communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-09103-7

Aberrant strand transfer events across multiple stages of reverse transcription shape the heterogeneous landscape of the HIV-1 reservoir

Jarryt Hardy[1], Virginie Mortier[1], Yiannis De Meersman[1,3,4,5], Delfien Staelens[1], Els Demecheleer[1], Marlies Schauvliege[1], Ciel Vermeersch[1], Elizaveta Padalko[1,2] & Chris Verhofstede[1]

Curative interventions for HIV-1 seek to reduce the size of the replication-competent viral reservoir. A comprehensive understanding of the composition of defective proviral sequences is essential for refining the methods used to characterize and measure this replication-competent viral reservoir. An in-depth analysis of proviral sequences with a single internal deletion was performed to better understand the mechanisms behind their origin. An in-house full-length individual proviral sequencing technique was adopted for HIV-1 sequencing. This sequence dataset was further supplemented with sequences from two published studies, adding up to 395 proviral sequences. Based on the deletion junction profile of each individual provirus, six hypothetical mechanisms of aberrant strand transfer during reverse transcription were modelled. Our findings show that proviral sequences with deletions do not exclusively result from erroneous strand transfer events during minus-strand synthesis but emerge through distinct mechanisms throughout the reverse transcription process.

Since the development of potent antiretroviral therapy (ART), human immunodeficiency virus type 1 (HIV-1) infection can be managed as a chronic, well-controlled disease. However, strict adherence to ART is required to suppress viral replication, as current ART cannot eradicate the virus in people living with HIV (PLWH). Interruption of treatment leads to the re-establishment of replication from a stable reservoir of resting memory $CD4^+$ T cells that harbour integrated, replication-competent HIV-1 DNA[1–7]. Curative interventions aim to reduce the size of this reservoir or even eradicate it.

Over the past decades, significant progress has been made in characterizing the viral reservoir, acknowledging that a comprehensive understanding of its genetic composition is crucial for the development of new strategies to disarm it. Studies have shown that only a minuscule proportion of proviral DNA isolated from PLWH on ART is intact and considered replication-competent. Several multiplex PCR-based assays have been developed to assess the genomic intactness of the provirus. Medium-throughput digital PCR (dPCR) assays, such as the IPDA, 5T-IPDA, Q4PCR, and RAINBOW assays, rely on the amplification of subgenomic regions of the proviral DNA to quantify the replication-competent HIV-1 reservoir[8–11]. However, amplification of subgenomic regions tends to overestimate the number of intact proviruses[12]. Moreover, the outcomes of these multiplex dPCR assays are highly susceptible to errors, including those caused by sequence polymorphisms in the primer/probe binding regions[13–15]. More in-depth characterization of the genomic composition of integrated HIV-1 DNA is typically performed using near full-length individual proviral amplification and sequencing techniques (FLIPS)[16–18]. Although FLIPS is time-consuming and expensive due to its low throughput, it provides valuable insights into the genetic landscape of the viral reservoir and its dynamics. An important shortcoming, however, is that the preferential amplification of proviruses with a large internal deletion, which is inherent to the PCR reaction, prevents a reliable quantitative understanding of the composition of the latent reservoir[19]. Based on the results of FLIPS, it is estimated that over 90% of all proviral sequences are defective, with the most common defects being internal deletions, small deletions at the packaging signal and the major splice donor site, and guanine-to-adenine hypermutations[7,10,16–18,20].

[1]AIDS Reference Laboratory, Department of Diagnostic Sciences, Ghent University, Ghent, Belgium. [2]Laboratory of Medical Microbiology, Department of Diagnostic Sciences, Ghent University, Ghent, Belgium. [3]Present address: Lung Research Center, Department of Internal Medicine and Pediatrics, Ghent University, Ghent, Belgium. [4]Present address: Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [5]Present address: Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands. e-mail: jarryt.hardy@ugent.be

Defects in the proviral sequences are thought to arise from the same mechanisms that drive the genomic adaptability of HIV-1. The driving force behind the high propensity for variation in the HIV genome is the error-prone nature and lack of proofreading activity of the reverse transcription enzyme. The forward mutation rate for HIV-1 is estimated to be approximately 10^{-5} mutations per base pair per replication cycle[21–23]. HIV-1 relies on the enzymatic characteristics of reverse transcriptase to copy its genetic information from single-stranded RNA into double-stranded DNA. Completion of this process requires two successful DNA strand transfers, in which reverse transcriptase transposes from one side of the template to the other[24]. The mechanisms behind strand transfer events can also drive intermolecular template switching between co-packaged RNA genomes. First, RNase H incisions downstream of pause sites in the reverse transcription process function as initiation sites for the invasion-driven strand transfer mechanism during minus-strand synthesis[25–27]. The acceptor RNA can then displace the cleaved fragments of the donor RNA template and propagate through a branch migration process. These intermolecular transfers are essential for overcoming genomic damage and, according to in vitro research, are estimated to occur 3 to 15 times per reverse transcription cycle[28–31]. Although these properties of the virus function as mechanisms of selective advantage, they also contribute to the formation of defective proviruses. Erroneous strand transfers during minus-strand cDNA synthesis can result in DNA with internal deletions or duplications. These aberrant strand transfers are often characterized by short stretches of homologous sequences at the deletion junctions. Imamichi et al. looked at repeated sequences at deletion junctions and suggested that 40% of the detected internal deletions were attributed to erroneous strand transfers during minus-strand synthesis[18]. Another source of defective proviruses is the deamination of cytosine on minus-strand cDNA by predominantly APOBEC3G and APOBEC3F. Conversion of cytosine to uracil in the minus-strand results in G-to-A hypermutations in the plus-strand DNA, leading to premature stop codons and other amino acid mutations[32–37].

Recent studies have shown that defective proviruses may retain some of their transcriptional and translational competence. Novel truncated, unspliced HIV-RNA transcripts with translationally competent open reading frames (ORFs) have been identified in individuals with prolonged viral suppression[18,38]. It is reported that protein production from defective proviruses drives viral immune escape mechanisms, persistent immune activation, and pathogenesis observed in individuals on prolonged ART[39–42]. Although these findings indicate that proviruses identified as defective may contribute to persistent comorbidities associated with HIV-1 pathogenesis in individuals on ART, most of the HIV reservoir studies centre solely on the genomically intact proviruses.

Under the assumption that understanding the mechanisms behind the emergence of defective proviral sequences may help optimize methods for quantifying the clinically relevant viral reservoir and improve our understanding of their role in disease persistence, we conducted an in-depth analysis of the structure of HIV-1 proviral sequences with large internal deletions. Proviral sequences were obtained from two published studies and supplemented with sequences generated in our own laboratory (Ghent Aids Reference Laboratory, ARL) using a comparable FLIPS method[16,18]. Our observations demonstrated that, relative to the size of the different genomic regions, aberrant strand transfers occur most frequently in the long terminal repeats (LTR) regions of the HIV genome. An additional hotspot was identified near the central polypurine tract (cPPT). Further analysis of the deletion junction positions led to the modelling of six hypothetical mechanisms of aberrant strand transfer during reverse transcription that contribute to the generation of a heterogeneous pool of proviruses with internal deletions.

Results

Length analysis of proviral sequences with an internal deletion

Three hundred ninety-five proviral sequences with a single internal deletion of more than 200 nucleotides were compared based on sequence length (Supplementary Data 1). Sequence lengths shorter than 4 kb were observed in 266 sequences (67.3%). A length distribution curve revealed a remarkably high prevalence of 2- and 5-kilobase amplicons (Fig. 1a). The distribution curve for the sequences collected from Imamichi et al. differed slightly from those of the two other studies (Fig. 1b). Kruskal-Wallis analysis confirmed significant differences in median sequence length between the studies, $H(2) = 20.12$, $p < 0.0001$. A post-hoc Dunn's test revealed a significantly higher mean rank size score for the sequences from Imamichi et al. compared to those retrieved from Hiener et al., $p < 0.0001$, and the ARL, $p = 0.0002$, while no significant difference was observed between the sequences retrieved from Hiener et al. and the ARL, $p = 0.4358$.

