to be $T_{CM} < T_N < T_{TM} < T_{EM}$ (Figure 2A). In all participants, T_{EM} contained the highest number of intact proviruses (range = 1–26; median, 13 per 10^6 cells). We did not identify any intact proviruses in the T_{CM} subset in any of the six participants, despite obtaining 125 sequences from seven million cells (Table 2); however, we calculated the upper bound of the confidence interval for T_{CM} to be 12 intact proviruses in 10^6 cells (Figure 2A). Moreover, the proportion of each cell subset to the total pool of $CD4^+$ T cells can differ between individuals (Figure 2B) (Josefsson et al., 2013b). When the contribution of each cell type to the overall reservoir was calculated for four participants, we found that the highest number of intact HIV-1 genomes was located in T_{EM} cells (range = 68%–100% of contribution; median, 76%; Figure 2C).

Detection of Identical Intact HIV-1 Proviruses in Participants on Long-Term ART

The stability and longevity of the HIV-1 reservoir are thought to be maintained via the proliferation of cells containing HIV-1 proviruses (Maldarelli et al., 2014; Wagner et al., 2014; Cohn et al., 2015; Lee et al., 2017), with antigen-driven and/or homeostatic proliferation of $CD4^+$ T cells containing an integrated HIV-1 provirus the likely mechanism (von Stockenstrom et al., 2015; Chomont et al., 2009). Phylogenetic trees of proviral contigs were prepared for each participant (Figure 3; Data S1). Genetically identical proviruses are indicative of cellular proliferation, as the error rate of the HIV-1 reverse transcriptase limits the transcription of identical proviruses by *de novo* replication. In all participants, we observed evidence of proliferation via the presence of identical (≥2) proviral sequences. The majority (92%) of identical HIV-1 sequence expansions included defective proviruses. Interestingly, in participant 2026, 32% of sequences were derived from an identical HIV-1 sequence expansion located in T_{EM}/T_{TM} cells, which was intact in all coding regions but had a deletion in the packaging loop and MSD site, rendering it defective and, likely, replication incompetent (Ho et al., 2013) (Figure 3). In three participants (SCOPE cohort participants 2026, 2046, and 2275; Figure 3, Data S1A, and Data S1C, respectively), we observed expansions of genetically identical intact proviruses in both T_{TM} and T_{EM} cells. The detection of identical intact proviruses supports previous findings (Simonetti et al., 2016; Lorenzi et al., 2016; Bui et al., 2017) and provides further evidence that the pool of latent replication-competent proviruses is maintained, at least in part, by cellular proliferation of memory $CD4^+$ T cells.

FLIPS Provides Advantages over Existing Assays that Genetically Characterize the Latent Reservoir

We compared the number of intact proviruses identified using FLIPS to integrated DNA measurements that were available for four participants and were obtained by the method described previously (Vandergeeten et al., 2014). As expected, we found no evidence of a correlation between the number of intact HIV-1 genomes and the number of integrated proviruses per million cells ($p = 0.86$). We did, however, find evidence that the number of intact HIV-1 proviruses per million cells measured with the FLIPS assay is lower than the frequency of cells harboring integrated HIV-1 DNA ($p < 0.01$; Figure 4A).

To determine whether SPS of sub-genomic regions and the FLIPS assay identified equivalent numbers of intact HIV-1 genomes, we compared the number of intact proviruses identified in each cell subset using the FLIPS assay to those that might have been identified if we had used SPS. To ensure that we were working with the same population of cells as the FLIPS assay, we simulated the SPS of *env* (V1–V3; HXB2 positions 6430–7374) and *gag-pol* (p6 through nt 1–900 of the gene encoding reverse transcriptase; HXB2 positions 1827–3528) (von Stockenstrom et al., 2015) rather than conducting the SPS assay on a different population of cells from each participant. We chose the *env* and *gag-pol* regions, as these gene regions have been

Table 2. Number of Cells and Sequences Obtained for Each Subset and Participant

	Naive			Central Memory			Transitional Memory			Effector Memory		
Participant (SCOPE ID)	Defective	Intact	No. of Cells Analyzed	Defective	Intact	No. of Cells Analyzed	Defective	Intact	No. of Cells Analyzed	Defective	Intact	No. of Cells Analyzed
2302	15	4	3,287,933	35	0	704,122	31	3	820,287	32	6	728,025
2115	0	0	0	28	0	2,294,141	29	0	859,259	22	1	1,888,889
2275	0	0	0	3	0	251,852	21	0	1,057,688	33	3	1,330,640
2046	0	0	6,465	24	0	190,647	1	0	4,048	23	2	153,236
2452	35	2	641,975	34	0	3,440,763	37	0	225,734	33	3	459,055
2026	26	0	152,727	1	0	47,685	10	1	133,448	32	1	51,046
Total	76	6	4,089,100	125	0	6,929,210	129	4	3,100,464	175	16	4,610,891

