



HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2022 December 13.

Published in final edited form as:

Cell Rep. 2022 April 26; 39(4): 110739. doi:10.1016/j.celrep.2022.110739.

Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia before and during analytical treatment interruption

Basiel Cole^{1,10}, Laurens Lambrechts^{1,2,10}, Zoe Boyer³, Ytse Noppe¹, Marie-Angélique De Scheerder⁴, John-Sebastian Eden³, Bram Vrancken⁵, Timothy E. Schlub⁶, Sherry McLaughlin⁷, Lisa M. Frenkel^{7,8}, Sarah Palmer^{3,11}, Linos Vandekerckhove^{1,4,9,11,*}

¹HIV Cure Research Center, Department of Internal Medicine and Pediatrics, Ghent University Hospital, Ghent University, 9000 Ghent, Belgium

²BioBix, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, 9000 Ghent, Belgium

³Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney 2145, NSW, Australia

⁴Department of General Internal Medicine and Infectious Diseases, Ghent University Hospital, Corneel Heymanslaan 10, 9000 Ghent, Belgium

⁵Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory of Evolutionary and Computational Virology, KU Leuven-University of Leuven, 3000 Leuven, Belgium

⁶University of Sydney, Faculty of Medicine and Health, Sydney School of Public Health, Sydney 2000, NSW, Australia

⁷Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA

⁸Departments of Global Health, Laboratory Medicine, Medicine, and Pediatrics, University of Washington, Seattle, WA, USA

⁹Lead contact

¹⁰These authors contributed equally

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence: linos.vandekerckhove@ugent.be.

AUTHOR CONTRIBUTIONS

B.C., L.L., L.F., S.P., and L.V. conceptualized the experiments. M.A.D.S. processed the samples from the initial HIV-STAR study, including cell isolation from peripheral blood and tissue, and performed cell sorting and single-genome sequencing. B.C. and Y.N. performed experiments involving cell sorting, multiple displacement amplification, single-genome sequencing, and integration site sequencing. L.L. and Z.B. performed experiments involving near full-length proviral sequencing. B.C., L.L., B.V., J.S.E., and T.S. analyzed data and performed associated analyses. B.C., L.L., T.S., and B.V. made figures and tables. B.C. and L.L. wrote the manuscript. All co-authors edited and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare that no conflict of interest exists.

SUPPLEMENTAL INFORMATION

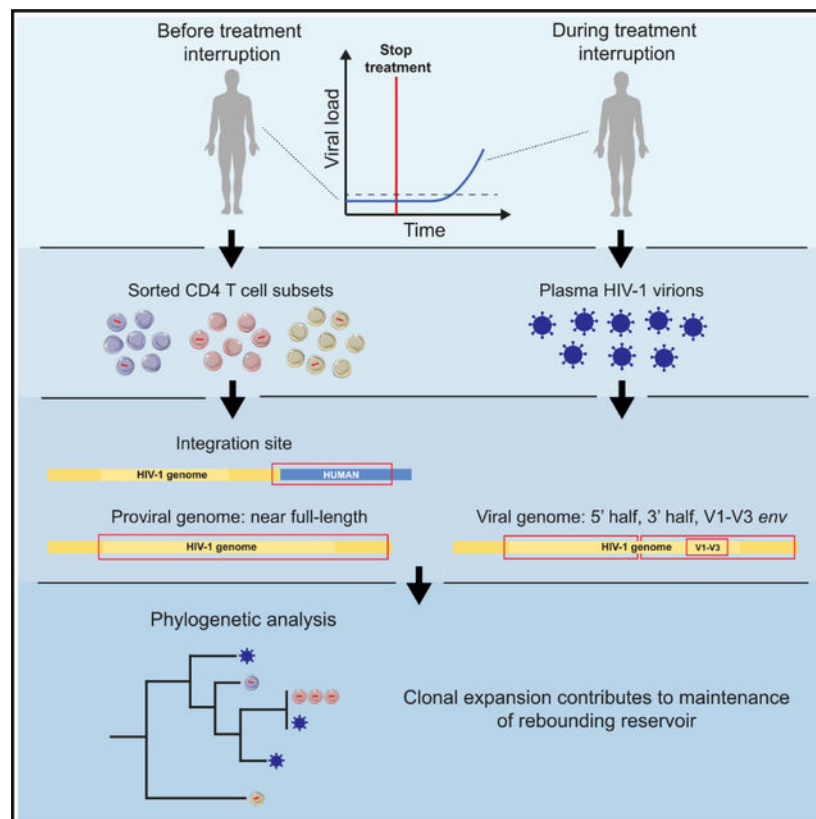
Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2022.110739>.

¹¹These authors contributed equally

SUMMARY

The HIV-1 reservoir is composed of cells harboring latent proviruses that have the potential to contribute to viremia upon antiretroviral treatment (ART) interruption. While this reservoir is known to be maintained by clonal expansion of infected cells, the contribution of these cell clones to residual viremia and viral rebound remains underexplored. Here, we conducted an extensive analysis on four ART-treated individuals who underwent an analytical treatment interruption (ATI), characterizing the proviral genomes and associated integration sites of large infected clones and phylogenetically linking these to plasma viremia. We show discrepancies between different assays in their ability to assess clonal expansion. Furthermore, we demonstrate that proviruses could phylogenetically be linked to plasma virus obtained before or during an ATI. This study highlights a role for HIV-infected cell clones in the maintenance of the replication-competent reservoir and suggests that infected cell clones can directly contribute to rebound viremia upon ATI.

Graphical Abstract



In brief

Cole, Lambrechts, et al. study the proviral HIV-1 landscape in four participants undergoing an antiretroviral treatment interruption. Matches between plasma viruses recovered during and before

the interruption and proviruses found in cell clones highlight a role for clonal expansion in the maintenance of the clinically relevant viral reservoir.

INTRODUCTION

HIV-1 infection remains incurable because of the presence of a persistent viral reservoir, capable of contributing to viral rebound upon treatment interruption (TI) (Chun and Fauci, 1999; Chun et al., 1997, 1998; Finzi et al., 1997). Despite efforts to better understand the dynamics and persistence of the HIV-1 viral reservoir, pinpointing the origins of rebounding viruses remains elusive (De Scheerder et al., 2019). Previously, it was shown that infected CD4 T cells can undergo clonal expansion, contributing to the long-term persistence of the HIV-1 viral reservoir during antiretroviral therapy (ART) (Boritz et al., 2016; Cohn et al., 2015; Einkauf et al., 2019; Hosmane et al., 2017; Maldarelli et al., 2014; Salantes et al., 2018; Simonetti et al., 2016; Wagner et al., 2014; Wang et al., 2018). The observation that low-level viremia (LLV) during ART (Aamer et al., 2020; Bailey et al., 2006; Brennan et al., 2009; Halvas et al., 2020; Tobin et al., 2005; Wagner et al., 2013) and rebound viremia upon TI (Aamer et al., 2020; Kearney et al., 2016; Lu et al., 2018; De Scheerder et al., 2019) often consist of monotypic populations of viruses, suggests that HIV-1-infected cell clones are potentially key contributors to refueling viremia during TI. Clonality of infected cells has historically been demonstrated by recovering identical proviral sequences or identical integration sites (IS) in multiple cells (Cohn et al., 2015; Hiener et al., 2017; Lee et al., 2017; Maldarelli et al., 2014; Pinzone et al., 2019; Von Stockenstrom et al., 2015; Wagner et al., 2014). While the former method allows for qualitative assessment of the proviral genome, it is often not adequate to confidently predict clonal expansion of HIV-1-infected cells, especially when evaluating a short subgenomic region (Lambrechts et al., 2020; Laskey et al., 2016). On the other hand, analysis of IS provides direct proof of clonal expansion, though it typically leaves the proviral sequence uncharacterized. Recently, several techniques to link near full-length (NFL) proviral sequences to IS were developed (Artesi et al., 2021; Einkauf et al., 2019; Patro et al., 2019). These assays combine the qualitative strength of NFL HIV-1 sequencing with IS analysis, shedding light on the integration profile of genetically intact versus defective proviruses.

Analytical treatment interruption (ATI) studies allow for the investigation of the dynamics and genetic makeup of rebounding viruses (Clarridge et al., 2018; Garner et al., 2017; Kearney et al., 2016; Pannus et al., 2020). To identify the source of rebounding viruses, we previously conducted the HIV-STAR (HIV-1 sequencing before ATI to identify the anatomically relevant HIV reservoir) study (De Scheerder et al., 2019). During this study, in-depth sampling was performed on 11 chronically treated HIV-1-infected participants prior to ATI. Cells were isolated from different anatomical compartments and sorted into several CD4 T cell subsets. Subgenomic proviral sequences (V1–V3 region of *env*) were recovered and phylogenetically linked to sequences from rebounding plasma virus collected during different stages of the ATI. This study suggested that HIV-1 rebound is predominantly fueled by genetically identical viral expansions, highlighting the potentially important role of clonal expansion in the maintenance of HIV-1 proviruses with the potential to rebound. While this study yielded a total of 4,329 V1–V3 *env* sequences from peripheral

blood mononuclear cells (PBMCs), lymph node (LN), and gut-associated lymphoid tissue (GALT), enabling a detailed investigation of the viral reservoir and its relation to rebound viremia, it left some questions unanswered. Most importantly, the evaluation of a short subgenomic region (V1–V3 *env*) to link proviral sequences to rebounding plasma virus made it impossible to investigate the entire genome structure of proviruses linked to rebound. Furthermore, the lack of IS analysis did not allow for the study of the chromosomal location of the rebounding proviruses.

To address these points, we performed a combination of multiple displacement amplification (MDA), IS analysis, and NFL proviral sequencing on four participants that were enrolled in the HIV-STAR study, with special attention to clonally expanded HIV-1 infected cells. We demonstrate that HIV-1 proviral sequences and corresponding IS of clonally expanded infected cells could be retrieved, and in rare cases, these could be linked to LLV during ART and rebound viremia upon ATI, further highlighting the role of clonal expansion in the maintenance of HIV-1 proviruses with the potential to rebound.

RESULTS

Experimental setup

To investigate the genetic composition and chromosomal location of proviruses within clonally expanded cells and their relationship to rebound viremia, several qualitative assays were performed on samples from four chronically treated HIV-1-infected individuals undergoing an ATI (Figure 1A; Table S1). Samples from these individuals were obtained longitudinally before (T1) and during the ATI (T2, T3, T4), as summarized in Figure 1B.

First, the overall landscape of HIV-1-infected cell clones prior to ATI (T1; Figure 1B) was determined by subgenomic single-genome sequencing (SGS) of V1–V3 *env* and full-length individual proviral sequencing (FLIPS) and by integration site loop amplification (ISLA) (Figures 1A; Table S1). This yielded three datasets that were used independently to identify potential clonally expanded infected cell populations.