Hypermutation analysis revealed 65 hypermutated sequences among the 395 proviral sequences. A length distribution curve demonstrates that the hypermutated sequences are uniformly dispersed along the length axis (Supplementary Information 1a). In addition, omission of the hypermutated sequences did not alter the median of the sequence lengths across the distinct studies (Supplementary Information 1b). A Mann-Whitney test was performed to further determine the repercussions of excluding hypermutated sequences on the length distribution of the dataset. Results indicated that the omission of hypermutated sequences did not significantly alter the distribution of sequence lengths, $U = 64{,}173$, $p = 0.7214$. We decided to retain the hypermutated sequences for subsequent analysis.

Deletion junction analysis

The start and end positions of the deletion junctions were mapped to their respective genomic regions. Deletion start positions were predominantly observed in the *gag* and *pol* regions, while deletion end positions were most frequently observed in the 3' LTR (Fig. 2a and Supplementary Data 2 and 3). When accounting for genomic region size to assess the frequency of deletion start and end sites within specific genomic regions, the LTR regions emerge as the most prominent hotspots, with 56.6% and 62.6% of the deletion junctions observed in the 5' and 3' LTR regions, respectively (Fig. 2b and Supplementary Data 2 and 3).

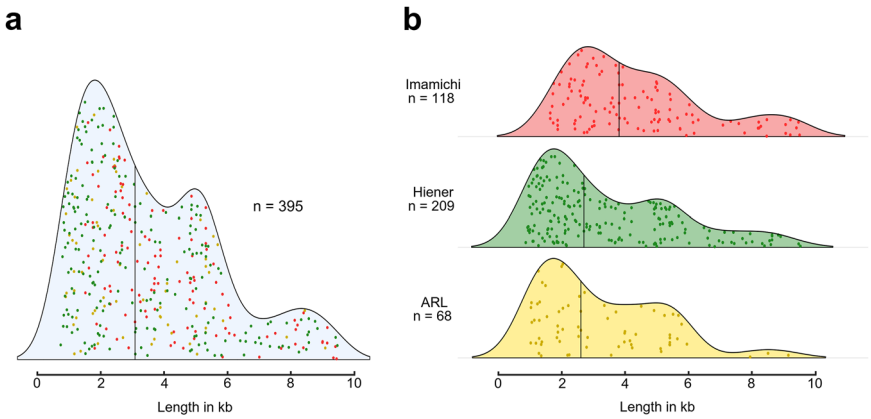


Fig. 1 | Length distribution curves of proviral sequences with a deletion of more than 200 nucleotides. Proviral sequences derive from an in-house FLIPS technique (ARL protocol) or were retrieved from two published studies (16,18). Length distribution curves of **a** the aggregated FLIPS-generated sequences and **b** each dataset individually. Each dot represents a single provirus and is coloured according to the study from which it originates: Hiener (green), Imamichi (red), in-house ARL protocol (yellow). The number of sequences examined for each subject (*n*) is reported. The median sequence length is depicted in each distribution curve. Created in BioRender. Hardy, J. (2025) https://BioRender.com/c0nf6h4.

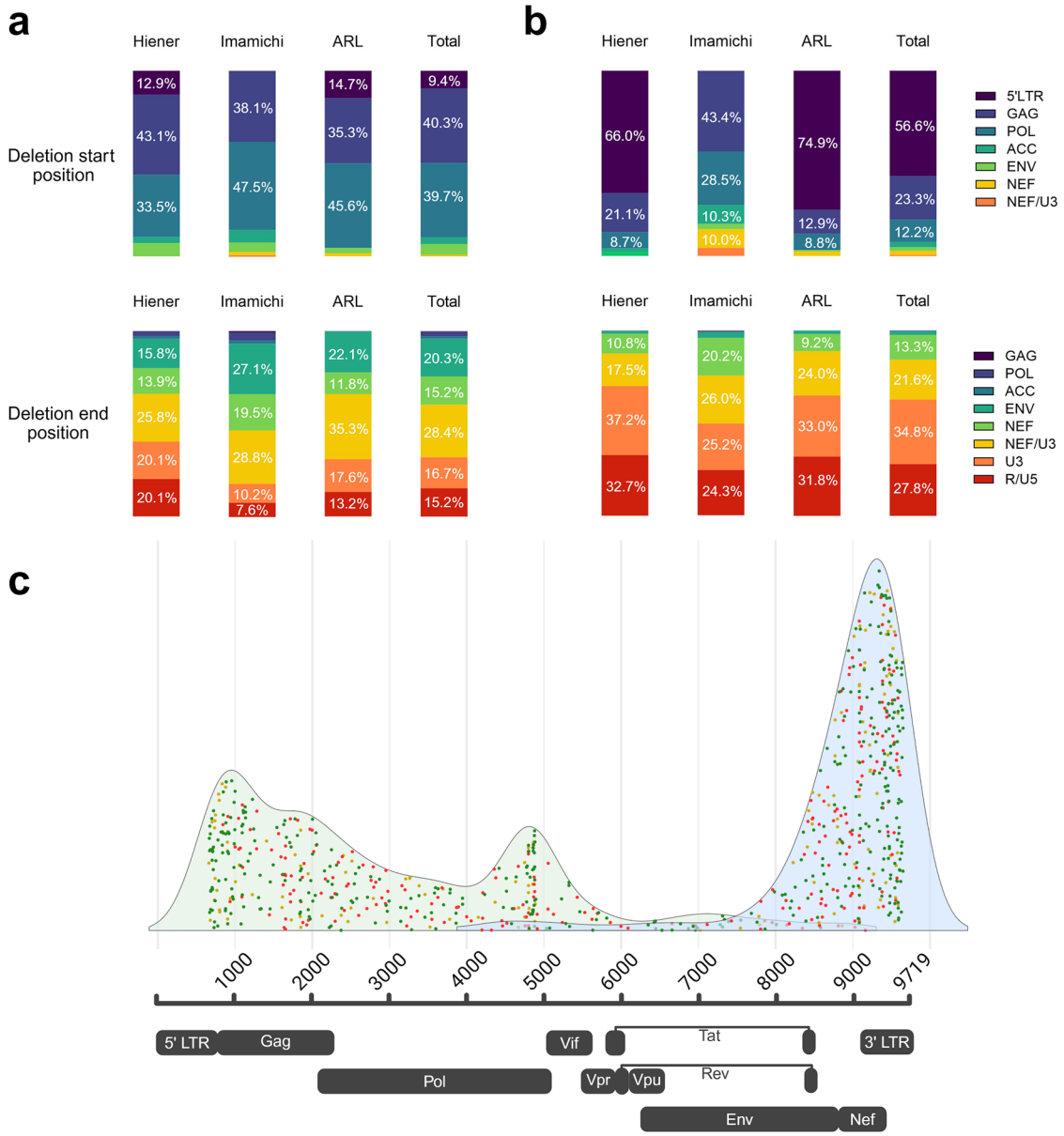


Fig. 2 | Analysis of the start and end positions of the deletion junctions in 395 FLIPS-generated sequences. Start and end positions were allocated to their respective genomic regions for each dataset individually and combined. Percentages represent the proportion of deletion joints assigned to each genomic region for the respective datasets: Hiener ($n = 209$), Imamichi ($n = 118$), ARL ($n = 68$), and the combined dataset ($n = 395$). The proportions per genomic region are depicted **a** irrespective of, and **b** in proportion to the respective genomic region sizes. The 3' LTR was deliberately divided into R/U5 and U3, as these regions correspond to distinct mechanisms associated with large deletions in proviral sequences. **c** A distribution curve of the deletion start (green) and end (blue) positions is superimposed on a schematic representation of the HIV-1 genome. Each dot represents a single provirus and is coloured according to the study from which it originates: Hiener (green), Imamichi (red), in-house ARL protocol (yellow). Created in BioRender. Hardy, J. (2025) https://BioRender.com/5o6xc9v.