analyzed in a number of studies (Barton et al., 2016; Josefsson et al., 2011, 2012, 2013a, 2013b; von Stockenstrom et al., 2015). From our full-length contigs, we identified *env* and/or *gag-pol* sequences with complete sequencing primer sites (primers E20, E30, E115, and E125 [*env*]; and primers 1849, 3500, 1870, and 3410 [*gag-pol*]) (Josefsson et al., 2013b; von Stockenstrom et al., 2015). This method of simulation may have overestimated the number of *env* and/or *gag-pol* sequences obtained, as it did not take into account primer mismatches that may have affected PCR amplification prior to sequencing (Kearney et al., 2008). Nonetheless, we identified intact and defective proviruses from our populations of *env* and *gag-pol* sequences using the same stringent screening method as the FLIPS assay. When comparing the number of intact proviruses identified using FLIPS to the number of those identified using SPS for *env* and *gag-pol*, we found that the number of intact proviruses identified by simulated SPS for these sub-genomic regions increased, on average, over patients and cell subsets by a factor of 13.2, $t(14) = 10.9$, $p < 0.001$; and a factor of 17.7, $t(13) = 10.7$, $p < 0.001$, respectively (Figure 4B). We identified a number of sequences derived from T_{CM} that were intact in the *env* and *gag-pol* regions but were defective when we examined the entire proviral genome using FLIPS. Additionally, as the regions chosen for SPS did not cover the packaging signal and MSD site in the 5′ non-coding region, we were unable to identify sequences that were defective due to deletions or mutations in this region.

DISCUSSION

We have developed a high-throughput and efficient method for genetically characterizing latent proviruses that overcomes the limitations of previous full-length sequencing assays and SPS. Our NGS-based FLIPS approach is able to sequence ~9 kb of the proviral genome without the use of multiple internal Sanger sequencing primers, some of which may not match all viruses within the population. Therefore, unlike the full-length sequencing assays, which use a Sanger approach (Ho et al., 2013; Bruner et al., 2016), FLIPS does not carry the risk of erroneously identifying deletions within variant virus sequences that do not match internal sequencing primers. By eliminating the need to use multiple internal sequencing primers, FLIPS is also able to resolve the entire sequence of an intact or defective provirus, a feat that is technically challenging and has not always been achieved using a Sanger-based approach (Bruner et al., 2016; Ho et al., 2013). It is important to note that, although FLIPS uses primers targeted to highly conserved regions of the HIV-1 LTR, it too may be affected by primer mismatches in this region. Clearly, FLIPS has simplified the method for obtaining full-length HIV-1 proviral sequences and improved the quality of the sequences obtained, and this may explain the larger number of sequences obtained per participant in this study (average, 88 sequences per participant; range = 50–144) compared to previous studies that used the Sanger approach (average, 27 sequences per participant [Ho et al., 2013]; and average, 16 sequences per participant [Bruner et al., 2016]). It is important to note that the proportion of intact proviruses and the different types of defective proviruses obtained in this study were similar to those obtained by Bruner et al. (2016), indicating the reproducibility of obtaining the genetic sequence of HIV-1 proviruses from participants on long-term ART. Recent studies also utilized NGS for sequencing near-full-length HIV-1 proviruses (Lee et al., 2017; Imamichi et al., 2016); however, the advantages of this approach over SPS and existing Sanger-based assays were not presented or discussed.

FLIPS overcomes the limitations of SPS by sequencing near full-length proviral sequences as opposed to sub-genomic regions. We have shown that the number of proviruses identified by SPS as intact in sub-genomic regions is 13- to 18-fold higher than the number identified by FLIPS. Therefore, FLIPS provides a more robust estimate of the number of intact proviruses in a cell population by providing a genome-scale perspective of defective changes. Additionally, nucleotide homology between full-length proviral sequences provides a better prediction of expansions of identical intact sequences than the sequencing of sub-genomic regions.

The FLIPS assay provides a stringent approach to predicting the replication competency of HIV-1 proviruses without requiring these proviruses to be reactivated and spread *in vitro* or *in vivo*. The majority of proviruses were characterized as defective due to the presence of large internal deletions or deleterious stop codons in multiple coding regions. However, a number of proviruses were defined as defective due to a unique deletion in the packaging signal and MSD site. We believe that *in vitro* studies similar to those performed by Ho et al. (2013) are required to determine the true replication competency of some sequenced proviruses. Additionally, as we did not determine the integration sites of the sequenced proviruses, it is possible that some