Next, to identify both the NFL proviral genome and associated IS from these infected clones, MDA was performed at limiting dilution on sorted cell lysates from peripheral blood obtained before ATI (T1). MDA wells were subjected to V1–V3 *env* SGS and ISLA, and wells that yielded a V1–V3 *env* sequence and/or an IS corresponding to a suspected cellular clone were further investigated. Potential clonality was determined by either (1) an exact link to ISLA/FLIPS/SGS data generated in the first step or (2) by identical V1–V3 *env* sequences and/or IS shared between several MDA wells. The NFL genomes of the proviruses in these selected MDA wells were sequenced and mapped back to NFL FLIPS sequences and historic V1–V3 *env* SGS proviral sequences from specimens collected prior to ATI (T1; Figure 1B), as well as V1–V3 *env* plasma-derived RNA sequences retrieved prior to (T1) and during the ATI (T2–T4, Figure 1B). In addition, the remaining T4 plasma samples from all four participants were subjected to 5′- and –3′-half genome amplification to complement existing V1–V3 *env* plasma SGS data.

The generated datasets allowed for the assessment of the genetic structure and IS of proviruses in clonally expanded infected cells, their placement across cellular subsets, and their contribution to residual viremia on ART and rebound viremia during an ATI.

Integration site analysis and NFL proviral sequencing

ISLA was performed on endpoint-diluted non-amplified cell lysate and on MDA-amplified cell lysates of central memory/transitional memory (TCM/TTM) and effector memory (TEM) CD4 T cell subsets from peripheral blood (T1) from three of the four participants: STAR 9, STAR 10, and STAR 11 (Figure 2A; Tables S2 and S3). A total of 328 IS (171 TCM/TTM, 157 TEM) were recovered across these participants, with 42% (139/328) belonging to clonally expanded infected cells. Analysis of IS revealed a significantly higher degree of clonally expanded HIV-1-infected cells in the TEM subset (mean 65%) compared to the TCM/TTM fraction (mean 22%) from the peripheral blood ($p < 0.001$ for STAR 9 and STAR 11; $p < 0.05$ for STAR 10), as previously reported (Hiener et al., 2017). Identical IS between different subsets, suggestive of linear differentiation from an originally infected TCM/TTM into a TEM cell, were observed in rare instances, as eight of the 35 different IS found in infected cell clones (23%) were recovered in both subsets.

NFL proviral sequencing using the FLIPS protocol (spanning 92% of the proviral genome) was performed for all four participants on the TCM/TTM and TEM subsets in the peripheral blood collected prior to ATI (T1) (Figure 2B). In addition, based on sample availability, CD45+ cells from the GALT and other cell subsets from the peripheral blood and LN were assayed with FLIPS, as listed in Table S4. Across all participants, 364 NFL sequences were isolated from the peripheral blood (TCM/TTM and TEM subsets), of which 27 (7%) were genetically intact, while the majority (66%) contained large internal deletions (Figure 2B). Pooling NFL sequences isolated from all blood and tissues yielded a total number of 479 NFL sequences with a mean of 120 sequences per participant (Figures S1 and S2A; Table S4). The genetically intact NFL sequences were primarily found within the TEM fraction (Figures S1 and S2A–S2D). In addition, the overall HIV-1 infection frequency in the peripheral blood was significantly higher in the TEM fraction compared to the TCM/TTM fraction ($p < 0.001$), except for participant STAR 4 (Figure S2C). Expansions of identical sequences (EIS), suggestive of clonal expansion of infected cells, were observed in the peripheral blood from all participants (Figure 2C). In accordance with the ISLA results, EIS were more frequent in the TEM fraction compared to the TCM/TTM fraction ($p < 0.05$ for STAR 4; $p = 0.23$ for STAR 9; $p = 0.061$ for STAR 10; $p < 0.001$ for STAR 11; Figure 2C).

Multiple displacement amplification-mediated characterization of infected cell clones

MDA-mediated HIV-1 provirus and IS sequencing offers the unique opportunity of linking NFL proviral sequences to their precise location in a chromosome. Starting from MDA wells that presumably contained a provirus belonging to an infected clone (as defined in the experimental setup), we identified in total 93 IS (37 different). From those wells with a successfully amplified IS, we managed to recover 15 complete NFL sequences (12 different). Among those sequences, we detected 12 HIV-infected cell clones for which both the complete proviral NFL sequence and IS could be retrieved as represented in Figure 3A

(Table S3). In total, eight clonal cell populations with a defective provirus were identified: one with hypermutation, one with a frameshift, and six with a defect in the packaging signal and/or major splice donor site (Figure 3A). Clones with a putatively intact proviral genome were less prevalent (4/12), three of which were retrieved in STAR 11 (Figure 3A). Interestingly, five out of eight defective clones were found in both the TCM/TTM and the TEM fractions, which suggests linear differentiation of clonal cell populations (Figures 3B–3D). In contrast, all the clones with genetically intact proviruses were found exclusively in the TEM fraction (Figures 3C and 3D).

Three out of four genetically intact proviruses were integrated into a gene of the zinc finger nuclease (ZNF) family (*ZNF141*, *GGNBP2*, *ZNF274*; Figure 3A), which have recently been identified as integration hotspots for intact proviruses in expanded clones (Huang et al., 2021). Moreover, we identified an expanded clone with an IS in *STAT5B*, harboring a provirus with a 25-bp deletion in the packaging signal (Figure 3A). Previous work has shown that proviral integration into the first intron of this gene can lead to HIV LTR-driven dysregulation of *STAT5B*, resulting in cellular proliferation (Cesana et al., 2017).

We conclude that a large fraction of the identified clonally expanded infected cell populations harbored defective proviruses and were frequently found across different cellular subsets, suggesting differentiation of infected cell populations.

Large discrepancies between suspected clonal HIV-1-infected cell populations identified with ISLA, SGS, and FLIPS

ISLA, SGS, and FLIPS can independently be used to assess clonality of infected cells, the former based on the IS and the two latter on a subgenomic (V1–V3 *env*) or NFL sequence of the proviral genome. In Figure 4, these three individual datasets are represented, and unambiguous links between these assays, as identified by MDA, are displayed in colors (see STAR Methods for additional information). Clonal populations that were only retrieved in one of the datasets, with no evident link to the other datasets, are represented in gray. In addition, both the nucleotide diversity of the V1–V3 *env* region and the clonal prediction score (CPS, capacity of the V1–V3 *env* region to distinguish between different NFL proviruses) were calculated for each participant (Laskey et al., 2016) (Table S5).

Three distinct sources of discrepancy and/or bias in their ability to detect specific clones could be identified. First, we found two instances of presumably clonal EIS as detected by SGS that were linked to multiple distinct NFL sequences (Figure 4; STAR 4, STAR 10, indicated with green and blue arrows). This shows that the V1–V3 *env* region does not always differentiate between distinct proviral genomes, resulting in a CPS lower than 100% (Table S5). In accordance with this, STAR 4 and STAR 10 displayed lower nucleotide diversity (0.010 and 0.014, respectively) in the V1–V3 *env* region compared to STAR 9 and STAR 11 (0.017 and 0.016, respectively), which could potentially explain these results (Table S5).

Second, a clear case of primer bias could be observed in STAR 9 (Figure 4). One major clone, with a hypermutated provirus integrated at an intergenic region on chromosome 11, was detected with both MDA and FLIPS. However, this provirus was never detected with

SGS, which can be explained by the fact that V1–V3 *env* primers did not anneal to this hypermutated sequence due to excessive mismatches (total of 10 mismatches over all four primers).

Third, STAR 4, 9, and 10 displayed only limited overlaps between V1–V3 *env* sequences recovered by SGS and FLIPS sequences, which is the result of an inherent sampling bias due to deletions. Indeed, the FLIPS assay frequently amplified proviruses with internal deletions covering the V1–V3 *env* region, which cannot be amplified by the SGS assay. As a result, there is clear discrepancy between the proviral populations identified by V1–V3 *env* SGS and the ones identified by FLIPS (Figure 4, clones in dark gray). In addition, the largest clone within STAR 11 based on ISLA data, with an IS in the *ZFC3H1* gene (Figure 2A), could not be linked to SGS and FLIPS data, which was probably the result of a large internal deletion spanning the majority of the proviral genome. Indeed, a total of seven different subgenomic PCRs, together spanning the NFL proviral genome, consistently failed to amplify this provirus (Table S3).

In conclusion, we demonstrate that the choice of assay to assess clonal expansion of infected cells dictates the outcome, driven by three non-mutually exclusive phenomena: lack of genomic variability within a subgenomic region, polymorphisms, and deletions.

Plasma viral sequences match intact proviruses and proviruses with large internal deletions or defects in the packaging signal

We next investigated the genomic structure and matched IS of proviruses that could phylogenetically be linked to plasma viremia before (T1) and during the ATI (T2–T4). Phylogenetic trees were made using the V1–V3 *env* region obtained from NFL DNA sequences (T1), SGS plasma sequences (T1–T4), and the 3′-half plasma sequences (T4) (Figure 5). In addition, we constructed phylogenetic trees based on alignments of the NFL DNA sequences (T1) and either 3′- or 5′-half plasma sequences (T4) (Figure S3). These latter alignments were checked for sequence recombination, and no evidence for such events was observed. Overall, a total of eight V1–V3 *env* regions from proviral sequences matched plasma sequences, suggestive of links to residual or LLV ($n = 3$, matching T1–T2 plasma) or rebound viremia ($n = 5$, matching T3–T4 plasma), as indicated with arrows in Figure 5.

The three matching V1–V3 *env* regions between proviruses and plasma viruses retrieved at T1 (before ATI) or at T2 (during the ATI but before a detectable viral load) were all found in STAR 11. The first match was between a defective provirus with a frameshift (IS at chr17:7545670, found in both blood TCM/TTM and TEM subsets) and a plasma virus at T1 (Figure 5D). The other matches were found (1) between a genetically intact provirus (blood TCM/TTM subset) and a T1 plasma virus and (2) between an intact provirus (blood TEM subset) and a T2 plasma virus, respectively (Figure 5D). The latter provirus was found to be integrated in the *ZNF141* gene, which belongs to the Krüppel-associated box domain (KRAB)-containing zinc finger nuclease family. Interestingly, this viral sequence was not detected in the plasma from rebounding time points (T3, T4), suggesting that it might have been obscured by outgrowth of more fit variants at these time points (Figures 5D and S3D).

The five remaining matches were observed between proviruses and plasma viruses obtained during the rebound time points (T3, T4). Interestingly, only one of these matches involved a genetically intact proviral genome found in STAR 9 (Figures 5A–5C). This intact provirus (blood TEM subset) matched three out of four plasma sequences found at T2 prior to rebound but only matched one sequence at T4 during rebound. For participants STAR 10 and STAR 4, one or more plasma sequences obtained during rebound (T3, T4) were identical to defective proviral genomes (Figures 5A and 5B). STAR 10 displayed three such matches (Figure 5A): (1) one plasma sequence matched a unique provirus with a large internal deletion (blood TCM/TTM subset), (2) one plasma sequence matched a provirus with a packaging signal (PSI)/major splice donor (MSD) deletion belonging to an infected cell clone (IS at chr8:100792121, blood TEM subset), and (3) one plasma sequence matched a provirus with a probable deletion (IS in *ZBTB20*, blood TEM subset). In STAR 4, a provirus from the GALT with a PSI/MSD deletion matched plasma V1–V3 *env* sequences from all three ATI time points (T2–T4) and two different sets of 3′-half genomes (T4) with an identical V1–V3 *env* (Figure 5B). When looking at extended alignments based on either 5′- or 3′-half T4 plasma genomes and NFL proviral sequences, this GALT provirus and some of the plasma sequences seem to be closely related (Figure S3A, arrows). In fact, the proviral sequence and the 5′-half plasma sequence were identical besides a 105-bp deletion in the PSI/MSD region, while the two sets of 3′ plasma genomes matching the V1–V3 *env* were either 8 bp or 19 bp different from the proviral sequence, suggesting a common ancestor.