Figure 2c depicts the nucleotide positions of the deletion start and end positions. Deletion start positions are scattered throughout the 5' half of the genome but tend to accumulate in the 5' LTR. However, there is a second hotspot for 5' deletion breakpoints in the central region of the genome, specifically in the integrase coding region (Figs. 2c and 3a). Forty-one out of 395 (10.4%) deletion start sites are positioned in a 100-nucleotide region between nucleotide positions 4800 and 4900 (according to HXB2), near the cPPT. Nine of those amplicons even had deletion junctions that coincided with both the cPPT and 3' PPT. Deletion end positions are less scattered throughout the genome and occur primarily in the utter 3' end of the genome (Fig. 3b). Figures 2c, 3a, b further demonstrate that the proviral sequences with a single deletion of more than 200 nucleotides predominantly lack the 3' half of their genome. Based on the amino acid sequences, deleterious stop codons and frameshift mutations were identified to determine the integrity of the *gag* and *nef* ORFs within each provirus. Amino acid and sequence analysis demonstrated that 154 of the 395 (39.0%) defective FLIPS-generated sequences had an intact *gag* ORF, while 61 (15.4%) retained a functional *nef* ORF. Additionally, sequences derived from samples with detectable viral load were omitted from the dataset, and the prevalence of intact *gag* and *nef* ORFs was reassessed. Upon reassessment, 109 of 307 (35.5%) defective proviral sequences had intact *gag* ORFs, and 37 of 307 (12.1%) had intact *nef* ORFs, respectively. Of the 37 proviruses that encompassed an intact *nef* ORF, 11 also contained an intact *gag* ORF.

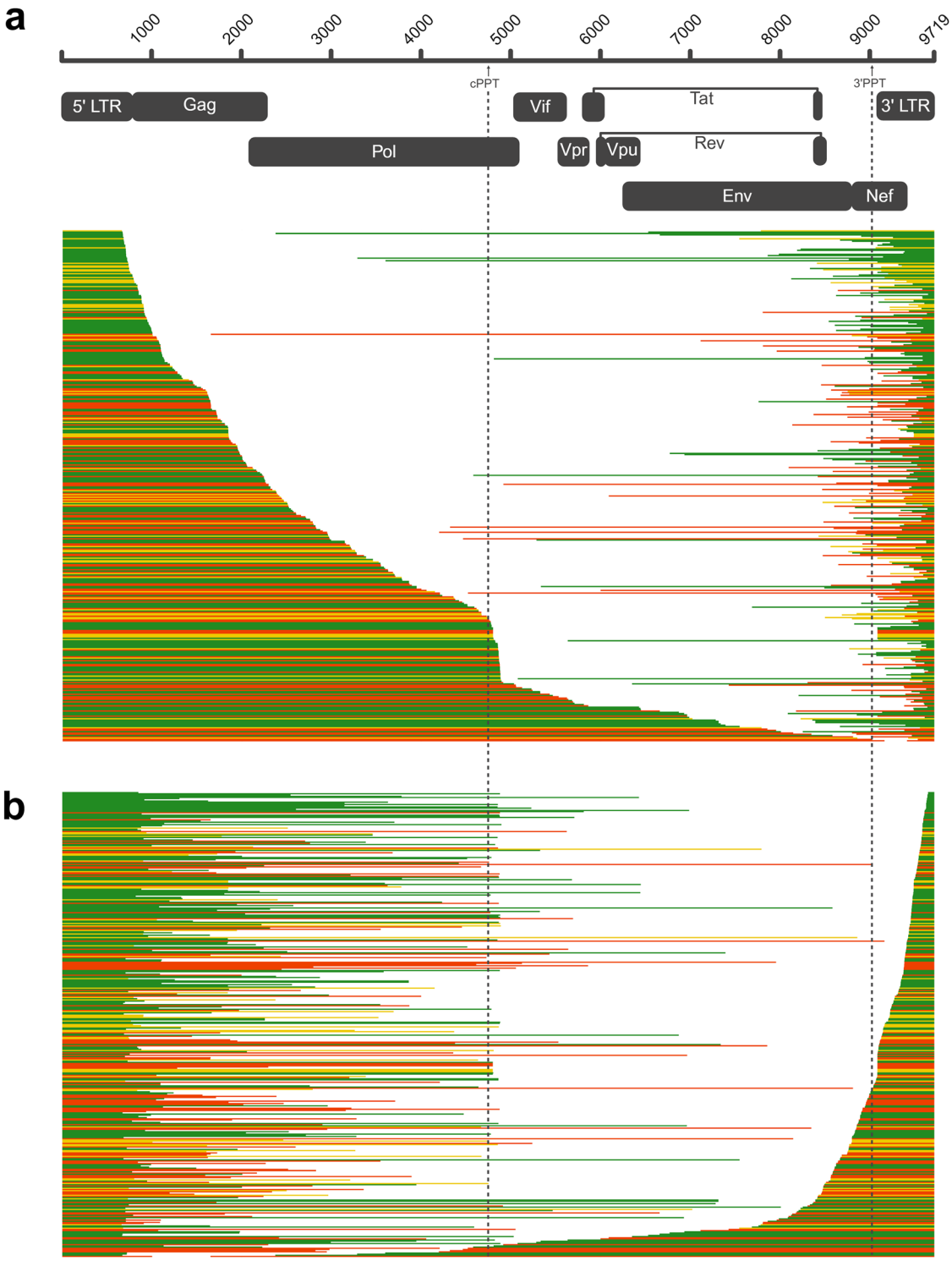
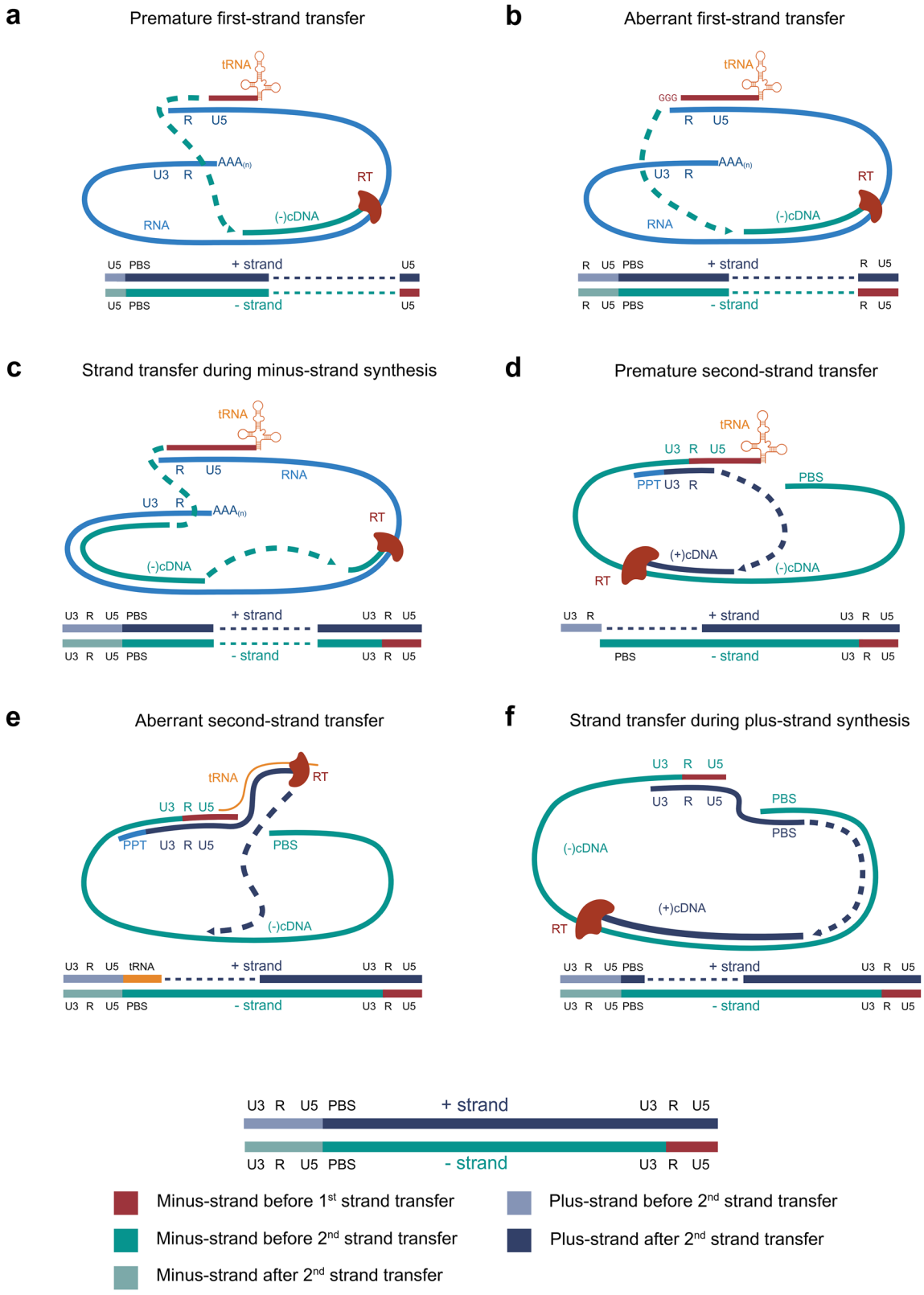