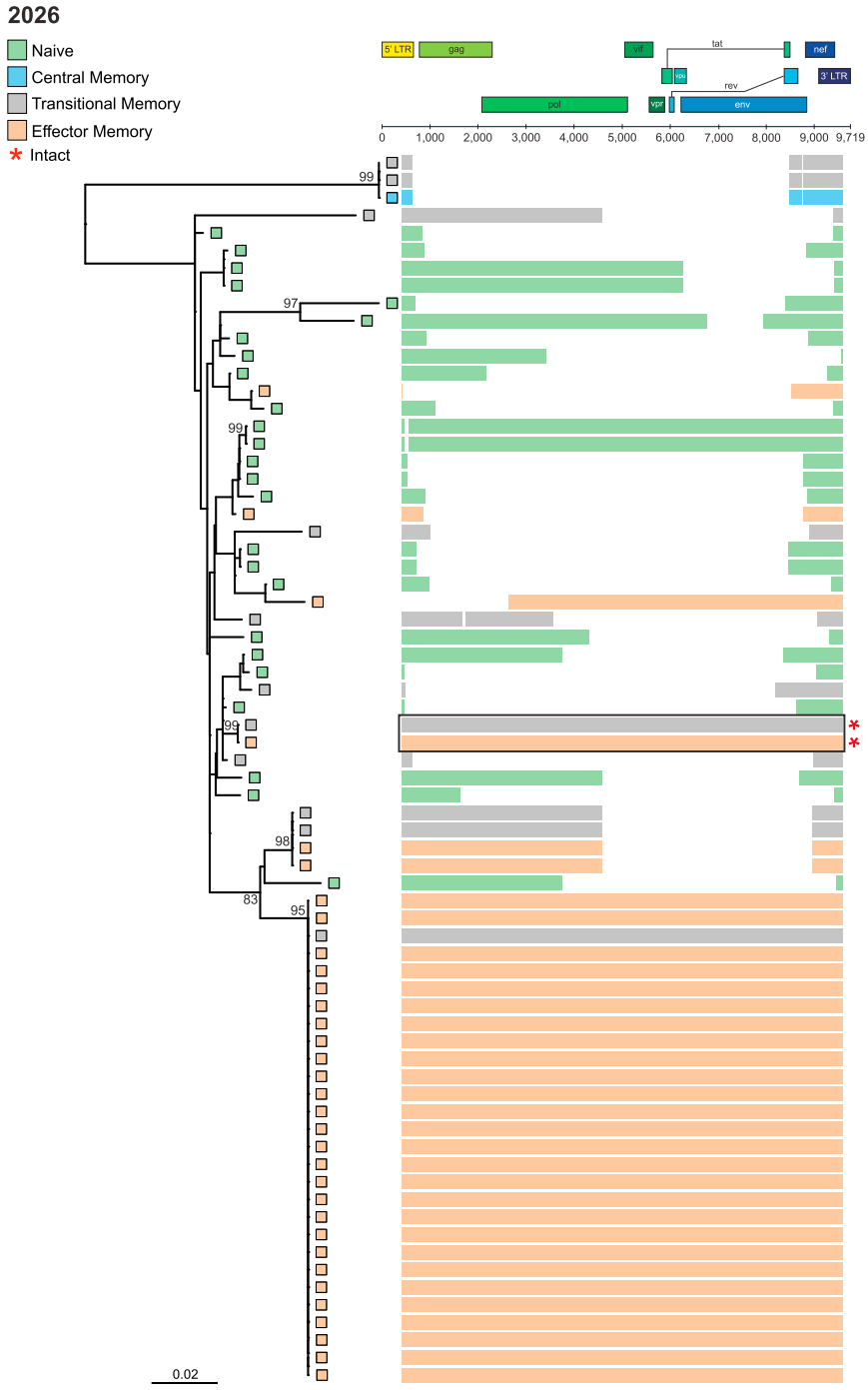


Figure 3. Representative Phylogenetic Tree of HIV-1 Sequences from a Participant on Long-Term ART

Participant 2026. Horizontal lines adjacent to tips represent the individual contig sequence aligned to HXB2. Red asterisks denote intact sequences; black box indicates identical intact sequences. See also Data S1A–S1E.

proviruses defined as genetically intact may be replication incompetent due to integration into a centromere (Lewinski et al., 2005).

Genetic characterization of HIV-1 proviruses provides information about the stability, distribution, and dynamics of the latent reservoir. Using FLIPS, we identified a number of identical intact proviruses, which would not have been possible using traditional SPS methods. The detection of identical intact proviruses suggests a role for cellular proliferation in the maintenance and stability of the latent replication-competent reservoir. Unlike Hosmane et al. (2017), Bui et al. (2017) and Lorenzi et al. (2016), who used the definitive viral outgrowth assay to identify identical replication-competent proviruses in participants on ART, we did not use *in vitro* studies to confirm the replication competency of intact proviruses identified by FLIPS. Therefore, we cannot be certain that the identical intact proviruses we identified are truly replication competent. Additionally, as integration site analysis of identical proviruses was not performed on these participants, we cannot exclude focal infection prior to ART initiation as the source of these identical proviruses. However, the presence of identical defective proviruses in every participant suggests that proliferation of resting memory $CD4^+$ T cells is a common occurrence in participants on long-term ART.

Using the FLIPS assay, we observed the distribution of genetically intact HIV-1 proviruses among naive and memory $CD4^+$ T cell populations. The number of intact proviruses was statistically different across the cell subsets, and we observed T_{EM} to have the highest number of intact proviruses. We did not directly sequence any intact proviruses from the T_{CM} subset; however, using a binomial test, we calculated a conservative estimate of the number of intact proviruses in T_{CM} to be, at most, 12 per 10^6 cells. The presence of intact HIV-1 proviruses in T_{CM} is supported by previous studies that have identified T_{CM} as a reservoir of replication-competent HIV-1 (Soriano-Sarabia et al., 2014; Chomont et al., 2009). Additionally, as T_{CM} cells differentiate into T_{EM} cells after T cell receptor activation and in response to homeostatic cytokines (Sallusto et al., 1999, 2004; Geginat et al., 2003), this may explain why T_{EM} harbors more intact proviruses than T_{CM}. Although we observed no intact provirus in the T_{CM} subset, intact virus may exist at a lower level than the sensitivity of this sample size; therefore,

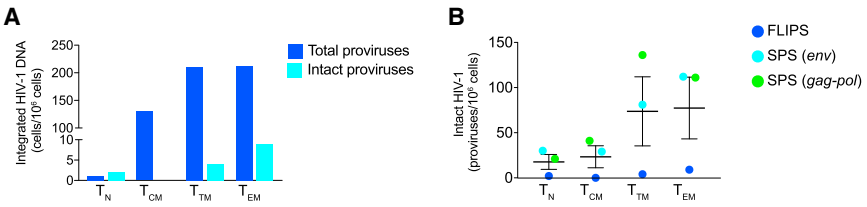


Figure 4. Comparison of FLIPS to Other Assays

(A) Frequency of cells containing either intact or integrated HIV-1 DNA.