In conclusion, by performing MDA-mediated NFL and IS analysis, we identified several proviruses with their linked IS, predominantly belonging to peripheral blood CD4 memory subsets, that matched sequences from plasma before and/or during an ATI. Interestingly, most of these proviruses were classified as defective. Future studies will be needed to define whether these defective proviruses are still capable of producing viremia or whether the observed matches are the result of insufficient power of the V1–V3 *env* region to distinguish viral sequences, despite the high CPS.

DISCUSSION

Integration of HIV-1 genomes into the DNA of host cells leads to the establishment of a persistent HIV-1 reservoir. While most of these integrated proviruses are defective, a small proportion are genetically intact and capable of producing infectious virions (Hiener et al., 2017). The number of genetically intact HIV-1 proviruses has been shown to decay slowly, with an estimated half-life of 4 years during the first 7 years of suppression and 18.7 years thereafter (Peluso et al., 2020). This long half-life can in part be explained by continuous clonal expansion of infected cells harboring these genetically intact HIV-1 proviruses (Liu et al., 2020; Patro et al., 2019). While this phenomenon is well established, the contribution of clonally expanded HIV-1-infected cells to residual viremia under ART and rebound viremia upon ATI remains underexplored. Here, we used a combination of NFL sequencing, IS analysis, and MDA-mediated IS/NFL sequencing to define the source of rebounding virus detected during an ATI.

We first showed large discrepancies between different techniques to assess clonal expansion of HIV-1 infected cells. We identified at least three distinct sources of discrepancy/bias, driven either by the lack of genomic variability within a subgenomic region or by extensive polymorphisms leading to primer mismatches and/or deletions, preventing amplification of certain proviruses. Indeed, we demonstrated that the use of the relatively variable short subgenomic region of the HIV-1 genome (V1–V3 *env*) to assess clonality of infected cells can lead to inaccurate results. This was shown by the recovery of distinct NFL proviruses, integrated at different sites, displaying identical V1–V3 *env* sequences. Similar observations were made in a recently published study, where P6-PR-RT sequences were compared to matched NFL/IS sequences (Patro et al., 2019). They found multiple instances of identical proviral P6-PR-RT sequences, a region known to be less variable, with distinct IS. Primer binding issues due to polymorphisms and/or deletions are an important source of bias in HIV-1 molecular research because of the high variability of the HIV genome (Lambrechts et al., 2020). While this problem can partially be overcome by using subtype-specific or even participant-specific primer sets, it is unlikely that these will be able to account for all genomic variabilities or deletions. In conclusion, we demonstrate that the use of subgenomic regions to assess clonality should be done with caution, and researchers should be aware of several biases introduced by genomic variability.

We next set out to find links between NFL proviral sequences and RNA sequences from the plasma during different stages of an ATI. First, we identified several matches between defective NFL proviruses and plasma V1–V3 *env* sequences. Some of these proviruses harbored small deletions in the 5' UTR region, covering the MSD/PSI region. For example, in participant STAR 4, we found a match between a provirus containing a 105-bp deletion in the PSI region (including stem loop 1 and 2) and plasma sequences recovered during rebound. It has been shown that proviruses with such defects are still capable of producing infectious virions, though with significantly lower efficiency (Pollack et al., 2017). We and others previously showed that proviruses with PSI/MSD defects are capable of producing viral proteins, highlighting their clinical relevance (Cole et al., 2021; Imamichi et al., 2020; Pollack et al., 2017). In order to discern whether these viruses could truly contribute to viral rebound, follow-up experiments to assess their replication fitness will have to be conducted.

Three other defective proviruses linked to rebound viruses, all from participant STAR 10, contained large internal deletions, making it unlikely that these are the actual source of the virus rebounding during an ATI. Rather, these are probably phylogenetically related proviruses, as they share an identical V1–V3 *env* sequence. Two previous studies that tried to link proviral sequences to rebound plasma sequences, based on full *env* sequences, concluded that while they were not able to directly link the proviral sequences to the rebounding ones, the rebounding sequences could often be accounted for by recombination (Cohen et al., 2018; Vibholm et al., 2019). Because we assessed only a small portion of the *env* gene (V1–V3 region) at time points T1–T3, we were not able to comprehensively study recombination events, though we hypothesize that recombination may be a probable cause of identical overlap between defective proviral sequences and rebounding virus sequences. At T4, we recovered half-genome plasma sequences, though these did not show any sign of recombination. Interestingly, in the 5'-half genome plasma dataset (T4), two plasma sequences matched a proviral genome besides a 105-bp deletion observed in the proviral

DNA sequence. While this suggests a close phylogenetic relationship, the shared common ancestor with intact PSI driving the rebound of this plasma virus remained unsampled.

We further identified two links between genetically intact NFL proviruses and plasma viruses emerging upon ATI. On both occasions, the intact provirus was located in the peripheral blood TEM subset, suggesting that these might be easier to reactivate because of higher activation status. Alternatively, this could reflect the higher degree of clonality observed in the TEM subset compared to other memory subsets, which in turn increases the chances of detecting links. The first link was found in participant STAR 9, where an intact provirus obtained with FLIPS could be linked to plasma virus at T2 and T4. Interestingly, this virus was first sampled at T2 and persisted into T4, which suggests that it emerged during the phase of the ATI when the viral load was still undetectable. In participant STAR 11, an intact provirus integrated in the *ZNF141* gene could be linked to plasma virus at T2 during an ATI. Another recent publication found a clonal infected cell population with IS in the *ZNF721/ABCA11P* gene, which contributed to persistent residual viremia that was not suppressed by ART (Halvas et al., 2020). This *ZNF721/ABCA11P* gene is located on chromosome 4 and belongs to the KRAB-containing zinc finger nuclease (ZNF) family. This integration event shows great similarities with the provirus we identified in the *ZNF141* gene, which also belongs to the KRAB-ZNF family and which is located on chromosome 4, just upstream of the *ZNF721/ABCA11P* gene. Interestingly, three other studies also described infected cell clones harboring a genetically intact provirus integrated in the *ZNF721/ABCA11P* gene, suggesting that this region is a particular hotspot for the persistence of genetically intact proviruses (Einkauf et al., 2019; Halvas et al., 2020; Jiang et al., 2020). Because the plasma virus that was linked to the *ZNF141* clone stems from T2 (latest time point with undetectable viral load during the ATI) but did not persist in the later time points (T3 and T4), we cannot exclude that the virus we sampled emerged as a result of continuous virus shedding, as described by Aamer et al. (Aamer et al., 2020), rather than “true” rebounding virus. Alternatively, it is possible that this viral strain was not identified at T3 and T4 because it was obscured by other rebound viruses so not included among the variants we sequenced. Of note, we also found a clone in the peripheral blood TEM fraction (STAR 11) harboring an intact virus integrated into the *ZNF274* gene, which is also a member of the KRAB-ZNF gene family. A previous study on elite controllers identified a large clone harboring an intact provirus integrated into this gene and showed that this genomic region is associated with dense chromatin (Jiang et al., 2020). Despite the rather large size of the clone, we did not observe the emergence of the corresponding viral sequence in the plasma during the ATI, which is in agreement with its presumed “deep latent” state. These observations add to the emerging body of evidence that intact proviruses in clonally expanded CD4 T cells are preferentially integrated into genes of the KRAB-ZNF family, as recently described by Huang et al. (2021).

In conclusion, our data show that reservoir characterization using multiple methods, including IS analysis, NFL proviral sequencing, and a combination of both, enables the identification of matches between proviral sequences and plasma sequences recovered before and/or during an ATI, but these matches are rare. While our findings confirm that expanded HIV-infected cell clones present in the peripheral blood can contribute to both residual and rebound plasma viremia, the origins of a large fraction of rebounding

viruses remain unknown. Future studies should focus on in-depth characterization of tissue reservoirs to further investigate their relative contribution to rebound viremia.

Limitations of the study

We acknowledge several limitations in this study. The first is the limited sampling from tissue compartments, possibly causing us to miss important rebound lineages. Indeed, it has been shown that tissues, including LNs and GALT, harbor most of the HIV-1 latent reservoir, orders of magnitude higher than the peripheral blood compartment (Estes et al., 2017). Whether there is compartmentalization between different anatomical compartments is under debate. Several studies, including our previously conducted HIV-STAR study, have suggested that there is limited compartmentalization between the HIV-1 proviral sequences recovered from LNs and from peripheral blood (Josefsson et al., 2013; Mcmanus et al., 2019; De Scheerder et al., 2019; Von Stockenstrom et al., 2015; Vibholm et al., 2019) based on identical proviral sequences and/or IS shared between both compartments. In addition, our previous HIV-STAR study did not show evidence of any enrichment of rebounding sequences stemming from specific anatomical compartments (De Scheerder et al., 2019), justifying our decision to focus the current study primarily on the peripheral blood compartment. The second limitation of the current study is that the link to plasma sequences at T1–T3 is based on the V1–V3 *env* region rather than on NFL plasma sequences. This means that we cannot exclude the possibility that links between proviral sequences and rebounding plasma sequences are the result of matches in V1–V3 *env* but with genetic variation outside of this region, but the CPS for the V1–V3 *env* region for participants STAR 9 and STAR 11 was calculated at 100%. Furthermore, the lack of matching sequences between half-genome plasma sequences at T4 and proviral sequences from T1 might be because the plasma at T4 was dominated by a genetically oligoclonal pool of viruses, which might have obscured less fit rebounding viruses that match T1 proviruses. Furthermore, a study by Bertagnolli et al. showed that the outgrowth of a substantial fraction of viruses of the latent reservoir is blocked by autologous IgG antibodies against HIV-1 *envelope* (Bertagnolli et al., 2020). This mechanism might explain the discrepancy between proviruses recovered *ex vivo* and viruses recovered from the plasma. Indeed, the population of viruses that rebound might have been shaped by immune pressure, which is absent when assessing proviral sequences recovered from extracted DNA. This phenomenon further complicates finding links between proviruses and plasma viruses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Linos Vandekerckhove (linos.vandekerckhove@ugent.be).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- HIV-1 proviral sequence data have been deposited at GenBank and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- HIV-1 proviral integration site data are listed in Table S2.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of four HIV-1 infected, ART treated adult male participants were included in this study. All had an undetectable viral load (<20 copies/mL) for at least 1 year prior to treatment interruption, and all initiated ART during the chronic phase of infection. The participants characteristics are summarized in Table S6. Participants were sampled longitudinally, prior to and during an ATI (Figure 1B). Anatomical compartments that were sampled, and corresponding cell subsets sorted from these, are summarized in Table S1. This study was approved by the Ethics Committee of the Ghent University Hospital (Belgian registration number: B670201525474). Written informed consent was obtained from all study participants.