Fig. 3 | Proviral sequences aligned to the HIV HXB2 (K03455) reference strain. Sequences are arranged according to the start (**a**) and end (**b**) positions of each individual deletion junction. Each horizontal bar represents an individual proviral sequence generated through FLIPS and coloured according to the study from which it was derived: Hiener (green), Imamichi (red), in-house ARL protocol (yellow). The plots are accompanied by a schematic representation of the HIV-1 genome, the dotted lines highlight the positions of the cPPT and 3'PPT. Created in BioRender. Hardy, J. (2025) https://BioRender.com/wa00q4u.

Aberrant first-strand transfer

Proviral deletion junctions were matched to potential errors that may arise at different stages of the in vivo reverse transcription process. Minus-strand DNA synthesis initiates at the utter end of the 5' untranslated region of the viral RNA, followed by a first-strand transfer event facilitated by sequence identity between the R regions at both ends of the RNA template. Deletions that indicate erroneous events prior to or during first-strand transfer were observed in 61 of the 395 sequences (15.4%). In 7 of these 61 sequences, the deletion junction was found in the U5 region, indicative of a premature first-strand transfer, which gives rise to double-stranded proviral DNA lacking intact R regions at both ends (Fig. 4a). Additionally, 50 out of 61 sequences showed a deletion junction in the R region and 4 proviral sequences had a deletion junction that coincided with the



homopolymer G triplet defining the 5′ end of the R region, both suggesting an aberrant strand transfer during first-strand transfer. The latter are indicative of completed strong-stop minus-strand cDNA synthesis, followed by aberrant transfer of the minus-strand cDNA to a more downstream position on the RNA template (Figs. 4b and 5a).

Strand transfer during minus-strand synthesis

The most widely accepted mechanism for the generation of proviral DNA with large internal deletions is aberrant strand transfer during minus-strand DNA synthesis, following the completion of the first-strand transfer. Such erroneous strand transfer events will result in proviral DNA with intact LTR regions (Fig. 4c). We identified 334 sequences (84.6%) in which the deletion

Fig. 4 | Schematic representation of six mechanisms of aberrant strand transfer during reverse transcription. a Reverse transcription is initiated through the interaction between human tRNA(Lys-3) (orange) and the viral primer binding site on the RNA (light blue). Minus-strand cDNA (red: prior to first-strand transfer; green: post first-strand transfer) synthesis is disrupted and first-strand transfer occurs before the 5′ end of the viral RNA is reached. Premature first-strand transfers are characterized by a deletion junction in the U5 region. **b** Aberrant first-strand transfer events were identified either by a deletion junction within the R region or by a deletion junction coinciding with the homopolymer G triplet that defines the 5′ end of the R region. Aberrant first-strand transfers (**a**, **b**) give rise to double-stranded DNA (dsDNA) with incomplete LTR regions. **c** Following successful first-strand transfer, strand transfer events during minus-strand synthesis generate dsDNA with intact LTR regions and predominantly large internal deletions. **d** Second-strand synthesis initiates at the polypurine tract and proceeds to transcribe the 5' end of the plus-strand DNA (dark blue). As with premature first-strand synthesis, plus-strand synthesis can also be prematurely disrupted, giving rise to defective 5′ LTR regions. **d**–**f** Strand transfer events emerging during plus-strand synthesis result in mismatched double-stranded DNA molecules. **e** In contrast to premature second-strand transfer, second-strand transfer can also be delayed. In this scenario, reverse transcriptase begins transcribing fragments of tRNA(Lys-3) before switching to the minus-strand template. Although LTR regions remain intact in this scenario, the plus-strand carries an insertion that originates from the tRNA(Lys-3)-primer. **f** Strand transfer events during plus-strand synthesis represent an additional mechanism associated with the generation of internal deletions that, for technical reasons, could not be confirmed using the analysed dataset. All deletions are represented by dotted lines. Created in BioRender. Hardy, J. (2025) https://BioRender.com/78oz5fu.

junctions are indicative of this mechanism of origin. In 65 of these sequences (19.5%), the deletion end positions are found in the U3 region, while 112 (33.4%) sequences had deletion end sites in the U3/*nef* overlapping region. Deletion start positions were predominantly observed in *gag* (39.2%) and *pol* (38.6%).

Premature second-strand transfer

Proviral sequences were screened for indicators of errors arising during plus-strand DNA synthesis. Second-strand transfer is directed by sequence identity at the primer binding site (PBS). An aberrant second-strand transfer will result in plus-strand DNA lacking essential motifs of the 5' LTR (Fig. 4 d). These proviral sequences will be missed by the currently used FLIPS protocols due to the positioning of the sense primers. We therefore developed two alternative amplification protocols using sense primers positioned upstream of the FLIPS primers and antisense primers either just downstream of *pol* (5' PCR) or at the 3' end of the *nef* coding region (5' XL PCR) (Supplementary Data 4). The 5' PCR protocol generated 41 amplicons, of which 24 were intact (58%), 2 had a deletion of less than 200 nucleotides (4.8%), and 15 had a deletion of more than 200 nucleotides (36.6%). The 5' XL PCR protocol generated 19 amplicons, 11 of which were intact (57.9%) and 8 had a deletion greater than 200 nucleotides (42.1%). Of the 23 sequences with large deletions generated using these alternative PCR protocols, 20 (87.0%) showed deletion start sites positioned upstream of the PBS, indicative of a premature second-strand transfer (Supplementary Information 2 and Data 5).