(B) Number of intact proviruses identified using the FLIPS assay compared to simulated SPS of *gag-pol* and *env* regions.

Data are represented as mean ± confidence interval (CI) for all graphs.

sequencing of HIV-1 from a larger number of T_{CM} cells may be required to detect intact proviruses.

Future studies into HIV-1 latency and eradication will benefit from utilization of the FLIPS assay. For example, further genetic characterization of latent proviruses could be crucial to determining particular coding or non-coding regions that make the virus more responsive to latency reversing agents (LRAs). Similarly, sequencing of the virus before and after administration of ART could elucidate whether there are any sequence characteristics that could make the virus more likely to become latent. FLIPS could assist in the identification of sites targeted by CRISPR-Cas technology. Lastly, by adapting the FLIPS assay to sequence plasma HIV-1 RNA, the particular subsets that contribute to viral rebound after administration of a LRA or ART interruption could be identified.

EXPERIMENTAL PROCEDURES

Participant Cohort

Leukapheresis or peripheral blood samples were obtained from six HIV-1 subtype-B-positive individuals on long-term suppressive ART (3.2–17.7 years). Participant characteristics, including age and sex are available in Table 1. Participants 2026, 2046, and 2452 initiated therapy during chronic infection, and participants 2115, 2302, and 2275 initiated therapy during acute/early infection. Participants were enrolled in the SCOPE (participants 2026, 2046, 2115, and 2275) or OPTIONS (participants 2452 and 2302) cohorts that are both ongoing longitudinal observational cohorts based at the University of California, San Francisco. Viral loads were monitored every 4 months, and all participants maintained viral suppression (<40–75 copies per milliliter) for at least 3 years prior to this study.

Written informed consent was obtained from all participants. The study was approved by the institutional review boards at the University of California, San Francisco, and the Western Sydney Local Health District, which includes the Westmead Institute for Medical Research.

Isolation of CD4$^+$ T Cell Subsets

Naive and memory CD4$^+$ T cell subsets were isolated by fluorescence-activated cell sorting. For the OPTIONS cohort participants (2452 and 2302), cell subsets were isolated from peripheral blood as previously described (von Stockenstrom et al., 2015). For the SCOPE cohort participants (2026, 2046, 2115, and 2275), cell subsets were isolated from leukapheresis samples, as described later. Both methods sorted memory CD4$^+$ T cell subsets on expression of CD45RO followed by CD27+/CCR7+ (T_{CM}), CD27−/CCR7− (T_{EM}), and CD27+/CCR7− (T_{TM}) to obtain the relevant memory subpopulations.

For the SCOPE cohort participants, naive and memory CD4$^+$ T cell subsets were isolated as follows (Figure S2). Peripheral blood mononuclear cells (PBMCs) were obtained from leukapheresis samples by Ficoll-Hypaque density gradient centrifugation followed by isolation of total CD4$^+$ T cells by magnetic negative selection, according to the manufacturer's protocol (Stem Cell Technologies). Cells were initially sorted using the following antibodies: CD3-Alexa Fluor 700 (clone UCHT1, BD Biosciences #557943), CD4-APC (clone RPA-T4, BD Biosciences #555349), CD14-V500 (clone M5E2, BD Biosciences #561391), LIVE/DEAD Fixable Aqua Dead Cell Stain marker (Invitrogen #L34957), CD45RA-BV650 (clone HI100, BioLegend #304135), and CD45RO-PerCP-eFluor 710 (clone UCHL1, eBioscience #40-0457-42). CD45RO+/CD45RA− cells were classified as memory cells and further sorted into T_{CM}, T_{TM}, and T_{EM} subsets based on their expression of CD27 and CCR7 using the antibodies CD27-APCeFluor780 (clone O323, eBioscience #47-0279-42) and CCR7-PECF564 (clone 150503, BD Biosciences #562381), with a purity for each T cell subset of 95%–100%, 67%–92%, and 85%–98%, respectively. CD45RO−/CD45RA+ cells were further sorted into T_N based on the expression of CD27, CCR7, CD127, and CD95 using the additional antibodies CD127-PE (clone HIL-7R-M21, BD Biosciences #557938) and CD95-PE-Cy7 (clone DX2, BD Biosciences #561633), with a purity of 100%.

The number of cells obtained ranged from 5×10^4 to 12×10^6 for both sorting strategies. Cell pellets were stored at −80°C. Cells were lysed in 100 μL lysis buffer (10 mM Tris-HCl [Invitrogen], 0.5% Nonidet P-40, 0.5% Tween-20 [Sigma], and 0.3 mg/mL proteinase K [Ambion]) per 10^6 cells by incubating at 55°C for 1 hr followed by incubation at 85°C for 15 min to obtain genomic DNA for PCR amplification.

2-LTR Circles and Integrated DNA Quantification

2-LTR circles from total CD4$^+$ T cells and integrated DNA from naive and memory T cell subsets were quantified as previously described (Vandergeeten et al., 2014).

