AIDS Reviews

PERMANYER

Check for updates

REVIEW ARTICLE

Integration site patterns, clonal expansion, and co-infection effects in HIV reservoirs: a systematic review of studies in people on antiretroviral therapy

*Bright Forkuo-Forkuo[1,2], John Atanniba-Atanniba[2], Evelyn Y. Bonney-Bonney[2], Richard Donkor-Donkor[3], Reuben Adjei-Adjei[4], Mary Amoakoh-Coleman-Amoakoh-Coleman[2], Kwamena W. C. Sagoe-Sagoe[3] and George B. Kyei-Kyei[2,5,6]**

[1]University of Ghana Medical School, Accra, Ghana; [2]Department of Virology, Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana; [3]Department of Microbiology, University of Ghana Medical School, Accra, Ghana; [4]College of Health Sciences, University of Ghana, Accra, Ghana; [5]Medical and Scientific Research Centre, University of Ghana Medical Centre, Accra, Ghana; [6]Department of Medicine and Molecular Microbiology, Washington University School of Medicine in St. Louis, St. Louis, USA

Abstract

HIV integrated into specific loci of the host's genome plays a significant role in clonal expansion and the persistence of the HIV reservoir. Certain factors, such as the HIV subtype, co-infections, inflammatory, and immune activation signaling, have been found to influence the integration sites (ISs). This systematic review examines IS patterns, clonal expansion, and the effects of co-infection on HIV reservoirs. A comprehensive search was conducted across four databases. We identified 18 articles published between 2016 and 2025 that met our criteria. We described the methods used for the IS analysis studies, features of the ISs, and their impact on clonal expansion. We also described the effects of cytomegalovirus (CMV), Mycobacterium tuberculosis (MTB), and influenza virus on the ISs and clonal dynamics. Eleven (61.1%) of the studies were conducted in the USA. Subtypes were mainly HIV-1 B. Ligand-mediated polymerase chain reaction or its variants were the most reported methods. Most HIV integration events occurred within genic regions, with reported proportions ranging from 77.0% to 100.0% of all sites. Most of the ISs were within introns, were in opposite orientation, occurred in cancer-associated genes, and influenced clonal expansion. CMV, MTB, and influenza did not significantly affect IS features but were associated with increased clonality. Expanded HIV-infected cell clones are integrated into genes regulating proliferation and immune signaling. Viral and immune co-factors may influence clonal persistence through other mechanisms. Understanding these will be critical for reservoir-targeted cure strategies.

Keywords: *HIV. Integration site. Clonal expansion. Co-infection. Subtype. Antiretroviral therapy.*

***Correspondence:**
George B. Kyei-Kyei
E-mail: g.kyei@wustl.edu

Received: 07-04-2026
Accepted: 22-09-2026
DOI: 10.24875/AIDSRev.26000005

Available online: 11-11-2026
AIDS Rev. (ahead of print)
www.aidsreviews.com

1139-6121 / © 2026 AIDS Reviews. Published by Permanyer. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Despite the success of antiretroviral therapy (ART) in inhibiting HIV replication, the virus persists in reservoir cells when ART is interrupted.[1] Understanding the source and characteristics of this persistent viral reservoir is critical for designing interventions for achieving cure. The sources of viral rebound upon ART interruption are thought to be due to the release of the virus from latent reservoirs in blood and sanctuary sites, and clonal expansion.[2] HIV-1 has tropism for activated CD4+ T cells,[3] but most of these cells die within a few days of infection, with a small fraction transitioning into latency.[4] There is evidence that a substantial portion of the reservoir results from the initial infection of memory cells.[5,6] It is now widely accepted that the proviruses in memory CD4+ T cells are maintained by clonal expansion.[1,4,7] Three main mechanisms for this clonal expansion have been identified: h omeostatic e xpansion, antigenic stimulation, and HIV integration site (IS) proliferation.[8] HIV-1-infected memory CD4+ T cells undergo homeostatic proliferation without inducing HIV-1 reactivation through homeostatic cytokines such as interleukin-7.[8] Antigenic stimulation of HIV-1-infected memory CD4+ T cells is driven by common antigens such as cytomegalovirus (CMV), Epstein–Barr virus, human papillomavirus, or bacterial products.[1,9] Chronic or repeated exposure to these antigens may contribute to the persistence of the HIV-1 reservoir via clonal expansion.[9,10] The antigen stimulation leads to activation and proliferation of the CD4+ T-cells into clones through Nuclear factor kappa B (NF-kb) and nuclear factor of activated T-cells (NFAT) signaling.[8] Antigenic stimulation may induce HIV-1 reactivation because the HIV-1 long terminal repeat (LTR) has binding sites for NF-kb and NFAT transcription factors.[8] Simonetti et al.[11] found that proliferation of infected cells under antigenic stimulation occurred regardless of the IS. They concluded that the response to antigens is the major driver of clonal expansion.