Aberrant second-strand transfer

In one of the defective proviral sequences generated at the ARL, a non-HIV insert that corresponded to the human tRNA(Lys-3) sequence was observed at the deletion junction. This sequence was initially removed from the dataset due to the presence of a non-HIV sequence. Nevertheless, it was suggestive of a distinct aberrant plus-strand transfer mechanism, in which strand transfer was deferred and reverse transcription continued over the attached human tRNA(Lys-3) before transferring to a downstream position on the minus-strand DNA (Fig. 4e). The role of this mechanism in the formation of defective proviral sequences was further examined by searching the NIH genetic sequence database (GenBank) for sequence stretches that consisted of the 5' end of HIV-1 (5'-AAATCTCTAGCAG-3'), followed by the HIV PBS sequence (5'-TGGCGCCCGAACAGGGAC-3') and the first nucleotides of the tRNA(Lys-3) sequence (5'-TTGAACCCTGG-3'). This query returned 43 matches with ≥ 90% query coverage, including sequences from 9 different studies. All of these sequences were classified as defective HIV-1 proviruses[14,20,43–48]. The tRNA inserts defined the deletion junction and varied in length from 7 up to 54 nucleotides (Fig. 5b).

Discussion

A thorough understanding of the genomic landscape of the HIV-1 reservoir and its establishment will help refine methods for its characterization and quantification. Most proviruses in PLWH on suppressive ART contain internal deletions, which are generally thought to result from erroneous strand transfer during minus-strand synthesis. We conducted an in-depth analysis of deletion junctions in 418 proviral sequences with large internal deletions to better understand the mechanisms underlying their formation. Our results revealed that internal deletions may arise at different stages of the reverse transcription process. We modelled six aberrant strand transfer mechanisms based on the observed deletion patterns.

Length distribution curves of FLIPS-generated proviral sequences revealed a high prevalence of short amplicons, predominantly 2- and 5-kilobases in length. Although it is well established that proviruses with large internal deletions outnumber intact ones in PLWH on ART, the high prevalence of short amplicons might also be biased by the length-dependent PCR amplification rate, which selects against longer templates in a heterogeneous pool of proviruses[19]. Consequently, low amplification rates for long templates may reduce the median sequence length in the distribution curves.

Analysis of the two-kilobase amplicons indicated that they predominantly originated from either premature, aberrant first-strand transfer or minus-strand transfer occurring shortly after successful first-strand transfer. Aberrant first-strand transfer events result in proviruses with incomplete LTR regions, compromising their integration into the host genome. Integration requires 3' end processing of both viral DNA strands, which is considered highly sequence dependent[49]. Proviruses that derive from both models A and B lack the 5' LTR U3 terminus and, therefore, do not include the invariant CA dinucleotide that is recognized in 3'end processing. Attenuated or even lost integrase activity has been observed when mutations were introduced at the termini[50]. However, the complexes formed on either the U3 or U5 termini require only 15 base pairs and 3' end processing has been observed when the cleavage site is positioned a few nucleotides away from the DNA terminus[51]. Hence, despite the absence of the 5' LTR U3 terminus, integration may occur under certain conditions in which an alternative CA dinucleotide is positioned near to the erroneous 5' LTR U3 terminus. Oh et al. showed that, in the Rous sarcoma virus, there is no strict requirement for the canonical CA[52]. Oh et al. further demonstrated that once one end of the vDNA is integrated into the host genome by the integrase enzyme, the host machinery can efficiently integrate the other mutated or truncated end. In addition, Joseph et al. matched individual HIV proviral sequences to their respective 5'-adjacent host sequences and identified a 24-bp deletion at the 5' LTR U3 terminus, suggesting that integration of incomplete LTR regions can still take place[53]. However, integration of reverse transcription products that derive from models A and B cannot be demonstrated without sequencing the viral DNA ends and their matched integration sites.

In this study, erroneous first-strand transfer was identified as a key contributor to the formation of defective proviruses. It is important to note that the 3' LTR is not fully covered by the amplification protocols used. Consequently, the number of deletion junctions observed in the 3' LTR and, therefore, the role of premature and aberrant first-strand transfer in the formation of defective proviruses is most likely underestimated. Although sequence identity

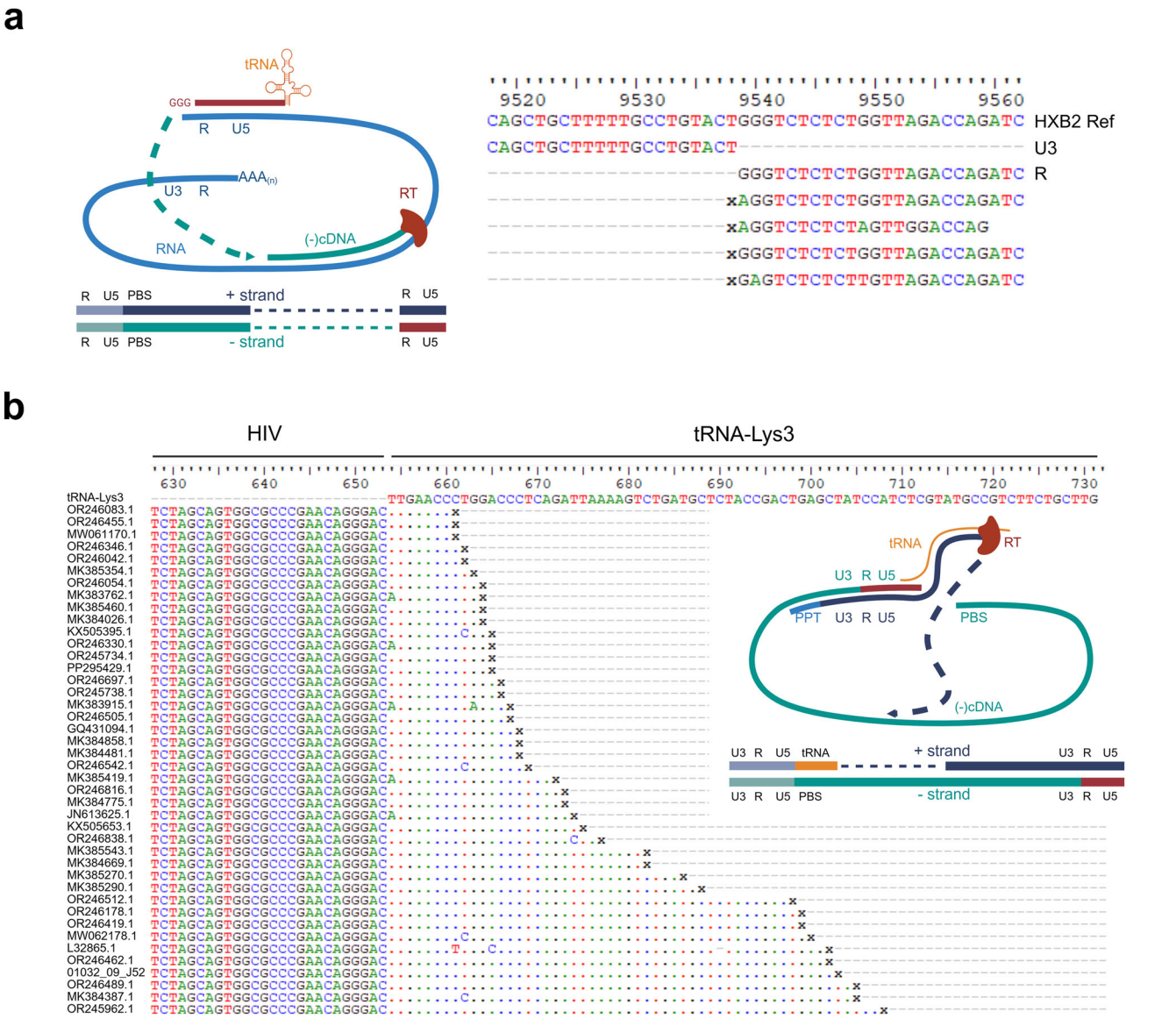


Fig. 5 | Sequences related to aberrant first- and second-strand transfer mechanisms. a Four proviral sequences showed deletion junctions that coincided with the homopolymer G triplet defining the 5' end of the R region. Both the proviral sequences and the U3 and R genomic regions are aligned to the HIV HXB2 (K03455) reference strain. **b** One of the proviral sequences, generated at the ARL (01032_09_J52), harboured an insertion of a fragment of human tRNA(Lys-3). Further evidence of aberrant second-strand transfer is provided by a search in GenBank (≥90% query coverage), which resulted in 43 similar cases. Genbank accession numbers are provided for each sequence. The shift from HIV to tRNA (Lys-3) is displayed by plotting the identities to the tRNA(Lys-3) sequence as dots. Each deletion junction is indicated with an 'x'. Created in BioRender. Hardy, J. (2025) https://BioRender.com/4ci7zzx.