METHOD DETAILS

Sample processing and sorting—Cryopreserved PBMCs were thawed and CD4 T cell enrichment was carried out with negative magnet-activated cell sorting (Beckton Dickinson, BD IMag, #557939). CD4 T cells were stained with the following monoclonal antibodies: CD3 (Beckton Dickinson, #564465), CD8 (Beckton Dickinson, #557746), CD45RA (Beckton Dickinson, #550855), CD45RO (Beckton Dickinson, #555493), CD27 (Beckton Dickinson, #561400), CCR7 (Beckton Dickinson, #560765), and a fixable viability stain (Beckton Dickinson, #565388). In addition, the GALT cells were stained with the CD45 (Beckton Dickinson, #564105) monoclonal antibody. Fluorescence-activated cell sorting was used to sort stained peripheral blood-derived CD4 T cells into naïve CD4 T cells (CD45RO⁻, CD45RA⁺), central memory/transitional memory CD4 T cells (CD3⁺ CD8⁻ CD45RO⁺ CD27⁺) and effector memory CD4 T cells (CD3⁺ CD8⁻ CD45RO⁺ CD27⁻), GALT cells into CD45⁺ cells and cells from lymph nodes into central memory/transitional memory CD4 T cells (CD3⁺ CD8⁻ CD45RO⁺ CD27⁺) and effector memory CD4 T cells (CD3⁺ CD8⁻ CD45RO⁺ CD27⁻), using a BD FACSJazz (Beckton Dickinson, #655490) machine, as previously described (De Scheerder et al., 2019). The gating strategy used for the aforementioned sorts can be found in Figure S4. A small fraction of each sorted cell population was analyzed by flow cytometry to check for purity, which was over 95% on average. Flow cytometry data was analyzed using FlowJo software v10.6.2 (Tree-Star).

Droplet digital PCR (ddPCR)—Sorted cells were pelleted and lysed in 100 μ L lysis buffer (10 mM TrisHCl, 0.5% NP-40, 0.5% Tween-20 and proteinase K at 20 mg/mL) by incubating for 1 h at 55°C and 15 min at 85°C. HIV-1 copy number was determined by

a total HIV-1 DNA assay on droplet digital PCR (Bio-Rad, QX200 system), as described previously (Rutsaert et al., 2019). PCR amplification was carried out with the following cycling program: 10 min at 98°C; 45 cycles (30 s at 95°C, 1 min at 58°C); 10 min at 98°C. Droplets were read on a QX200 droplet reader (Bio-Rad). Analysis was performed using ddpcRquant software (Trypsteen et al., 2015).

Whole genome amplification (WGA)—Cell lysates were diluted according to ddPCR HIV-1 copy quantification, so that less than 30% of reactions contained a single genome. Whole genome amplification was performed by multiple displacement amplification with the REPLI-g single cell kit (Qiagen, #150345), according to manufacturer's instructions. The resulting amplification product was split for downstream IS analysis, single genome/proviral sequencing, and, for selected reactions, near full-length HIV-1 sequencing.

Single-genome sequencing from DNA samples—Single-proviral sequencing of the V1-V3 region of *env* was performed on DNA from lysed cells as described before (Josefsson et al., 2012; Von Stockenstrom et al., 2015), with a few adaptations. The amplification consists of a nested PCR with the following primers: Round 1 forward (E20) 5'-GGGCCACACATGCCTGTGTACCCACAG-3' and reverse (E115) 5'-AGAAAAATTCCCCTCCACAATTAA-3'; round 2, forward (E30) 5'-GTGTACCCACAGACCCCAGCCCACAAG-3' and reverse (E125) 5'-CAATTCTGGGTCCCCTCCTGAGG-3'. The 25 μ L PCR mix for the first round is composed of: 5 μ L 5X MyTaq buffer, 0.375 μ L MyTaq DNA polymerase (Bioline, #BIO-21105), 400 nM forward primer, 400 nM reverse primer and 1 μ L REPLI-g product. The mix for the second round has the same composition and takes 1 μ L of the first-round product as an input. Thermocycling conditions for first and second PCR rounds are as follows: 2 min at 94°C; 35 cycles (30 s at 94°C, 30 s at 60°C, 1 min at 72°C); 5 min at 72°C. Resulting amplicons were visualized on a 1% agarose gel and Sanger sequenced (Eurofins Genomics, Ebersberg, Germany) from both ends, using second round PCR primers.

Single-genome sequencing from plasma samples—Both 5' - and -3' -half genome amplicons were generated from plasma samples collected at time point T4. RNA was extracted from pelleted virions and cDNA was generated as follows: (1) Plasma samples were thawed at 37°C. (2) Remove debris by centrifuging the plasma for 10 min at 3,600 rpm and discarding the pellet. (3) Transfer supernatant to ultracentrifuge tube and adjust volume to 9 mL with PBS (Thermo Fisher, #20012027). (4) Centrifuge at >85,000 g for 70 min at 4°C. (5) 240 μ L of the supernatant is subjected to viral RNA extraction using the QIAamp Viral RNA Mini Kit (Qiagen, #52904), according to manufacturer's instructions. (6) Half of the RNA was used to generate cDNA for 5' -half sequencing using the R5968 primer (5'-TGTCTYCKCTTCTCCTGCCATAG-3'), while the other half was used to generate cDNA for 3' -half sequencing using the primer R9665 (5'-GTCTGAGGGATCTCTAGWTACCAGA-3'). Two mastermixes were prepared. Mix 1 consisted of 25 μ L RNA, 2.5 μ L 10 mM dNTP, 1.25 μ L 20 μ M oligo-dT (SuperScript III First Strand synthesis system, Invitrogen, #18080051) and 0.5 μ L 50 μ M primer. Mix 2 consisted of 0.75 μ L UltraPure DNase/RNase-free water (Thermo Fisher, #10977023), 5X RT buffer (Invitrogen, #18080051), 2.5 μ L 100 mM DTT (Invitrogen, #18080051), 2.5 μ L

40U/μL RNase inhibitor (Takara, #2313B), 2.5 μL SuperScript III Reverse Transcriptase (Invitrogen, #18080051) and 2 μL ThermaSTOP RT (Sigma Aldrich, #TSTOPRT-250). Mix 1 was heated to 65°C for 5 min and then snap-chilled on ice for at least 2 min. Mix 2 was pre-warmed to 50°C and then added to the chilled mix 1. The mixture was incubated at 50°C for 90 min 1 μL SuperScript III Reverse Transcriptase was added to the reaction, followed by another incubation of 90 min at 50°C and then 70°C for 15 min. Finally, 1 μL RNase H (Invitrogen, #18080051) was added, followed by an incubation of 20 min at 37°C. Subsequently, the cDNA was used as template for half-genome long-range PCRs, as described previously (Cole et al., 2021). The cDNA was diluted to endpoint to ensure the presence of single HIV-1 copies per reaction. The 25 μL PCR mix for the first round was composed of: 5 μL 5X Prime STAR GXL buffer, 0.5 μL PrimeStar GXL polymerase (Takara, #R050B), 0.125 μL ThermaStop (Sigma Aldrich, #TSTOP-500), 250 nM forward and reverse primers, and 1 μL MDA product. The mix for the second round had the same composition and took 1 μL of the first-round product as an input. Thermocycling conditions for first and second PCR rounds were as follows: 2 min at 98°C; 35 cycles (10 s at 98°C, 15 s at 62°C, 5 min at 68°C); 7 min at 68°C. Reactions without reverse transcriptase were negative, ensuring that the RNA extracts were not contaminated by DNA. PCR products were checked on a 1% agarose gel and positives were sequenced by Illumina sequencing, as described below.

Integration site loop amplification—Integration site sequencing was carried out by integration site loop amplification (ISLA), as described by Wagner et al. (Wagner et al., 2014), but with a few modifications. Firstly, the *env* primer used during the linear amplification step was omitted, as it was not necessary to recover the *env* portion of the provirus at a later stage. Therefore, the reaction was not split after the linear amplification, and the entire reaction was used as an input into subsequent decamer binding and loop formation. For some proviruses, an alternative set of primers were used to retrieve the IS from the 5' end (Table S7). Resulting amplicons were visualized on a 1% agarose gel and positives were sequenced by Sanger sequencing. Analysis of the generated sequences was performed using the 'Integration Sites' webtool developed by the Mullins lab; <https://indra.mullins.microbiol.washington.edu/integrationsites/>.

Full-length individual proviral sequencing—Sequences were generated from the genomic DNA of sorted subsets using the Full-length Individual Proviral Sequencing (FLIPS) assay as first described by Hiener et al. (Hiener et al., 2017) with some minor adjustments. Briefly, the assay consists of two rounds of nested PCR at an endpoint dilution where <30% of the wells are positive, using the BLOuterF (5'-AAATCTCTAGCAGTGGCGCCCGAACAG-3') and BLOuterR (5'-TGAGGGATCTCTAGTTACCAGAGTC-3') primers for the first round, followed by a second round using primers 275F (5'-ACAGGGACCTGAAAGCGAAAG-3') and 280R (5'-CTAGTTACCAGAGTCACACAACAGACG-3'). The cycling conditions are 94°C for 2 m; then 94°C for 30 s, 64°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 61°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 58°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 55°C for 30 s, 68°C for 10 m for 21 cycles; then 68°C for 10 m. For the second round, 10 extra cycles at 55°C are included. The PCR products were

visualized using agarose gel electrophoresis. Amplified proviruses from positive wells were cleaned using AMPure XP beads (Beckman Coulter, #A63880), followed by a quantification of each cleaned provirus with Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, #P11496). Next, an NGS library preparation using the Nextera XT DNA Library Preparation Kit (Illumina, #FC-131-1096) with indexing of 96-samples per run was used according to the manufacturer's instructions (Illumina, #FC-131-2001), except that input and reagents volumes were halved and libraries were normalized manually. The pooled library was sequenced on a MiSeq Illumina platform via 2×150 nt paired-end sequencing using the 300 cycle v2 kit (Illumina, #MS-102-2002).

Provirus amplification from MDA reactions—MDA reactions containing a potentially clonal proviral sequence were subjected to near full-length sequencing, using either a single-amplicon approach (Hiener et al., 2017), a four-amplicon approach (Patro et al., 2019), or a five-amplicon approach (Einkauf et al., 2019), as previously described. In case of the multiple amplicon approaches, amplicons were pooled equimolarly and sequenced as described above.