However, HIV integration plays a significant r ole i n clonal expansion and the persistence of the latent reservoir. Integration is not a stochastic process, and HIV prefers to insert its DNA into transcriptionally active, gene-dense regions of the human genome.[1,12] Certain host factors guide this targeting. CPSF6 binds to the HIV capsid and facilitates the transport of the HIV pre-integrating complex through the nuclear pore into the nucleus.[12,13] LEDGF/p75 binds to HIV integrase, guiding it to active transcriptional regions for integration.[12-14] HIV-infected cells with proviruses integrating at specific genomic loci appear to proliferate and expand clonally. Factors such as HIV subtype, co-infections, inflammatory and immune activation signaling, and epigenetic modification have been found to influence ISs.[8,14] Although several studies have examined HIV ISs, there has been no comprehensive synthesis of evidence on the patterns of ISs and their influence on clonality. This systematic review aims to bridge these gaps and identify critical directions for research relevant to an HIV cure. We explored methodological approaches used for IS mapping; characterized reported genomic features of HIV ISs in individuals on long-term ART; and determined how infections such as CMV and *Mycobacterium tuberculosis* (MTB) influenced ISs and clonality.

Methods

This study used a systematic approach in reviewing and synthesizing evidence on IS patterns, clonal expansion, and co-infection effects in HIV reservoirs in individuals on long-term ART. It used the PICOS (P-population, I-intervention, C-comparator, O-outcomes, and S-study) to define the population, intervention, control group, outcome of the exposure, and the study design to be included. The entire study was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.

We included studies involving adults (≥ 18 years) living with HIV and receiving ART. The studies should have performed HIV IS analysis and assessed clonal expansion or the effects of co-infections on these parameters. We included only case series (n ≥ 2) done within a 10-year period (2016-2025). We excluded non-human and cell-line experimental studies.

Search strategy and data synthesis

All relevant studies that report on the subject matter were searched from the following databases: PubMed, SCOPUS, Web of Science, and ScienceDirect. The keywords that were combined in search strings in the databases during the literature search are presented in the supplementary material. Boolean operators such as “AND” and “OR” were employed to refine the search. Mendeley Reference Manager was used to manage, remove duplicates, and collate the search output. References were imported into the Rayyan software for detailed review. The first screening was done by reviewing the study titles and abstracts. The full reports

for studies that passed the first screening were obtained and examined to see if they met the predetermined inclusion criteria. Inconsistencies among reviewers were resolved through discussions.

We employed the PRISMA Chart to illustrate the selection method for the articles that were included and the reasons for excluding others. We used a pre-tested extraction tool to ensure comprehensive information was obtained from each study. Information on the study's author(s), setting, aim, study design, methodology, population characteristics, and all relevant variables was extracted. A domain-based methodological quality assessment was performed to evaluate the technical and reporting robustness of included IS studies (Supplementary Table 2). Domains were defined a priori to reflect key steps in HIV IS recovery, sequencing, and analysis. Each domain was assessed as low concern, some concern, high concern, or unclear. We used the narrative synthesis method to summarize the findings on IS features. We described the methods used for the IS features and the pattern of clonal expansion (number of clones and clone size). We also reported on whether studies involved co-infections and their effects on the IS and clonal dynamics. We did not perform any meta-analysis. Since no field data will be collected, ethical approval was not required. The study has been registered with PROSPERO with registration number PROSPERO 2026 CRD420261290007.

Results

Study selection

A total of 437 studies published from 2016 to 2025 were retrieved from four electronic databases. After removing duplicates, 305 studies remained for screening. We excluded 241 studies based on irrelevant titles and abstracts, leaving 64 articles for full-text review. Of the 64 studies, we were unable to retrieve the full text for 2. Among the remaining 62 studies whose full text was retrieved, 44 were excluded, and 18 were included in this systematic review (Fig. 1).