between the R regions is believed to be essential for accurate first-strand transfer, strand transfer events have been observed even when the identity between acceptor and donor is limited[54–57]. Sequences with a deletion site coinciding with the homopolymer G triplet that defines the 5' end of the R region suggest an alternative mechanism for aberrant first-strand transfer. In this scenario, strong-stop minus-strand synthesis is completed, after which reverse transcriptase transfers to a downstream position. In previous work on tRNA-primed reverse transcription of plasma HIV-1 RNA, similar erroneous first-strand transfer events were identified in vitro[58]. Another mechanism that contributes to the high prevalence of short amplicons is minus-strand transfer occurring shortly after successful first-strand transfer. Deletion junction analysis showed that the LTR regions are the primary hotspots for erroneous first-strand transfer events. The tendency of reverse transcriptase to dissociate from its template has been attributed to enzyme stalling caused by secondary structures[27,59,60]. Watts et al. studied the secondary structures of HIV-1 and demonstrated that both LTR regions are highly structured regions[61].

In addition to the high prevalence of 2-kilobase amplicons, 5-kilobase amplicons appeared to be abundant. These 5-kilobase amplicons result from erroneous strand transfer during minus-strand synthesis and showed comparable deletion patterns associated with a deletion hotspot in the central region of the genome, near the cPPT. Nine of the 5-kilobase amplicons across the three studies even had deletion junctions that coincided with both the cPPT and the 3' PPT. These have been linked to inter- and intramolecular G-quartet structures that may stall reverse transcription and, consequently, facilitate strand transfer. G-rich hotspots for strand transfer events have been described in the *gag*, U3, and cPPT regions of the viral RNA genome and may facilitate interactions between these regions[62–64].

Apart from erroneous minus-strand transfer, our analysis indicates that plus-strand synthesis does not always unfold as intended. Premature

second-strand transfer events give rise to proviral sequences with incomplete 5' LTR regions and will often be missed by FLIPS due to primer design. We therefore designed primers to cover the entire 5' LTR and characterized 23 additional sequences with deletions of over 200 nucleotides, 19 of which would not have been detected by FLIPS. Hence, FLIPS fails to detect numerous defective proviruses.

A second aberrant plus-strand transfer mechanism, in which second-strand transfer is deferred and tRNA(Lys-3) is accidentally reverse transcribed into cDNA, was observed in a single sequence. A GenBank search revealed similar sequences, providing further evidence for the occurrence of this aberrant plus-strand transfer mechanism. Read-through transcription of tRNA(Lys-3) indicates that the m^1A modification at position 58 (m^1A58), which serves as the termination site of plus-strand synthesis, is not recognized. A second termination site, at nucleotide position 54 (m5U54), has been identified and proposed to act upon hypomodification of m^1A58[65,66]. Sequence analysis demonstrated that 5 out of 44 sequences with a tRNA(Lys-3) insertion had a U58A mutation, while none were observed at the second position. These findings indicate that reverse transcriptional readthrough of termination sites is more likely attributable to modification profiles than to sequence mutations. Nonetheless, reverse transcriptional read-through at the plus-strand termination leads to non-complementarity of the minus- and plus-strand, resulting in deficient strand transfer.

In Fig. 4f, we suggest an additional mechanism for plus-strand transfer leading to mismatched double-stranded DNA molecules. An issue to consider is whether the mismatched double-stranded DNA remains stable once integrated into the host genome. Under these conditions, it remains uncertain whether DNA damage responses step in or whether genomic instability leads to apoptosis. Unfortunately, the current amplification and sequencing strategy does not allow us to distinguish these aberrant strand transfer events from those occurring during minus-strand synthesis.

Sequences obtained from Imamichi et al. appear to have significantly shorter deletions than those from Hiener et al. and from the ones generated at the ARL. Samples across the three studies were collected from 18 PLWH and varied in both the time of infection prior to treatment initiation and the duration of treatment[16,18]. Although early initiation of treatment is associated with a smaller HIV DNA reservoir size in PLWH on long-term ART, it is not believed to substantially shape the genomic landscape of the reservoir[20,67,68]. The duration of treatment, however, will shift the population of HIV-infected cells. Multiple studies have shown that cells that harbour intact HIV genomes decay faster than those harbouring defective proviruses[44,69–72]. Consequently, we hypothesize that sequences with internal deletions from Imamichi et al., which included samples collected at time points when HIV-RNA levels exceeded the quantification limit of 40 copies per millilitre, showed shorter deletions due to differences in sampling. However, we must note that 27 of the sequences generated at the ARL from an ART-naïve individual did not significantly affect the median amplicon length. Differences in sequence length may also shape the distribution of the observed aberrant strand transfer mechanisms, although this effect could not be discerned in the performed analysis.

Under prolonged immunological pressure, proviruses that remain transcriptionally and translationally active can be selectively cleared[38,42]. Limiting-dilution single-genome HIV RNA sequencing demonstrated that as many as 7% of HIV-1 proviruses in peripheral blood mononuclear cells (PBMCs) from ART-treated individuals are transcriptionally active[73]. Our analysis demonstrates a high prevalence of intact *gag* and *nef* open reading frames in defective proviruses obtained from PLWH on suppressive ART. In addition to the high prevalence of intact reading frames, truncated and alternative ORFs may contribute to the production of peptides that induce CTL responses, leading to $CD8^+$ T cell dysfunction and systemic inflammation[74]. Given that the transcriptionally and translationally active proviruses are selectively targeted, a discrepancy between the prevalence of intact open reading frames between the original dataset and the reassessed dataset was anticipated. However, no substantial differences in the prevalence of intact *gag* and *nef* ORFs were seen upon reassessing the dataset. In addition, the intact gag ORF prevalence in the *nef*-intact subset is comparable to the overall *gag* prevalence, which indicates the lack of evidence of positive co-selection of the two open reading frames.

In conclusion, our findings demonstrate that the composition of the proviral DNA in the cellular reservoir is more complex than so far thought and that our current understanding of how the defective reservoir arises remains incomplete. This has implications for assays that aim to characterize the reservoir, as these assays were developed based on insights derived from superficial methods. Additionally, we highlighted the complex nature of the reverse transcription process and its role in the formation of proviruses with large deletions. Beyond the well-known aberrant strand transfer during minus-strand synthesis, we modelled five alternative aberrant strand transfer mechanisms.