FLIPS Assay

Nested PCR

To amplify single full-length HIV-1 proviruses, lysed CD4$^+$ T cells were serially diluted 1:3, 1:9, 1:27, and 1:81, and 20 reactions at each dilution were performed with two rounds of nested PCR using primers specific to the HIV-1 5′ and 3′ U5 LTR regions. This PCR amplification has been modified from previously reported assays (Bruner et al., 2016; Ho et al., 2013; Li et al., 2007). Briefly, for the first round PCR, 2 μL diluted DNA was amplified in a 40-μL reaction containing 1 μM primers (BLOuter forward [BLOuterF]: 5′-AAATCTC TAGCAGTGGCGCCCGAACAG-3′, HXB2 position 623–649; and BLOuter reverse [BLOuterR]: 5′-TGAGGGATCTCTAGTTACCAGAGTC-3′, HXB2 position 9662–9686), 1× High Fidelity Buffer (600 mM Tris-SO_4 [pH 8.9], 180 mM $(NH_4)_2SO_4$; Invitrogen), 2 mM $MgSO_4$ (Invitrogen), 0.2 mM dNTPs (Promega), and 0.025 U/μL Platinum Taq High Fidelity (Invitrogen). Each PCR plate contained 85 PCR reactions: 80 sample reactions, 4 negative control reactions, and a positive control reaction containing pWT/BaL (obtained through the AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 HXB3/BaL Infectious Molecular Clone (pWT/BaL), catalogue number 11414, from Dr. Bryan R. Cullen). PCR conditions for the first round were 94°C for 2 min, then 94°C for 30 s, 64°C for 30 s, and 68°C for 10 min for 3 cycles; 94°C for 30 s, 61°C for 30 s, and 68°C for 10 min for 3 cycles; 94°C for 30 s, 58°C for 30 s, and 68°C for 10 min for 3 cycles; 94°C for 30 s, 55°C for 30 s, and 68°C for 10 min for 21 cycles; and then 68°C for 10 min. The first-round PCR reaction was diluted 1:3 in Tris-HCl (5 mM, pH 8), and 2 μL of the diluted reaction was transferred to the 30-μL second-round reaction (primers 275 forward [275F]: 5′-ACA GGGACCTGAAAGCGAAAG-3′, HXB2 positions 646–666; and 280 reverse [280R]: 5′-CTAGTTACCAGAGTCACACAACAGACG-3′, HXB2 positions 9650–9676). Second-round PCR conditions were identical to the first round, except that an additional 10 cycles at 94°C for 30 s, 55°C for 30 s, and 68°C for 10 min were added. Wells positive for amplified HIV-1 proviruses were identified by diluting the second-round PCR reaction 1:3 with Tris-HCl (5 mM, pH 8), followed by visualization on a 1% agarose gel. According to the Poisson distribution, the dilution at which 30% of PCR reactions are positive has an

80% probability of containing a single amplified provirus. Therefore, the dilution where approximately 30% of the amplicons were positive was selected, and additional PCRs were completed until 30–40 single proviruses were amplified per T cell subset where possible.

DNA Purification and Quantification

Proviral DNA from each single amplified HIV-1 provirus was purified using the QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions with minor modifications. These modifications included heating the elution buffer (Buffer EB) to 60°C and incubating the buffer on the column for 3 min prior to the final spin. Each HIV-1 provirus was then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and diluted in water to a final DNA concentration of 0.2 ng/μL as input for Nextera XT library preparation.

Nextera XT Library Preparation

Amplified proviruses were prepared for NGS using the Nextera XT DNA Library Preparation Kit (Illumina) with indexing of 96 samples per run. All tagmentation, PCR amplification, and clean-up steps were performed according to the manufacturer's instructions, except that reagent and input DNA volumes were halved, and libraries were normalized manually. To do this, the concentration of each individual provirus Nextera XT library was determined using the KAPA Illumina Universal NGS qPCR Kit (KAPA Biosystems) and then combined in equimolar amounts to a concentration of 4 nM. The final pooled 96-sample library was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) with the average fragment lengths determined by bioanalyzer (Agilent). The pooled library was denatured by combining equal volumes (5 μL each) with 0.2 N NaOH before neutralization with 5 μL of 200 mM Tris-HCl. The final library was then diluted to 12.5 pM with chilled Hybridization Buffer (HT1) immediately prior to sequencing. Pooled libraries underwent 2 × 150 nt paired-end sequencing on the Illumina MiSeq platform (300-cycle kit v2), which was performed by the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia). This yielded approximately 20 million paired-end reads per run or 200,000 reads per individual provirus library for analysis following demultiplexing.

De Novo Assembly of Unique Proviruses

Following amplification and sequencing, we assembled, *de novo*, individual proviruses from the paired-end reads using a custom designed workflow in CLC Genomics Workbench 9 (QIAGEN) (Figure S3). This workflow involved the following specific steps. (1) Quality control: removal of Illumina adaptor sequences and ambiguous nucleotides, trimming of 5′ and 3′ terminal nucleotides, and a stringent quality limit of 0.001 that corresponds to a quality control phred score of 30. (2) Merge overlapping pairs: paired forward and reverse reads with overlapping regions were merged to form single extended reads. (3) *De novo* assembly: 10,000 non-overlapping paired reads were subsampled to obtain an expected coverage of ~200× and were then *de novo* assembled using the native CLC Genomics assembler with a word size of 30 nt and a maximum bubble size of 65 nt. (4) Map reads to *de novo* assembled contig: the full read set was then re-mapped to the *de novo* assembled contig before extraction of a final majority consensus sequence. The final *de novo* assembled contig length was compared to the size of the corresponding band on the original gel. The minimum acceptable average coverage was 1,000 reads. The final consensus underwent further quality control measures before being confirmed as a unique HIV-1 proviral sequence as described later.