Characteristics of included studies

The sample sizes range from 2 to 12. One study[15] included a cell line in their work, while another study,[16] included secondary IS analysis data in addition to data from their study participants. Eleven (61.1%) of the studies were conducted in the USA; two studies originated from Belgium and Switzerland, and one study each from China, Italy, and South Africa (Table 1). All the studies involved participants with HIV-1. Subtypes mainly consisted of HIV-1 B. Only one study indicated that it also included HIV-1 subtypes CRF02_AG and CRF63_02A1.[26] The median ART duration of participants in the included studies ranged from 4.7 to 13 years. Only five of the studies involved participants who were not virologically suppressed, either because they had not yet initiated ART or were under Analytical Treatment Interruption for comparison with those who were fully suppressed.[16,19,25,27] The remaining studies included only participants who were fully suppressed. We included one study that infected CD4+ cells and monocyte-derived macrophages (MDMs) *ex vivo* with autologous primary HIV-1 isolates that mimicked the *in vivo* situation.[23]

Methodological approaches for IS mapping

CD4+ T-cells isolation method

All the studies used peripheral blood mononuclear cells from the participants. One study collected both blood and tissue samples (from the tonsils) simultaneously from three of the nine participants.[30] Most studies (55.7%) employed magnetic-based negative selection or $CD8^+$ depletion kits to enrich $CD4^+$ T cells, including EasySep, RosetteSep, Miltenyi, and Dynabeads platforms. Four studies (22.2%) used fluorescence-activated cell sorting (FACS)-based sorting (BD FACSAria II/III) to obtain highly purified $CD4^+$ T cells. In four of the studies (22.2%), the cell isolation method was not reported (Table 2).

IS amplification methods

Ligation-mediated polymerase chain reaction (LM-PCR) or its variants were the most commonly reported methods, appearing in 10 (55.5%) of the studies included (Table 1). The non-restrictive LM-PCR (nrLAM-PCR) was reported in three studies,[20,22,23] whereas the linker-tagged amplification PCR, a linear target amplification variant, was also reported in one study.[21] IS loop amplification (ISLA), a loop-mediated, restriction-free approach, was employed in 4 (22.2%) of the studies. One study employed a multi-method pipeline (LM-PCR + nrLAM-PCR + ISLA) to maximize the IS recovery.[24]

The LM-PCR uses the whole-genome-amplified DNA from cells positive for proviruses, which is fragmented