Methods

Selection of defective proviral sequences

Proviral sequences generated by Hiener et al. (Accession numbers KY766150 to KY766212 and KY778264 to KY778681; 481 sequences) and Imamichi et al. (Accession numbers KU677989 to KU678195; 208 sequences) were retrieved from the NIH genetic sequence database, Genbank[16,18]. One hundred thirty-three additional proviral sequences, generated in-house at the ARL, were added to this dataset. ARL sequences originated from blood samples collected from three PLWH and stored in the ARL/ARC Biobank (ethical approval reference: BC-4364; ethical committee of Ghent University Hospital). Proviral sequences retrieved from Hiener et al. and Imamichi et al. were derived from PBMCs obtained from six and nine individuals, respectively. In the study by Imamichi et al., samples were collected at time points when HIV-RNA levels exceeded the quantification limit of 40 copies per millilitre in five of the nine individuals[18]. One of the individuals sampled at the ARL also had a detectable viral load due to being ART-naive. Altogether, 822 proviral sequences were assembled, to which the following selection criteria were applied (Supplementary Information 3). First, intact sequences, defined as sequences that are no more than 200 nucleotides shorter than the primer coverage length, were removed. Sequences with incomplete amplicon coverage or sequences with regions in reverse complement or non-HIV-related sequences were also discarded. The retained 538 sequences with a single internal deletion of more than 200 nucleotides, were aligned in BioEdit. Using the sequence identity matrix tool provided by the BioEdit software, clones of identical sequences were identified. For each clone, only one representative was retained in the alignment, leaving 418 unique defective proviral HIV DNA sequences for final analysis: 91 generated at the ARL, 209 from Hiener et al., and 118 from Imamichi et al. Among those 418 defective proviral sequences, 395 were generated using FLIPS. The residual 23 were generated at the ARL using alternative PCR protocols: either the 5'XL PCR (8 of 23 sequences) or the 5' PCR (15 of 23 sequences).

Sequencing HIV-1 Proviral DNA

Comparable near full-length amplification and sequencing techniques were used across the three studies. For the in-house ARL protocol, DNA was first extracted from PBMCs using the QIAamp DNA Blood Mini Kit, according to the manufacturer's instructions. A real-time PCR assay directed at the 5' LTR, as described in Avettand-Fenoel et al.[75], was used to determine the concentration of total HIV-1 DNA. Using absolute quantification, the endpoint dilution for single-genome sequencing was assessed. Near full-length PCR resulted in amplicons of 8926 nucleotides. Briefly, the 25 µL reaction mixture consisted of 12.5 µL of 2x Platinum SuperFi II PCR Master Mix, 0.2 µM of sense and antisense primers, and 5 µL of DNA. First-round PCR products were diluted ninefold with nuclease-free water before transferring 1.5 µL to the second-round PCR, which was identical to the first, except for the primers (Supplementary Data 4). PCR products were

purified with the QIAquick PCR Purification Kit (Qiagen) after gel electrophoresis confirmed positive reactions. Sequence analysis was prepared with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) before being transferred to the 3500 Genetic Analyzer (Applied Biosystems).

In addition to FLIPS, two other nested PCR approaches were developed to extend sequence coverage to the very 5' end of the proviral DNA. The 5' and 5'XL PCR protocols yielded amplicons of 5062 and 7668 nucleotides, respectively. Primers used in both PCR approaches are displayed in Supplementary Data 4.

Deletion, hypermutation, and ORFs

Sequences were aligned to the HIV HXB2 (K03455) reference strain using the MAFFT multiple sequence alignment tool (EMBL-EBI). The genomic positions of the internal deletion junctions were determined using the QuickAlign tool of the Los Alamos National Laboratory (LANL). APOBEC-induced hypermutation was detected using Hypermut 2.0 from LANL, with a p-value < 0.05 considered indicative of hypermutation. Each FLIPS-generated sequence was submitted to the GeneCutter tool of the LANL. Genecutter translated the open reading frames into amino acid sequences. Stop codons and mutations that caused frameshifts were identified using the GeneCutter tool.

Statistics and reproducibility

All statistical tests were performed in Graphpad Prism (10.4.0). A Kruskal-Wallis test, and an additional post-hoc Dunn's test, with the significance level set at 0.05, was used to assess the differences in sequence length across the three studies. Both the shift in the distribution curve after omitting the hypermutated sequences and the decision to retain those hypermutated sequences were assessed based on a Mann-Whitney U test, with the significance level set at 0.05.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data underlying this article is available both in the article itself and in its online supplementary material. HIV-1 sequence data that support the findings of this study have been deposited in Genbank with the following accession numbers: PX241873 to PX241970 (https://www.ncbi.nlm.nih.gov/genbank/). The source values underlying Figs. 1–3 can be found in supplementary data 1-3.

Received: 6 June 2025; Accepted: 20 October 2025;
Published online: 26 November 2025

References

1. Chun, T. W. et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* **94**, 13193–13197 (1997).
2. Finzi, D. et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
3. Wong, J. K. et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295 (1997).
4. Davey, R. T. Jr. et al. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl. Acad. Sci. USA* **96**, 15109–15114 (1999).
5. Finzi, D. et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
6. Siliciano, J. D. et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med.* **9**, 727–728 (2003).
7. Ho, Y. C. et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
8. Bruner, K. M. et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120–125 (2019).
9. Cassidy, N. A. J. et al. HIV reservoir quantification using cross-subtype multiplex ddPCR. *iScience* **25**, 103615 (2022).
10. Gaebler, C. et al. Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. *J. Exp. Med.* **216**, 2253–2264 (2019).
11. Delporte, M. et al. Integrative assessment of total and intact HIV-1 reservoir by a 5-region multiplexed rainbow DNA Digital PCR assay. *Clin. Chem.* **71**, 203–214 (2025).
12. Reeves, D. B. et al. Impact of misclassified defective proviruses on HIV reservoir measurements. *Nat. Commun.* **14**, 4186 (2023).
13. Kinloch, N. N. et al. Author Correction: HIV-1 diversity considerations in the application of the Intact Proviral DNA Assay (IPDA). *Nat. Commun.* **12**, 2958 (2021).
14. Gaebler, C. et al. Sequence evaluation and comparative analysis of novel assays for intact proviral HIV-1 DNA. *J. Virol.* **95** https://doi.org/10.1128/JVI.01986-20 (2021).
15. Simonetti, F. R. et al. Intact proviral DNA assay analysis of large cohorts of people with HIV provides a benchmark for the frequency and composition of persistent proviral DNA. *Proc. Natl. Acad. Sci. USA* **117**, 18692–18700 (2020).
16. Hiener, B. et al. Identification of genetically intact HIV-1 Proviruses in Specific CD4(+) T cells from effectively treated participants. *Cell Rep.* **21**, 813–822 (2017).
17. Lee, G. Q. et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J. Clin. Investig.* **127**, 2689–2696 (2017).
18. Imamichi, H. et al. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* **113**, 8783–8788 (2016).
19. White, J. A. et al. Measuring the latent reservoir for HIV-1: Quantification bias in near full-length genome sequencing methods. *PLoS Pathog.* **18**, e1010845 (2022).
20. Bruner, K. M. et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat. Med.* **22**, 1043–1049 (2016).
21. Mansky, L. M. & Temin, H. M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**, 5087–5094 (1995).
22. Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J. Virol.* **84**, 9864–9878 (2010).
23. Sebastian-Martin, A., Barrioluengo, V. & Menendez-Arias, L. Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases. *Sci. Rep.* **8**, 627 (2018).
24. Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harbor Perspect. Med.* **2** https://doi.org/10.1101/cshperspect.a006882 (2012).
25. Hwang, C. K., Svarovskaia, E. S. & Pathak, V. K. Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proc. Natl. Acad. Sci. USA* **98**, 12209–12214 (2001).
26. Svarovskaia, E. S., Delviks, K. A., Hwang, C. K. & Pathak, V. K. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J. Virol.* **74**, 7171–7178 (2000).
27. Roda, R. H. et al. Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. *J. Biol. Chem.* **277**, 46900–46911 (2002).
28. Jetzt, A. E. et al. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**, 1234–1240 (2000).