De Novo assembly of HIV-1 sequences—The generated sequencing data from either FLIPS or multiple amplicon approaches was demultiplexed and used to *de novo* assemble individual likely proviruses. The code used to perform *de novo* assembly can be found at the following GitHub page: https://github.com/laulambr/virus_assembly. In short, the workflow consists of following steps: (1) check of sequencing quality for each library using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and removal of Illumina adaptor sequences and trimming of 5' and -3' terminal ends using BBtools v37.99 (sourceforge.net/projects/bbmap/). (2) The trimmed reads are *de novo* assembled using MEGAHIT v1.2.9 (Li et al., 2015) generating contigs for each library. (3) Per library, all *de novo* contigs were checked using blastn v2.7.1 against the HXB2 reference virus as a filter to exclude non-HIV-1 contigs in the following analysis steps. (4) Subsequently, the trimmed reads were mapped against the *de novo* assembled HIV-1 contigs to enable the calling of the final majority consensus sequence of each provirus using bbmap v37.99.

Analysis of HIV-1 sequences—Alignments of DNA sequences for each participant were made via MAFFT (Katoh et al., 2002) and manually inspected via MEGA7 (Kumar et al., 2016). The generated HIV-1 proviruses were categorized as intact or defective as described previously (Hiener et al., 2017). NFL proviruses and half-genome plasma sequences were screened for recombination by the “DualBrothers” software (Minin et al., 2005) and the “Recombinant Identification Program” webtool from the Los Alamos National Laboratory HIV sequence database (<https://www.hiv.lanl.gov>). Phylogenetic trees were constructed using PhyML v3.0 (Guindon et al., 2010) (best of NNI and SPR rearrangements) and 1000 bootstraps. MEGA7 (Kumar et al., 2016) and iTOL v5 (Letunic and Bork, 2019) were used to visualize phylogenetic trees.

Assessment of clones detected by ISLA/SGS/FLIPS—Sequences from presumed clonal origins, shown in Figure 4, were identified according to a fixed set of rules depending on the assay. For ISLA, an integration site (IS) was considered clonal if an identical IS

was observed more than once. For SGS, V1-V3 *env* sequences had to be 100% identical to be part of an expansion of identical sequences (EIS). For FLIPS, NFL sequences part of an EIS were allowed up to 3-bp differences to account for PCR-induced errors and sequencing errors. MDA-derived NFL/IS data from infected cell clones allowed to identify links between the 3 individual datasets (ISLA/SGS/FLIPS); unambiguous links were defined as follows: (1) for ISLA data, an MDA-derived IS had to be identical to a clonal ISLA IS, (2) for SGS data, a clonal V1-V3 *env* SGS sequence had to match 100% to a V1-V3 *env* region exclusively found in that MDA-derived NFL provirus and (3) for FLIPS data, the MDA-derived NFL sequence had to match up (3-bp differences allowed) with FLIPS NFL sequences part of an EIS. In addition to the clear links with MDA data, some ambiguous assay overlaps were observed where V1-V3 *env* SGS data matched different NFL genomes with the same V1-V3 *env* region (shown as discrepancies). Note that the detection of overlaps was restricted as (1) not all IS could be linked to a V1-V3 *env* or NFL genome and (2) not all NFL genomes did contain the V1-V3 *env* region.

QUANTIFICATION AND STATISTICAL ANALYSIS

p-values in Figures 2A and 2C test for a difference in the proportion of unique IS or FLIPS proviruses between TCM/TTM and TEM subsets. These were calculated using a Pearson's Chi squared test using the "prop.test" command in R version 3.6.2 ("R Core Team," 2020). Infection frequencies for FLIPS data shown in Figure S2C were calculated by expressing the total number of identified HIV positive cells as a proportion of all cells analyzed. The infection frequency was compared across cellular subsets using a logistic regression on the number of cells positive for HIV and total number of cells by fitting a generalized linear model using the "glm" function in R version 3.6.2. Interaction between participant and cellular subset was detected ($p < 0.001$) and included in the logistic regression. p-values in Figure S2C were calculated by a two-way ANOVA using the "Anova" function from the "car" package in R (John and Sanford, 2019).

ADDITIONAL RESOURCES

Clinical trial registry number: [NCT02641756](https://clinicaltrials.gov/ct2/show/study/NCT02641756).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to acknowledge and thank all participants who donated samples to the HIV-STAR study and all the clinicians and study nurses that assisted with the sample collection. We would also like to thank Marion Pardons, Tine Struyve, and Sofie Rutsaert for providing guidance during initial data analyses, for the constructive discussions, and for critically reading the manuscript. We are grateful for the discussions with and input from James Mullins, Rafick Sékaly, Susan Pereira Ribeiro, Hadega Aamer, Sam Kint, Oleg Denisenko, Katie Fisher, and Bethany Horsburgh. In addition, we would like to thank Kim De Leeneer, Céline Helmoortel, and Bram Parton for their assistance in performing MiSeq sequencing at UZ Ghent. This current research work was supported by the NIH (R01-AI134419, MPI: L.V. and L.F.) and the Research Foundation Flanders (S000319N and G0B3820N). L.V. was supported by the Research Foundation Flanders (1.8.020.09.N.00) and the Collen-Francqui Research Professor Mandate. S.P. was supported by the Delaney AIDS Research Enterprise to Find a Cure (1U19AI096109, 1UM1AI126611-01, and 1UM1AI164560-01) and the Australian National Health and Medical Research Council (APP1061681 and APP1149990). The sample collection at UZ Ghent was supported by an MSD investigator grant (ISS 52777). B.C. and L.L. were supported by FWO Vlaanderen (1S28918N and 1S29220N). B.V. was supported

by a postdoctoral grant (12U7121N) of the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek). Additional support came from the Molecular Profiling and Computational Biology Core of the University of Washington Fred Hutch Center for AIDS Research (NIH P30 AI027757).

REFERENCES

- Aamer HA, McClure J, Ko D, Maenza J, Collier AC, Mullins JI, and Frenkel LM (2020). Cells producing residual viremia during antiretroviral treatment appear to contribute to rebound viremia following interruption of treatment. *PLoS Pathogens* 16, e1008791. [PubMed: 32841299]
- Artesi M, Hahaut V, Cole B, Lambrechts L, Ashrafi F, Marçais A, Hermine O, Griebel P, Arsic N, van der Meer F, et al. (2021). PCIP-seq: simultaneous sequencing of integrated viral genomes and their insertion sites with long reads. *Genome Biol.* 22, 97. [PubMed: 33823910]
- Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-rotolo M, Haggerty CM, Kamireddi AR, Liu Y, Lee J, et al. (2006). Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *J. Virol.* 80, 6441–6457. [PubMed: 16775332]
- Bertagnolli LN, Varriale J, Sweet S, Brockhurst J, Simonetti FR, White J, Beg S, Lynn K, Mounzer K, Frank I, et al. (2020). Autologous IgG antibodies block outgrowth of a substantial but variable fraction of viruses in the latent reservoir for HIV-1. *Proc. Natl. Acad. Sci. U S A* 117, 32066–32077. [PubMed: 33239444]
- Boritz EA, Darko S, Swaszek L, Wolf G, Wells D, Wu X, Henry AR, Laboune F, Hu J, Ambrozak D, et al. (2016). Multiple origins of virus persistence during natural control of HIV infection. *Cell* 166, 1004–1015. [PubMed: 27453467]
- Brennan TP, Woods JO, Sedaghat AR, Siliciano JD, Siliciano RF, and Wilke CO (2009). Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4+ T cells reveals a novel source of residual viremia in patients on antiretroviral therapy. *J. Virol.* 83, 8470–8481. [PubMed: 19535437]
- Cesana D, Santoni de Sio FR, Rudilosso L, Gallina P, Calabria A, Beretta S, Merelli I, Bruzzesi E, Passerini L, Nozza S, et al. (2017). HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. *Nat. Commun.* 8, 498. [PubMed: 28887441]
- Chun T, and Fauci AS (1999). Perspective Latent reservoirs of HIV: Obstacles to the eradication of virus. *Proc. Natl. Acad. Sci. U S A* 96, 10958–10961. [PubMed: 10500107]
- Chun TW, Engel D, Berrey MM, Shea T, Corey L, and Fauci AS (1998). Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc. Natl. Acad. Sci. U S A* 95, 8869–8873. [PubMed: 9671771]
- Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JAM, Baseler M, Lloyd AL, Nowak MA, and Fauci AS (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U S A* 94, 13193–13197. [PubMed: 9371822]
- Clarridge KE, Blazkova J, Einkauf K, Petrone M, Refsland W, Justement JS, Shi V, Huiting ED, Seamon CA, Lee GQ, et al. (2018). Effect of analytical treatment interruption and reinitiation of antiretroviral therapy on HIV reservoirs and immunologic parameters in infected individuals. *PLoS Pathogens* 14, e1006792. [PubMed: 29324842]
- Cohen YZ, Lorenzi JCC, Krassnig L, Barton JP, Burke L, Pai J, Lu CL, Mendoza P, Oliveira TY, Sleckman C, et al. (2018). Relationship between latent and rebound viruses in a clinical trial of anti-HIV-1 antibody 3BNC117. *J. Exp. Med.* 215, 2311–2324. [PubMed: 30072495]
- Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, et al. (2015). HIV-1 integration landscape during latent and active infection. *Cell* 160, 420–432. [PubMed: 25635456]
- Cole B, Lambrechts L, Gantner P, Noppe Y, Bonine N, Witkowski W, Chen L, Palmer S, Mullins JI, Chomont N, et al. (2021). In-depth single-cell analysis of translation-competent HIV-1 reservoirs identifies cellular sources of plasma viremia. *Nat. Commun.* 12, 3727. [PubMed: 34140517]
- De Scheerder M-A, Vrancken B, Dellicour S, Schlub T, Lee E, Shao W, Rutsaert S, Verhofstede C, Kerre T, Malfait T, et al. (2019). HIV rebound is predominantly fueled by genetically identical viral expansions from diverse reservoirs. *Cell Host Microbe* 26, 347–358. [PubMed: 31471273]

- Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ, et al. (2019). Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J. Clin. Invest.* 129, 988–998. [PubMed: 30688658]
- Estes JD, Kityo C, Ssali F, Swainson L, Makamdop KN, Del Prete GQ, Deeks SG, Luciw PA, Chipman JG, Beilman GJ, et al. (2017). Defining total-body AIDS-virus burden with implications for curative strategies. *Nat. Med.* 23, 1271–1276. [PubMed: 28967921]
- Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, 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. [PubMed: 9360927]
- Garner SA, Rennie S, Ananworanich J, Dube K, Margolis DM, Sugarman J, Tressler R, Gilbertson A, and Dawson L (2017). Interrupting antiretroviral treatment in HIV cure research: scientific and ethical considerations. *J. Virus Eradication* 3, 82–84.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, and Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. [PubMed: 20525638]
- Halvas EK, Hughes SH, Mellors JW, Halvas EK, Joseph KW, Brandt LD, Guo S, Sobolewski MD, Jacobs JL, Tumiotto C, et al. (2020). HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. *J. Clin. Invest.* 130, 5847–5857. [PubMed: 33016926]
- Hienor B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, von Stockenström S, Odeval L, Milush JM, Liegler T, et al. (2017). Identification of genetically intact HIV-1 proviruses in specific CD4⁺T cells from effectively treated participants. *Cell Rep.* 21, 813–822. [PubMed: 29045846]
- Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DIS, Keele BF, Ho Y-C, Siliciano JD, and Siliciano RF (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. [PubMed: 28341641]
- Huang AS, Ramos V, Oliveira TY, Gaebler C, Jankovic M, Nussenzweig MC, and Cohn LB (2021). Integration features of intact latent HIV-1 in CD4⁺ T cell clones contribute to viral persistence. *J. Exp. Med.* 218, e20211427. [PubMed: 34636876]
- Imamichi H, Smith M, Adelsberger JW, Izumi T, Scrimieri F, Sherman BT, Rehm CA, Imamichi T, Pau A, Catalfamo M, et al. (2020). Defective HIV-1 proviruses produce viral proteins. *Proc. Natl. Acad. Sci. U S A* 117, 3704–3710. [PubMed: 32029589]
- Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, Chen SMY, Hua S, Rhee B, Chang K, et al. (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature* 585, 261–267. [PubMed: 32848246]
- John F, and Sanford W (2019). *An R Companion to Applied Regression* (Sage).
- 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. [PubMed: 22536001]
- Josefsson L, von Stockenström S, Faria NR, Sinclair E, Bacchetti P, Killian M, Epling L, Tan A, Ho T, Lemey P, et al. (2013). 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. U S A* 110, E4987–E4996. [PubMed: 24277811]
- Katoh K, Misawa K, Kuma K, and Miyata T (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. [PubMed: 12136088]
- Kearney MF, Wiegand A, Shao W, Coffin JM, Mellors JW, Lederman M, Gandhi RT, Keele BF, and Li JZ (2016). Origin of rebound plasma HIV Includes cells with identical proviruses that are transcriptionally active before stopping of antiretroviral therapy. *J. Virol.* 90, 1369–1376. [PubMed: 26581989]
- Kumar S, Stecher G, and Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. [PubMed: 27004904]
- Lambrechts L, Cole B, Rutsaert S, Trypsteen W, and Vandekerckhove L (2020). Emerging PCR-based techniques to study HIV-1 reservoir persistence. *Viruses* 12, 1–12.

- Laskey SB, Pohlmeier CW, Bruner KM, and Siliciano RF (2016). Evaluating clonal expansion of HIV-infected cells: Optimization of PCR strategies to predict clonality. *PLoS Pathog.* 12, e1005689. [PubMed: 27494508]
- Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, 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 [PubMed: 28628034]
- Letunic I, and Bork P (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. [PubMed: 30931475]
- Li D, Liu C-M, Luo R, Sadakane K, and Lam T-W (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. [PubMed: 25609793]
- Liu R, Simonetti FR, and Ho Y (2020). The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol. J* 17.
- Lu C, Pai JA, Nogueira L, Mendoza P, Gruell H, Oliveira TY, and Barton J (2018). Relationship between intact HIV-1 proviruses in circulating CD4 + T cells and rebound viruses emerging during treatment interruption. *Proc. Natl. Acad. Sci. U S A* 115, 11341–11348.
- Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, et al. (2014). Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 345, 179–183. [PubMed: 24968937]
- Mcmanus WR, Coffin JM, Kearney MF, Mcmanus WR, Bale MJ, Spindler J, Wiegand A, Musick A, Patro SC, Sobolewski MD, et al. (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *J. Clin. Invest.* 129, 4629–4642. [PubMed: 31361603]
- Minin VN, Dorman KS, Fang F, and Suchard MA (2005). Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 21, 3034–3042. [PubMed: 15914546]
- Pannus P, Rutsaert S, De Wit S, Allard SD, Vanham G, Cole B, Nescoi C, Aerts J, De Spiegelaere W, Tsoumanis A, et al. (2020). Rapid viral rebound after analytical treatment interruption in patients with very small HIV reservoir and minimal on-going viral transcription. *J. Int. AIDS Soc.* 23, e25453. [PubMed: 32107887]
- Patro SC, Brandt LD, Bale MJ, Halvas EK, Joseph KW, Shao W, and Wu X (2019). Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. *Proc. Natl. Acad. Sci. U S A* 116, 25891–25899. [PubMed: 31776247]
- Peluso MJ, Laird GM, Deeks SG, Peluso MJ, Bacchetti P, Ritter KD, Beg S, Lai J, Martin JN, Hunt PW, et al. (2020). Differential decay of intact and defective proviral DNA in HIV-1 – infected individuals on suppressive antiretroviral therapy. *JCI Insight* 5, e132997. [PubMed: 32045386]
- Pinzone MR, Vanbelzen DJ, Weissman S, Bertuccio MP, Cannon L, Venanzi-rullo E, Migueles S, Jones RB, Mota T, Joseph SB, et al. (2019). Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat. Commun.* 10, 728. [PubMed: 30760706]
- Pollack RA, Jones RB, Perteza M, Bruner KM, Martin AR, Thomas AS, Capoferri AA, Beg SA, Huang SH, Karandish S, et al. (2017). 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. [PubMed: 28407485]
- R Core Team (2020). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Rutsaert S, De Spiegelaere W, De Clercq L, and Vandekerckhove L (2019). Evaluation of HIV-1 reservoir levels as possible markers for virological failure during boosted darunavir monotherapy. *J. Antimicrob. Chemother.* 74, 3030–3034. [PubMed: 31314108]
- Salantes DB, Tebas P, Bar KJ, Salantes DB, Zheng Y, Mampe F, Srivastava T, Beg S, Lai J, Li JZ, et al. (2018). HIV-1 latent reservoir size and diversity are stable following brief treatment interruption. *J. Clin. Invest.* 128, 3102–3115. [PubMed: 29911997]
- Simonetti FR, Sobolewski MD, Fyne E, Shao W, Spindler J, Hattori J, Anderson EM, Watters SA, Hill S, Wu X, et al. (2016). Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proc. Natl. Acad. Sci. U S A* 113, 1883–1888. [PubMed: 26858442]

- Tobin NH, Learn GH, Holte SE, Wang Y, Melvin AJ, McKernan JL, Pawluk DM, Mohan KM, Lewis PF, Mullins JI, et al. (2005). Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *J. Virol.* 79, 9625–9634. [PubMed: 16014925]
- Trypsteen W, Vynck M, De Neve J, Bonczkowski P, Kiselinova M, Malatinkova E, Vervisch K, Thas O, Vandekerckhove L, and De Spiegelaere W (2015). ddpcRquant: threshold determination for single channel droplet digital PCR experiments. *Anal. Bioanal. Chem.* 407, 5827–5834. [PubMed: 26022094]
- Vibholm LK, Lorenzi JCC, Pai JA, Cohen YZ, Oliveira TY, Barton JP, Garcia Noceda M, Lu C-L, Ablanedo-Terrazas Y, Del Rio Estrada PM, et al. (2019). Characterization of intact proviruses in blood and lymph node from HIV-infected individuals undergoing analytical treatment interruption. *J. Virol.* 93, e01920–18. [PubMed: 30700598]
- 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. [PubMed: 25712966]
- Wagner TA, McKernan JL, Tobin NH, Tapia KA, Mullins JI, Frenkel M, and Frenkel LM (2013). An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J. Virol.* 87, 1770–1778. [PubMed: 23175380]
- Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, and Frenkel LM (2014). Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 345, 570–573. [PubMed: 25011556]
- Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, Kumar MR, Beg SA, Capoferri AA, Ray SC, et al. (2018). Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc. Natl. Acad. Sci. U S A* 115, E2575–E2584. [PubMed: 29483265]

Highlights

- Assays to study clonal expansion of HIV-1-infected cells introduce distinct biases
- Sequence matches are detected between low-level/residual viremia and proviruses
- Infected cells that contribute to rebound are often maintained by clonal expansion

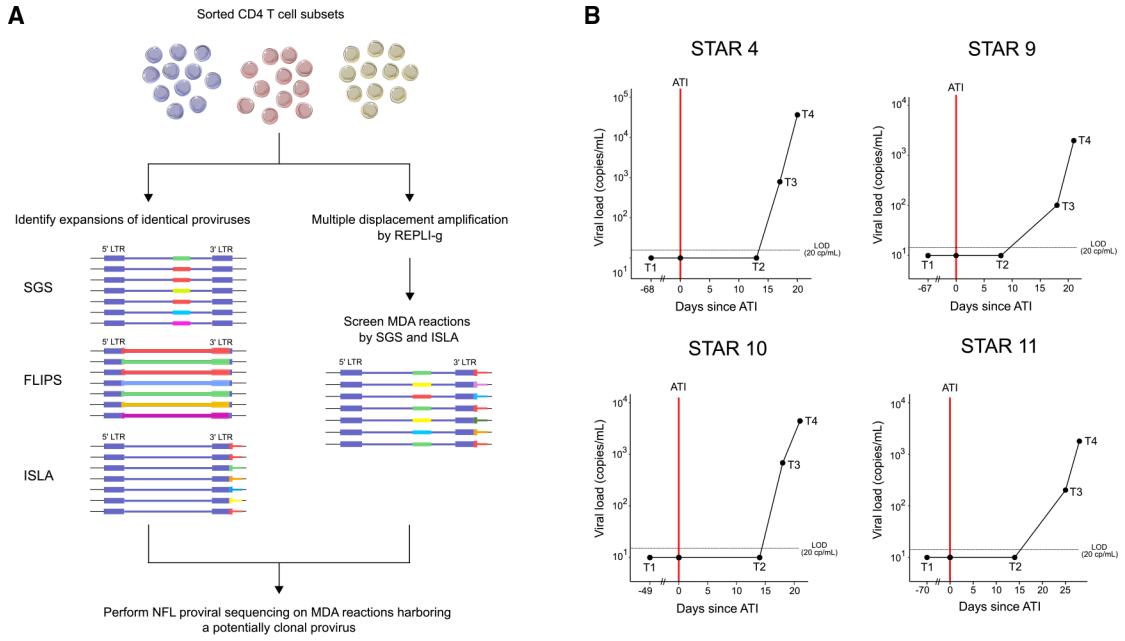


Figure 1. Overview of the workflow for HIV-1 reservoir characterization and viral loads at each time point of sample collection

(A) Workflow of HIV-1 reservoir characterization by single-genome sequencing (SGS), full-length individual proviral sequencing (FLIPS), integration site loop amplification (ISLA), and multiple displacement amplification (MDA). In a first step, potentially clonal HIV-1-infected cells were identified by SGS, FLIPS, and ISLA on lysed sorted CD4 T cell subsets. In a second step, MDA with subsequent SGS and ISLA was performed on selected sorted cell lysates. In the final step, MDA reactions containing a potentially clonal provirus were identified, and the near full-length (NFL) genome of the according provirus was amplified and sequenced.

(B) Viral load (copies/mL) at each time point of sample collection for all participants. The day of analytical treatment interruption (ATI) initiation is indicated with a vertical red line. The plasma was sampled during ART (time point 1, T1), 8 to 14 days after ATI (time point 2, T2), at the first detectable viral load (time point 3, T3), and at later rebound (time point 4, T4). Note that T1 is not shown to scale. The horizontal dashed lines indicate the limit of detection at 20 copies/mL. See also Figure S4 and Tables S1, S6, and S7.