The majority of the proviral contigs were assembled with the aforementioned workflow. In some circumstances, multiple contigs with a similar coverage were assembled. These contigs were then aligned to a full-length HIV-1 (9-kb) reference from the same participant using the Align Contigs tool in the CLC Genome Finishing Module, which allowed for manual assembly of a final single consensus. As before, all reads were re-mapped to this single consensus and then accepted as a final HIV-1 proviral sequence if the read coverage was even throughout the assembly and absent of SNPs with a frequency of >40%.

Further Quality Control Measures

(1) BLAST search: non-HIV contigs from non-specific amplification were identified through a BLAST search. The amount of non-HIV sequences identified ranged from 1% to 14% in the participants and they were mainly evident when the infection frequency of a cell subset was low. These contigs were not included in further analyses. (2) Identification of mixtures: to further ensure that each contig represented a single provirus, we screened all contigs and identified those assembled from multiple amplified HIV-1 proviral templates (mixtures) on the basis of read coverage and variant calling. Reads were mapped to a full-length HIV-1 (~9 kb) reference from the same participant. Mixtures were subsequently identified by uneven read coverage (e.g., when proviral contigs of different lengths were co-amplified) or by the presence of SNPs with a frequency of >40%. The percentage of amplified proviruses that were identified as mixtures ranged from 3% to 15% per participant. Mixed populations were not included in subsequent analyses.

Assembled proviral contigs were imported into Molecular Evolutionary Genetics Analysis (MEGA) 6 and aligned to the HIV-1 reference genome HXB2 using MAFFT v7 and manual editing where appropriate. To identify cases of inter-participant or positive control pWT/BaL contamination, an alignment and neighbor-joining tree were prepared containing all participants and the pWT/BaL sequence (MEGA 6). No contamination was observed, as each participant's sequences formed a unique monophyletic group distinct from the control sequence.

Identification of Intact Proviruses

To identify genetically intact proviruses, we developed a workflow that identifies HIV-1 proviral contigs with inversions, large internal deletions, deleterious stop codons and hypermutation, frameshift mutations, and defects in the packaging signal/MSD (Figure 1B).

Inversions

Inversions, which are regions of reverse complementarity, were identified manually in the alignment stage since they included regions that did not align to the reference sequence HXB2 but, following reverse complementation, were able to be aligned. All contigs containing inversions were subsequently omitted from the final alignment of full-length sequences for each T cell subset, as homologous alignments could not be prepared from these sequences in any orientation.

Large Internal Deletions

Contigs containing large internal deletions (>100 nt) were identified at the alignment stage and defined as defective.

Deleterious Stop Codons and Hypermutation

All contigs of a length of >8,900 nt were assumed to be nearly full length and were checked for the presence of deleterious stop codons and hypermutation. Deleterious stop codons were identified using the Gene Cutter tool (Los Alamos HIV Database), which divides the contigs into the open reading frames (*gag* to *nef*) and translates them to amino acids. Any contigs containing a stop codon in any gene, excluding *nef* (Foster and Garcia, 2007), were defined as defective. Additionally, contigs containing nucleotide deletions or insertions causing frameshift mutations were identified as defective. APOBEC-induced G-A hypermutation was identified in the remaining full-length proviruses using the Los Alamos HIV Database Hypermut tool (https://www.hiv.lanl.gov) and a consensus of the corresponding participant's full-length provirus as a reference.

Defects in cis-Acting Elements

We defined contigs with defects in the *cis*-acting elements of the non-coding 5′ LTR, including the packaging signal and MSD site, as defective. These sites are essential to HIV-1 replication, as proviruses with a deletion in the four stem loops (SL1, SL2, SL3, and SL4) of the packaging signal or a point mutation in the MSD (sequence GT) have previously been identified as replication incompetent (Ho et al., 2013).

Construction of Phylogenetic Trees

Maximum-likelihood phylogenetic trees were estimated for each participant using MEGA6. The best fit model of nucleotide substitution was determined for each tree using the MEGA6 model finder tool. Models used in this study included the generalized time-reversible model, as well as Tamura and Nei, incorporating gamma-distributed (gamma category 4) and/or invariant sites where appropriate. Our heuristic tree search strategy used the nearest neighbor interchanges branch-swapping algorithm. Branch support was inferred using 1,000 bootstrap replicates. Annotated tree images were constructed using the ggtree tool (Yu et al., 2016) and including bootstrap values >75.

Statistical Analyses

Statistical analysis was performed in R: A Language and Environment for Statistical Computing, v.3.3.2 (https://www.R-project.org). As per the Poisson distribution, the average number of HIV-1-positive cells per PCR plate was estimated as $n_{pcr} = w \log(p)$, where w is the total number of wells per plate, and p is the estimated true proportion of negative wells (number of negative wells adjusted for sequenced positive wells that were not HIV-1, a mixture, or were unable to be assembled). The proportion of cells positive for HIV-1 was then calculated as n_{pcr}/n_t, where n_t is the estimated number of cells used per plate. This proportion is used to estimate the frequency of HIV-1-positive cells per 10^6 cells. A linear mixed-effects model was used to compare the estimated frequency of HIV-1-positive cells per 10^6 cells across memory subsets and adjust for potential correlation of multiple observations within each patient (library *nlme*, function *lme*). The fixed effect of memory subset was assessed with a likelihood-ratio test on maximum likelihoods. The frequency of HIV-1-positive cells was log transformed to ensure normality of residuals, and the power variance function *varPower()* was used to ensure heteroscedasticity. A linear fixed-effect model with a fixed effect for each subject was also analyzed and provided similar conclusions (function *lme*).