29. Schlub, T. E., Smyth, R. P., Grimm, A. J., Mak, J. & Davenport, M. P. Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput. Biol.* **6**, e1000766 (2010).
30. Zhuang, J. et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* **76**, 11273–11282 (2002).
31. Cromer, D., Grimm, A. J., Schlub, T. E., Mak, J. & Davenport, M. P. Estimating the in-vivo HIV template switching and recombination rate. *AIDS* **30**, 185–192 (2016).
32. Mangeat, B. et al. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
33. Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**, 1112 (2003).
34. Zhang, H. et al. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94–98 (2003).
35. Suspene, R. et al. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res.* **32**, 2421–2429 (2004).
36. Pace, C. et al. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J. Virol.* **80**, 9259–9269 (2006).
37. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
38. Imamichi, H. et al. Defective HIV-1 proviruses produce viral proteins. *Proc. Natl. Acad. Sci. USA* **117**, 3704–3710 (2020).
39. Singh, K. et al. Long-term persistence of transcriptionally active 'defective' HIV-1 proviruses: implications for persistent immune activation during antiretroviral therapy. *AIDS* **37**, 2119–2130 (2023).
40. Scrimieri, F. et al. Transcriptionally active defective HIV-1 Proviruses and their association with immunological nonresponse to antiretroviral therapy. *J. Infect. Dis.* **229**, 1786–1790 (2024).
41. Dube, M. et al. Spontaneous HIV expression during suppressive ART is associated with the magnitude and function of HIV-specific CD4(+) and CD8(+) T cells. *Cell Host Microbe* **31**, 1507–1522.e1505 (2023).
42. Pollack, R. A. et al. Defective HIV-1 Proviruses are expressed and can be recognized by Cytotoxic T Lymphocytes, which shape the proviral landscape. *Cell Host Microbe* **21**, 494–506.e494 (2017).
43. Lambrechts, L. et al. HIV-PULSE: a long-read sequencing assay for high-throughput near full-length HIV-1 proviral genome characterization. *Nucleic Acids Res.* **51**, e102 (2023).
44. Pinzone, M. R. et al. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat. Commun.* **10**, 728 (2019).
45. Liang, B. et al. A comparison of parallel pyrosequencing and sanger clone-based sequencing and its impact on the characterization of the genetic diversity of HIV-1. *PloS one* **6**, e26745 (2011).
46. Cho, Y. K., Jung, Y., Sung, H. & Joo, C. H. Frequent Genetic Defects in the HIV-1 5' LTR/gag Gene in Hemophiliacs Treated with Korean Red Ginseng: Decreased Detection of Genetic Defects by Highly Active Antiretroviral Therapy. *J. Ginseng Res.* **35**, 413–420 (2011).
47. Council, O. D. et al. The persistent pool of HIV-1-infected cells is formed episodically during untreated infection. *J. Virol.* e0097924 (2024).
48. Lu, C. L. et al. Relationship between intact HIV-1 proviruses in circulating CD4(+) T cells and rebound viruses emerging during treatment interruption. *Proc. Natl. Acad. Sci. USA* **115**, E11341–E11348 (2018).
49. Yoshinaga, T. & Fujiwara, T. Different roles of bases within the integration signal sequence of human immunodeficiency virus type 1 in vitro. *J. Virol.* **69**, 3233–3236 (1995).
50. Esposito, D. & Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17**, 5832–5843 (1998).
51. Vink, C., Van Gent, D., Elgersma, Y. & Plasterk, R. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* **65**, 4636–4644 (1991).
52. Oh, J., Chang, K. W. & Hughes, S. H. Integration of rous sarcoma virus DNA: a CA dinucleotide is not required for integration of the U3 end of viral DNA. *J. Virol.* **82**, 11480–11483 (2008).
53. Joseph, K. W. et al. Deep sequencing analysis of individual HIV-1 proviruses reveals frequent asymmetric long terminal repeats. *J. Virol.* **96**, e00122–e00122 (2022).
54. Pfeiffer, J. K. & Telesnitsky, A. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. *J. Virol.* **75**, 11263–11274 (2001).
55. Berkhout, B., van Wamel, J. & Klaver, B. Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. *J. Mol. Biol.* **252**, 59–69 (1995).
56. An, W. & Telesnitsky, A. Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J. Virol.* **76**, 7897–7902 (2002).
57. Dang, Q. & Hu, W. S. Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. *J. Virol.* **75**, 809–820 (2001).
58. Hardy, J. et al. Reverse transcription of plasma-derived HIV-1 RNA generates multiple artifacts through tRNA(Lys-3)-priming. *Microbiol. Spectr.* **12**, e0387223 (2024).
59. Lanciault, C. & Champoux, J. J. Pausing during reverse transcription increases the rate of retroviral recombination. *J. Virol.* **80**, 2483–2494 (2006).
60. Moumen, A. et al. Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA. *J. Biol. Chem.* **278**, 15973–15982 (2003).
61. Watts, J. M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
62. Piekna-Przybylska, D., Sharma, G. & Bambara, R. A. Mechanism of HIV-1 RNA dimerization in the central region of the genome and significance for viral evolution. *J. Biol. Chem.* **288**, 24140–24150 (2013).
63. Shen, W., Gao, L., Balakrishnan, M. & Bambara, R. A. A recombination hot spot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer in vitro. *J. Biol. Chem.* **284**, 33883–33893 (2009).
64. Harpster, C., Boyle, E., Musier-Forsyth, K. & Kankia, B. HIV-1 genomic RNA U3 region forms a stable quadruplex-hairpin structure. *Biophys. Chem.* **272**, 106567 (2021).
65. Fukuda, H. et al. Cooperative methylation of human tRNA3Lys at positions A58 and U54 drives the early and late steps of HIV-1 replication. *Nucleic Acids Res.* **49**, 11855–11867 (2021).
66. Wu, T., Guo, J., Bess, J., Henderson, L. E. & Levin, J. G. Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *J. Virol.* **73**, 4794–4805 (1999).
67. Jain, V. et al. Antiretroviral therapy initiated within 6 months of HIV infection is associated with lower T-cell activation and smaller HIV reservoir size. *J. Infect. Dis.* **208**, 1202–1211 (2013).
68. Buzon, M. J. et al. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. *J. Virol.* **88**, 10056–10065 (2014).
69. Peluso, M. J. et al. Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI Insight* **5** https://doi.org/10.1172/jci.insight.132997 (2020).

70. Gandhi, R. T. et al. Selective decay of intact HIV-1 Proviral DNA on antiretroviral therapy. *J. Infect. Dis.* **223**, 225–233 (2021).
71. Anderson, E. M. et al. Dynamic shifts in the HIV proviral landscape during long term combination antiretroviral therapy: implications for persistence and control of HIV Infections. *Viruses* **12** https://doi.org/10.3390/v12020136 (2020).
72. Lian, X. et al. Progressive transformation of the HIV-1 reservoir cell profile over two decades of antiviral therapy. *Cell Host Microbe* **31**, 83–96.e85 (2023).
73. Wiegand, A. et al. Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proc. Natl. Acad. Sci.* **114**, E3659–E3668 (2017).
74. Kuniholm, J., Coote, C. & Henderson, A. J. Defective HIV-1 genomes and their potential impact on HIV pathogenesis. *Retrovirology* **19**, 13 (2022).
75. Avettand-Fenoel, V. et al. LTR real-time PCR for HIV-1 DNA quantitation in blood cells for early diagnosis in infants born to seropositive mothers treated in the HAART area (ANRS CO 01). *J. Med. Virol.* **81**, 217–223 (2009).

Author contributions

Conceptualization: C.V., J.H.; Methodology: C.V., J.H., Y.D.M.; Data Collection: Y.D.M., Ciel.V.; Formal Analysis: C.V., J.H., Y.D.M.; Technical Support: D.S., E.D., M.S., Ciel.V.; Writing – Original Draft: C.V., J.H.; Writing – Review and Editing: C.V., J.H., V.M., E.P.; Supervision: C.V., E.P. All authors reviewed and agreed with the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-09103-7.

Correspondence and requests for materials should be addressed to Jarryt Hardy.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Harry Taylor and Tobias Goris. [A peer review file is available].