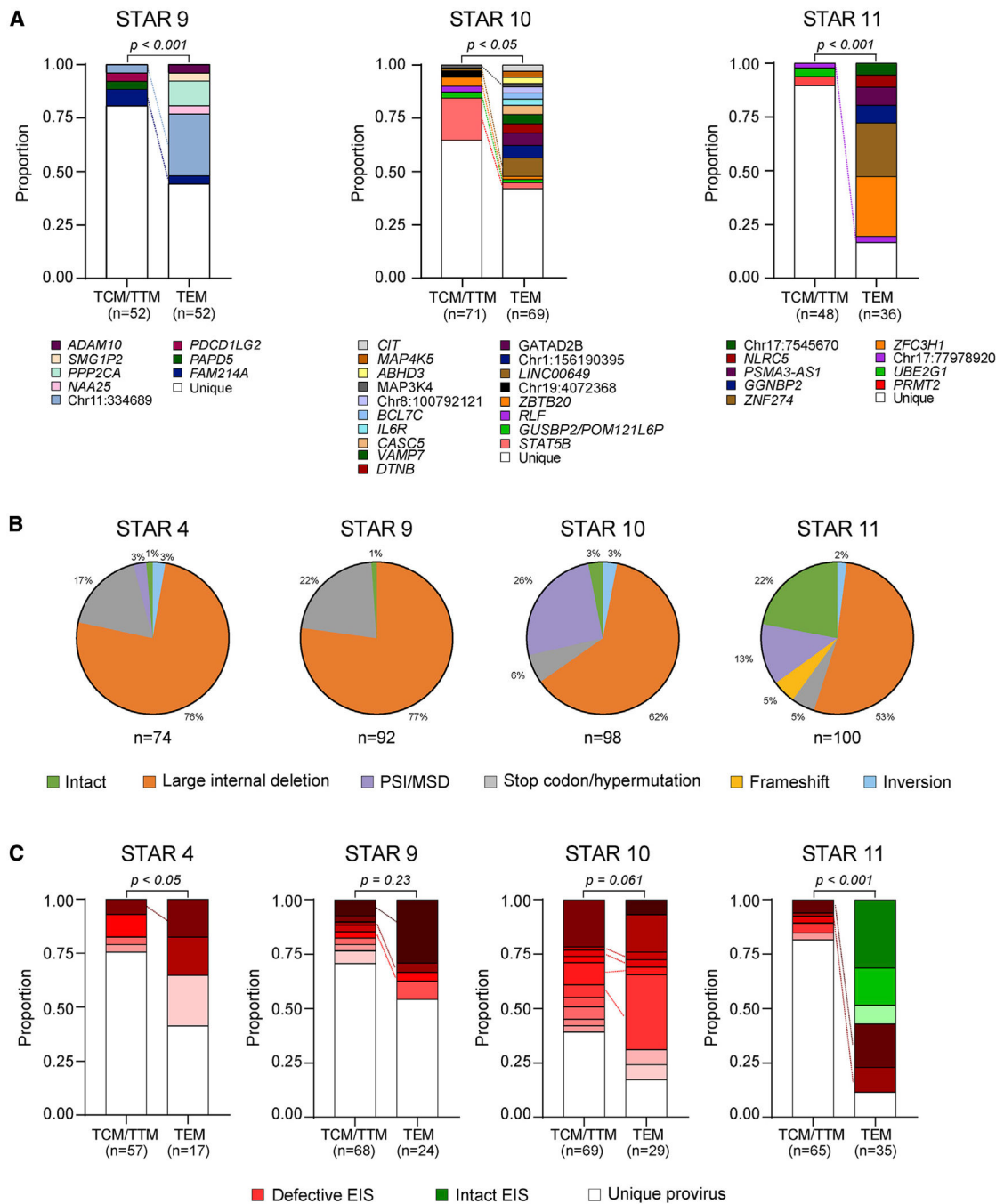


Figure 2. Classification of HIV-1 integration sites and proviral near full-length genome sequences across different cell subsets before ATI

(A) Proportions of integration sites (IS) retrieved by integration site loop amplification (ISLA) for participants STAR 9, STAR 10, and STAR 11 from TCM/TTM and TEM subsets from peripheral blood. IS found more than once are defined as “clonal” and are shown in color as proportion of all IS. Identical IS found in both subsets are linked with dashed lines. p values test was used for a difference in the proportion of unique IS between TCM/TTM and TEM by “prop.test” in R.

(B) Proportions of intact and defective near full-length sequences from full-length individual proviral sequencing (FLIPS) within TCM/TTM and TEM peripheral blood for each participant. See Figures S1 and S2A for figures including all FLIPS data.

(C) Proportions of expansions of identical sequences (EIS) found in TCM/TTM and TEM peripheral blood subsets based on FLIPS data for each participant. EIS consisting of a defective and intact provirus are shown in shades of red and green, respectively, while unique proviruses are represented in white. The “prop.test” in R was used to test for differences in the proportion of unique proviruses between TCM/TTM and TEM. TCM/TTM, central/transitional memory CD4 T cell; TEM, effector memory CD4 T cell; PSI, packaging signal; MSD, major splice donor. See also Figures S1 and S2 and Tables S2 and S4.

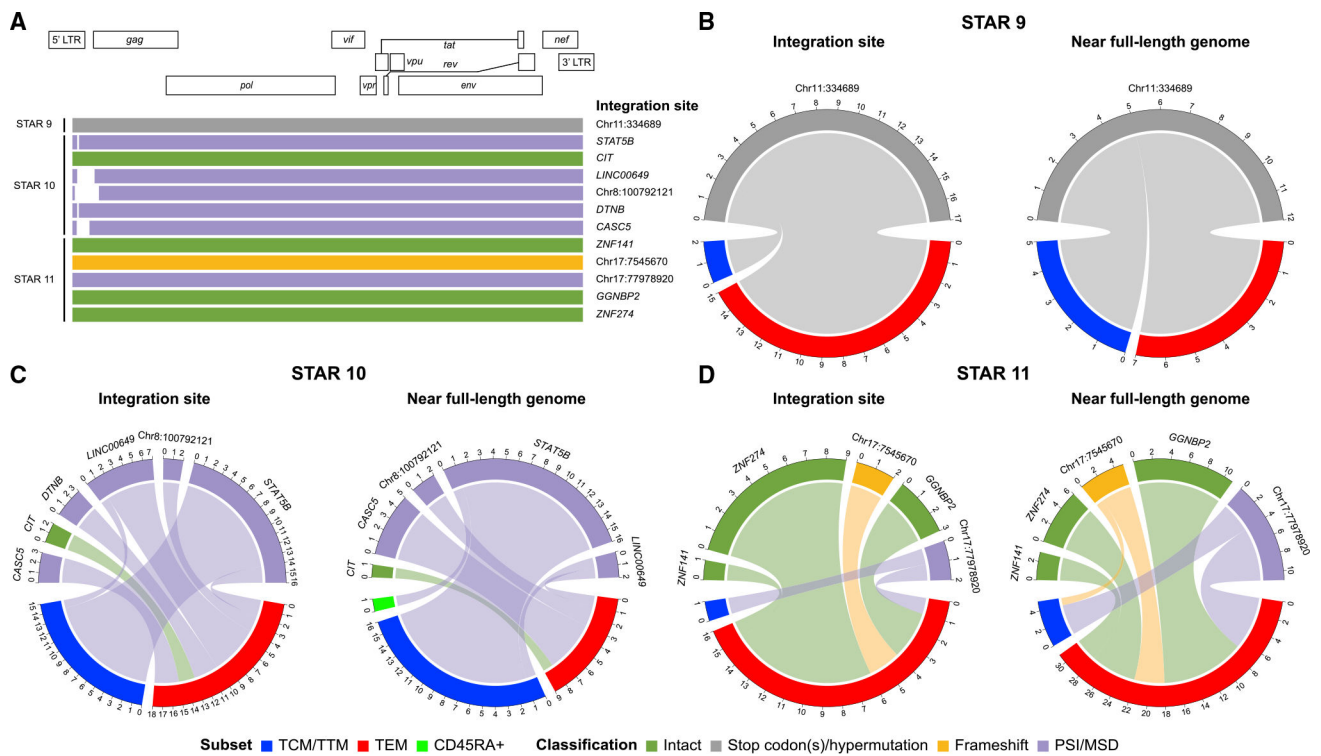


Figure 3. Near full-length proviral HIV-1 genomes and associated integration sites recovered from the peripheral blood by MDA

(A) For each participant, the recovered proviral genome structures are shown aligned to the HXB2 reference genome, and corresponding integration sites are listed on the right. For more information, see Table S3. Each provirus is colored according to their structural category.

(B–D) For each participant, the number of integration site (IS) and near full-length (NFL) proviruses generated via full-length individual proviral sequencing (FLIPS) linked to clones characterized by multiple displacement amplification (MDA) is shown together with their corresponding cellular subset. TCM/TTM, central/transitional memory CD4 T cell; TEM, effector memory CD4 T cell; PSI, packaging signal; MSD, major splice donor. See also Table S3.

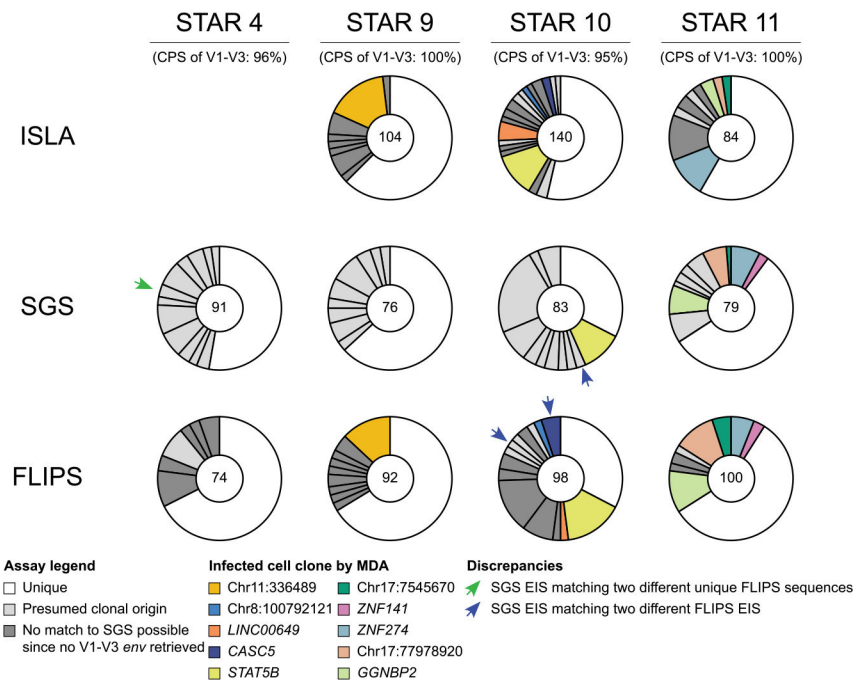


Figure 4. Comparison of assays to identify potentially clonal HIV-1-infected cell populations
 Clonality assessments are shown for integration site loop amplification (ISLA), single-genome sequencing (SGS), and full-length individual proviral sequencing (FLIPS). Identical sequences found multiple times within the same assay are shown as light gray slices. The total number of examined integration sites (IS), V1–V3 *env* sequences, and near full-length (NFL) proviral sequences is noted in the middle of each donut plot. When an IS, V1–V3 *env* SGS, or NFL FLIPS sequence could be unambiguously linked to data from previously identified infected cell clones by multiple displacement amplification (MDA), they were given a distinct standout color and chromosome designation as indicated in the legend. Dark gray slices show identical sequences that did not allow for overlap detection as (1) not all IS could be linked via MDA to a V1–V3 *env* or NFL sequence and (2) not all NFL genomes contained the V1–V3 *env* region. Arrows are used to indicate discrepancies between the different assays. EIS, expansion of identical sequences. See also Table S5.