The proportion of sequenced HIV-1 proviruses that were intact was calculated with a mixed logistic model to account for potential correlation of multiple observations within each patient (library lme4, function glmer). A likelihood-ratio test was used to assess the fixed effect of memory subset. A fixed-effect logistic regression with a fixed effect for each subject was also analyzed and provided similar conclusions (function *glm*). To calculate the frequency of intact HIV-1 cells per 10^6 cells, the proportion of HIV-1-positive cells per 10^6 cells was multiplied by the proportion of HIV-1 intact genomes. The comparison of the FLIPS estimate of intact virus to the SPS estimate was analyzed using a paired t test on the logarithm of the ratio of estimates within each patient and memory subset. The logarithm of the ratio of estimates was approximately normally distributed.

Data and Software Availability

The accession numbers for all sequences reported in this study are GenBank: KY778264–KY778681 and KY766150–KY766212.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.09.081.

AUTHOR CONTRIBUTIONS

B.H. and J.-S.E. designed the assay; B.H. and B.A.H. conducted FLIPS on participant samples and analyzed data; R.H., S.G.D., F.M.H., E.S., J.M.M., S.v.S., L.O., N.C., and R.F. enrolled the participants and collected and/or sorted cell subsets from the participant samples; N.C. and R.F. conducted the 2-LTR and integrated DNA assays; E.L., T.L., E.A.B., D.D., S.v.S., and L.O. prepared participant samples; K.B. and J.-S.E. designed analysis and data visualization workflows; T.E.S. conducted and interpreted the statistical analysis and contributed to the review and editing of the results and methods; B.H. wrote the original manuscript; and S.P. designed the study, supervised the work performed, and edited the manuscript.

ACKNOWLEDGMENTS

This work was supported the Delaney AIDS Research Enterprise (DARE) to Find a Cure (1U19AI096109 and 1UM1AI126611-01), an amfAR Research Consortium on HIV Eradication (ARCHE) Collaborative Research grant from The Foundation for AIDS Research (amfAR 108074-50-RGRL), the Australian Centre for HIV and Hepatitis Virology Research (ACH2) (2015-43), the UCSF-GIVI Center for AIDS Research (P30 AI027763), and the Australian National Health and Medical Research Council (AAP1061681). We would like to thank Dr. Joey Lai, Genomics Facility manager at The Westmead Institute for Medical Research, for his training in Nextera XT library preparation and use of his facility and the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia) for conducting MiSeq. We acknowledge with gratitude the participants who donated samples for this study.

Received: May 21, 2017
Revised: September 1, 2017
Accepted: September 25, 2017
Published: October 17, 2017

REFERENCES

Abram, M.E., Ferris, A.L., Shao, W., Alvord, W.G., and Hughes, S.H. (2010). Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J. Virol. *84*, 9864–9878.

Barton, K., Hiener, B., Winckelmann, A., Rasmussen, T.A., Shao, W., Byth, K., Lanfear, R., Solomon, A., McMahon, J., Harrington, S., et al. (2016). Broad activation of latent HIV-1 in vivo. Nat. Commun. *7*, 12731.

Bruner, K.M., Murray, A.J., Pollack, R.A., Soliman, M.G., Laskey, S.B., Capoferri, A.A., Lai, J., Strain, M.C., Lada, S.M., Hoh, R., et al. (2016). Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat. Med. *22*, 1043–1049.

Bui, J.K., Sobolewski, M.D., Keele, B.F., Spindler, J., Musick, A., Wiegand, A., Luke, B.T., Shao, W., Hughes, S.H., Coffin, J.M., et al. (2017). Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLoS Pathog. *13*, e1006283.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.R., Ghattas, G., Brenchley, J.M., et al. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat. Med. *15*, 893–900.

Chun, T.W., Justement, J.S., Murray, D., Hallahan, C.W., Maenza, J., Collier, A.C., Sheth, P.M., Kaul, R., Ostrowski, M., Moir, S., et al. (2010). Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. AIDS *24*, 2803–2808.

Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L., McElrath, M.J., Lehmann, C., et al. (2015). HIV-1 integration landscape during latent and active infection. Cell *160*, 420–432.

Davey, R.T., Jr., Bhat, N., Yoder, C., Chun, T.W., Metcalf, J.A., Dewar, R., Natarajan, V., Lempicki, R.A., Adelsberger, J.W., Miller, K.D., et al. (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. Proc. Natl. Acad. Sci. USA *96*, 15109–15114.

Evering, T.H., Mehandru, S., Racz, P., Tenner-Racz, K., Poles, M.A., Figueroa, A., Mohri, H., and Markowitz, M. (2012). Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. PLoS Pathog. *8*, e1002506.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science *278*, 1295–1300.

Foster, J.L., and Garcia, J.V. (2007). Role of Nef in HIV-1 replication and pathogenesis. Adv. Pharmacol. *55*, 389–409.

Geginat, J., Sallusto, F., and Lanzavecchia, A. (2003). Cytokine-driven proliferation and differentiation of human naïve, central memory and effector memory CD4+ T cells. Pathol. Biol. (Paris) *51*, 64–66.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803–809.

Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell *155*, 540–551.

Hosmane, N.N., Kwon, K.J., Bruner, K.M., Capoferri, A.A., Beg, S., Rosenbloom, D.I., Keele, B.F., Ho, Y.C., Siliciano, J.D., and Siliciano, R.F. (2017).

Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J. Exp. Med. *214*, 959–972.

Imamichi, H., Dewar, R.L., Adelsberger, J.W., Rehm, C.A., O'Doherty, U., Paxinos, E.E., Fauci, A.S., and Lane, H.C. (2016). Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. Proc. Natl. Acad. Sci. USA *113*, 8783–8788.

Josefsson, L., King, M.S., Makitalo, B., Brännström, J., Shao, W., Maldarelli, F., Kearney, M.F., Hu, W.S., Chen, J., Gaines, H., et al. (2011). Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. Proc. Natl. Acad. Sci. USA *108*, 11199–11204.

Josefsson, L., Eriksson, S., Sinclair, E., Ho, T., Killian, M., Epling, L., Shao, W., Lewis, B., Bacchetti, P., Loeb, L., et al. (2012). Hematopoietic precursor cells isolated from patients on long-term suppressive HIV therapy did not contain HIV-1 DNA. J. Infect. Dis. *206*, 28–34.

Josefsson, L., Palmer, S., Faria, N.R., Lemey, P., Casazza, J., Ambrozak, D., Kearney, M., Shao, W., Kottilil, S., Sneller, M., et al. (2013a). Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. PLoS Pathog. *9*, e1003432.

Josefsson, L., von Stockenstrom, S., Faria, N.R., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Tan, A., Ho, T., Lemey, P., et al. (2013b). The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. Proc. Natl. Acad. Sci. USA *110*, E4987–E4996.

Kearney, M., Palmer, S., Maldarelli, F., Shao, W., Polis, M.A., Mican, J., Rock-Kress, D., Margolick, J.B., Coffin, J.M., and Mellors, J.W. (2008). Frequent polymorphism at drug resistance sites in HIV-1 protease and reverse transcriptase. AIDS *22*, 497–501.

Laskey, S.B., Pohlmeyer, C.W., Bruner, K.M., and Siliciano, R.F. (2016). Evaluating clonal expansion of HIV-infected cells: optimization of PCR strategies to predict clonality. PLoS Pathog. *12*, e1005689.

Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A.J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science *300*, 1112.

Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.H., Hua, S., Chen, H.R., Ouyang, Z., et al. (2017). Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J. Clin. Invest. *127*, 2689–2696.

Lewinski, M.K., Bisgrove, D., Shinn, P., Chen, H., Hoffmann, C., Hannenhalli, S., Verdin, E., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2005). Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J. Virol. *79*, 6610–6619.

Li, B., Gladden, A.D., Altfeld, M., Kaldor, J.M., Cooper, D.A., Kelleher, A.D., and Allen, T.M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J. Virol. *81*, 193–201.

Lorenzi, J.C., Cohen, Y.Z., Cohn, L.B., Kreider, E.F., Barton, J.P., Learn, G.H., Oliveira, T., Lavine, C.L., Horwitz, J.A., Settler, A., et al. (2016). Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. Proc. Natl. Acad. Sci. USA *113*, E7908–E7916.

Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science *345*, 179–183.

Palella, F.J., Jr., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., Aschman, D.J., and Holmberg, S.D.; HIV Outpatient Study Investigators (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N. Engl. J. Med. *338*, 853–860.

Palmer, S., Kearney, M., Maldarelli, F., Halvas, E.K., Bixby, C.J., Bazmi, H., Rock, D., Falloon, J., Davey, R.T., Jr., Dewar, R.L., et al. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J. Clin. Microbiol. *43*, 406–413.

Palmer, S., Maldarelli, F., Wiegand, A., Bernstein, B., Hanna, G.J., Brun, S.C., Kempf, D.J., Mellors, J.W., Coffin, J.M., and King, M.S. (2008). Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. Proc. Natl. Acad. Sci. USA *105*, 3879–3884.

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

Sallusto, F., Geginat, J., and Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: function, generation, and maintenance. Annu. Rev. Immunol. *22*, 745–763.

Simonetti, F.R., Sobolewski, M.D., Fyne, E., Shao, W., Spindler, J., Hattori, J., Anderson, E.M., Watters, S.A., Hill, S., Wu, X., et al. (2016). Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. Proc. Natl. Acad. Sci. USA *113*, 1883–1888.

Soriano-Sarabia, N., Bateson, R.E., Dahl, N.P., Crooks, A.M., Kuruc, J.D., Margolis, D.M., and Archin, N.M. (2014). Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. J. Virol. *88*, 14070–14077.

Vandergeeten, C., Fromentin, R., Merlini, E., Lawani, M.B., DaFonseca, S., Bakeman, W., McNulty, A., Ramgopal, M., Michael, N., Kim, J.H., et al. (2014). Cross-clade ultrasensitive PCR-based assays to measure HIV persistence in large-cohort studies. J. Virol. *88*, 12385–12396.

von Stockenstrom, S., Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Shao, W., Hoh, R., Ho, T., et al. (2015). Longitudinal genetic characterization reveals that cell proliferation maintains a persistent HIV type 1 DNA pool during effective HIV therapy. J. Infect. Dis. *212*, 596–607.

Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science *345*, 570–573.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2016). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36.