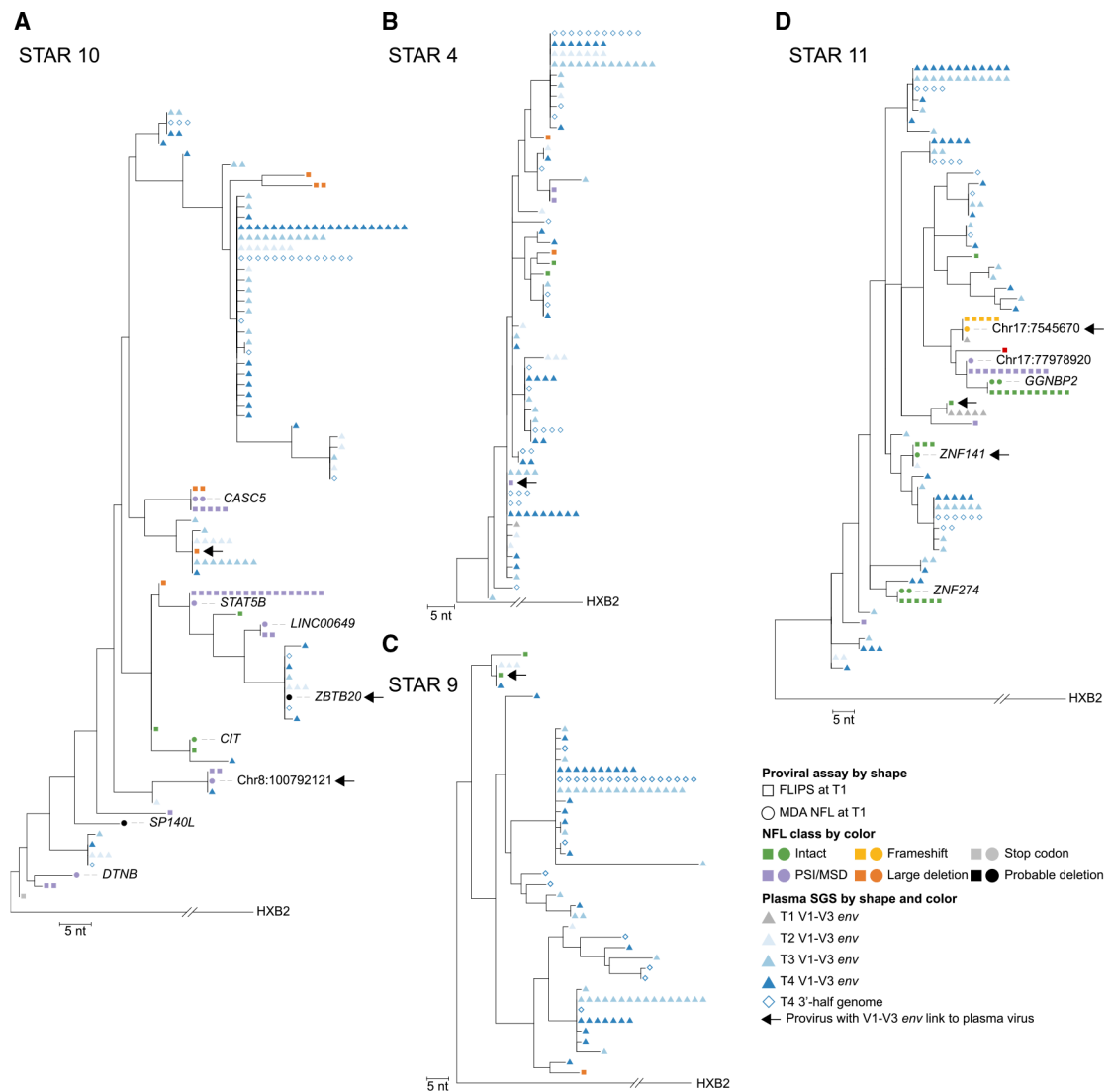


Figure 5. Maximum-likelihood phylogenetic trees based on the V1–V3 *env* region from proviral and plasma sequences before and during ATI

Proviral sequences collected before ATI from full-length individual proviral sequencing (FLIPS) and multiple displacement amplification (MDA) are shown as squares and circles, respectively. The integration sites (IS) associated with MDA-derived proviruses are noted if available. Plasma sequences are shown as triangles (V1–V3 *env*) or diamonds (3′-half genome) where the color indicates the time point during analytical treatment interruption (ATI). Arrows indicate identical matches between proviral and plasma V1–V3 *env* sequences. All trees are rooted to the HXB2 reference sequence.

(A) In participant STAR 10, three identical matches between defective proviral and plasma rebound sequences were found. For two, the corresponding IS *ZBTB20* and Chr8:100792121 could be recovered.

(B) In participant STAR 4, only one match between a unique major splice donor (MSD) deleted provirus and plasma sequences was observed.

(C) In STAR 9, a match between a unique intact provirus and multiple plasma sequences from different time points were found.

(D) In STAR 11, a rebounding plasma sequence could be linked to an expansion of identical intact near full-length (NFL) genomes located in the *ZNF141* gene. One unique intact provirus can be linked to a residual plasma sequence from T1. SGS, single-genome sequencing; PSI, packaging signal. See also Figure S3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD4 neg selection kit for IMag; human CD4 T cell	BD Biosciences	Cat#557939; RRID: AB_2802162
Anti-human CD3 BB	BD Biosciences	Cat#564465; RRID: AB_2744386
Anti-human CD8 PeCy7	BD Biosciences	Cat#557746; RRID: AB_396852
Anti-human CD45 PerCPCy5.5	BD Biosciences	Cat#564105; RRID: AB_2744405
Anti-human CD45 RO PE	BD Biosciences	Cat#555493; RRID: AB_395884
Anti-human CD45 RO PerCPCy5.5	BD Biosciences	Cat#560607; RRID: AB_1727500
Anti-human CD45 RA APC	BD Biosciences	Cat#550855; RRID: AB_398468
Anti-human CD27 APC	BD Biosciences	Cat#561400; RRID: AB_10645790
Anti-human CD197 (CCR7) PE	BD Biosciences	Cat#560765; RRID: AB_2033949
Viability stain 780	BD Biosciences	Cat#565388; RRID: AB_2869673
Biological samples		
Patient derived PBMCs (HIV-STAR)	Ghent University Hospital	N/A
Patient derived Plasma (HIV-STAR)	Ghent University Hospital	N/A
Chemicals, peptides, and recombinant proteins		
Proteinase K 20mg/mL	Ambion/Lifetechnologies	Cat#AM2546
Ultrapure 1M TrisHCL PH 8	Thermo Fisher Scientific	Cat#15568-025
Nonidet P 40 Substitute solution	Sigma Aldrich	Cat#98379
Tween-20	Sigma Aldrich	Cat#P9416
PBS, pH 7.2	Thermo Fisher Scientific	Cat#20012027
dNTP PCR nucleotide mix 10 mM 1mL	Promega	Cat#C1145

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UltraPure DNase/RNase-free water	Thermo Fisher Scientific	Cat#10977023
RNase inhibitor	Takara	Cat#2313B
ThermaSTOP RT	Sigma Aldrich	Cat#TSTOPRT-250
RNase H	Invitrogen	Cat#18080051
PrimeStar GXL polymerase	Takara	Cat#R050B
ThermaStop	Sigma Aldrich	Cat#TSTOP-500
AMPure XP	Beckman Coulter	Cat#A63880
Critical commercial assays		
REPLI-g single cell kit	Qiagen	Cat#150345
MyTaq DNA polymerase	Bioline	#BIO-21105
QIAamp Viral RNA Mini Kit	Qiagen	Cat#52904
SuperScript III First Strand synthesis system	Invitrogen	Cat#18080051
Quant-iT PicoGreen dsDNA Assay Kit	Invitrogen	Cat#P11496
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
Nextera XT Index Kit v2 Set A	Illumina	FC-131-2001
MiSeq Reagent Kit v2 (300-cycles)	Illumina	MS-102-2002
Deposited data		
RNA and DNA sequencing – V1-V3 <i>env</i> SGS data	GenBank	GenBank: MH639149–MH643573
RNA and DNA sequencing – V1-V3 <i>env</i> SGS data	GenBank	GenBank: MH648157–MH648606
RNA and DNA sequencing – FLIPS NFL data	GenBank	GenBank: MZ041268–MZ041657
RNA and DNA sequencing – 5' and 3' end plasma data	GenBank	GenBank: MZ955650–MZ955856

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
See Table S7 for a list of used oligonucleotides.		
Software and algorithms		
Flow Jo software	BD Biosciences	https://www.flowjo.com/
ddpcRquant software	Trypsteen et al., 2015	http://statapps.ugent.be/dPCR/ddpcrquant/
'Integration Sites' webtool	Mullins Lab	https://indra.mullins.microbiol.washington.edu/integrationsites/
De novo assembly pipeline	Cole et al., 2021	https://github.com/laulamb/virus_assembly
FastQC	Babraham Bioinformatics	http://www.bioinformatics.babraham.ac.uk/projects/fastqc
bbmap	BBMap	sourceforge.net/projects/bbmap/
MEGAHIT	Li et al., 2015	https://github.com/voutcn/megahit
MAFFT	Katoh et al., 2002	https://mafft.cbrc.jp/alignment/software/
MEGA7	Kumar et al., 2016	https://www.megasoftware.net/
DualBrothers	Minin et al., 2005	https://msuchard.faculty.biomath.ucla.edu/DualBrothers/index.html#:~:text=DualBrothers%20is%20a%20recombination%20detection,in%20a%20multiple%20sequence%20alignment
Recombinant Identification Program	Los Alamos HIV Database	https://www.hiv.lanl.gov
PhyML v3.0	Guindon et al., 2010	https://github.com/stephaneguindon/phyml/
iTOL v5	Letunic and Bork, 2019	https://itol.embl.de/
R version 3.6.2	R Core Team, 2020	https://www.R-project.org/
Companion to Applied Regression, "car"	R package	https://cran.r-project.org/web/packages/car/index.html
Other		
BD FACSJazz	BD Biosciences	Cat#655490
QX200 Droplet Digital PCR System	Biorad	Cat#1864001
QX200 Droplet Generator	Biorad	Cat#1864002
MiSeq System	Illumina	Cat#SY-410-1003