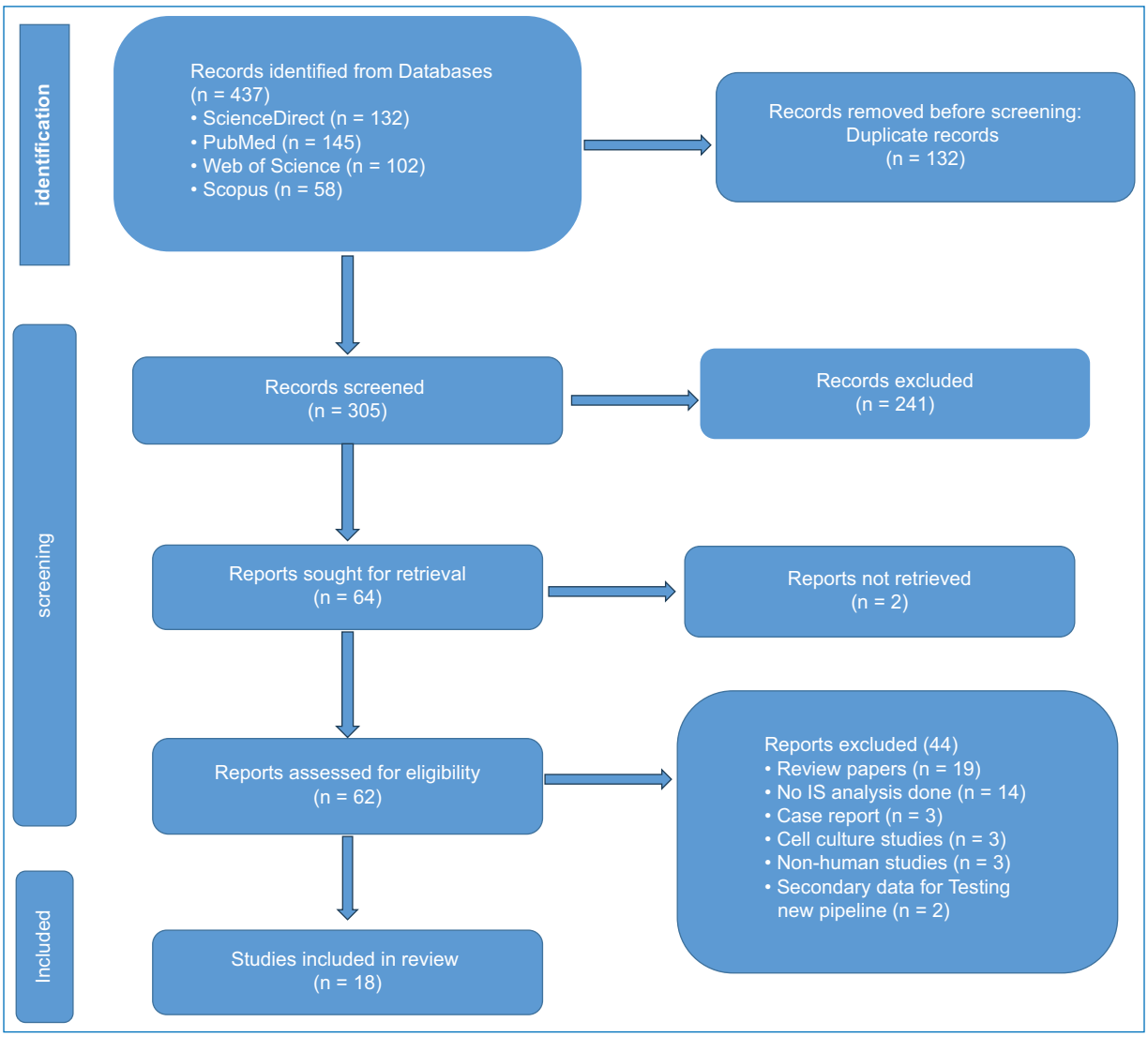


Figure 1. Preferred reporting items for systematic reviews and meta-analyses diagram of included studies.

either enzymatically or by sonication. DNA adapters are then ligated to the ends of the fragmented DNA. Nested PCR is then carried out between the adaptor and the LTR ends, and the PCR products are sequenced from both ends. The enzymatic method is subject to restriction enzyme bias, which may limit the detection of ISs located in regions lacking suitable restriction sites. The nrLAM-PCR methods begin with linear amplification using primers that bind to the viral LTR, followed by linker ligation and nested PCR. The nrLAM-PCR eliminates the need for restriction enzymes, thereby reducing site-selection bias and improving the recovery of proviruses in challenging genomic regions. The ISLA method uses linear priming and loop formation to efficiently amplify ISs.

Sequencing platform

Across the included studies, Illumina platforms were the most commonly used sequencing methods (Table 1). Fifteen of the included studies employed Illumina MiSeq, whereas one study used Illumina Nextera. Four studies used Sanger sequencing methods to verify the Illumina sequencing. One study used the Ion Torrent S5 platform for sequencing.[26]

Unique ISs

The number of unique IS ranged from 22 to 10523. Nine studies identified fewer than 500 unique IS (low yield), 2 identified between 500 and 1000 (moderate yield), 4 identified more than 1000 (high yield), and 3

Table 1. Characteristics of included studies

Studies	Country	Median ART (years)	HIV subtype	Sample size	IS amplification method	Sequencing platform	Metric for clonality
Cole et al., 2021[17]	Belgium	N/A	HIV-1 B	8	ISLA	Illumina MiSeq	1. Proviral sequencing. 2. TCR-Sequencing
Huang et al., 2021[18]	USA	13	HIV-1B	6	LM-PCR	Illumina MiSeq	1. Proviral sequencing 2. IS recurrence
Simonetti et al., 2021[11]	USA	8	HIV-1B	10	LM-PCR	1. Illumina MiSeq 2. Sanger	1. Proviral sequencing 2. IS recurrence
Halvas et al., 2020[19]	USA	11	HIV-1 B	4	LM-PCR	1. Illumina MiSeq 2. Sanger	1. Proviral sequencing 2. IS recurrence
Kok et al., 2021[20]	Switzerland	N/A	HIV-1B	12	nr-LAM PCR	Illumina MiSeq	IS recurrence
Brandt et al., 2021[15]	USA	10	HIV-1B	3	LM-PCR	Illumina MiSeq	N/A
Bale et al., 2021[21]	South Africa	8.1	HIV-1B	11	LTA-PCR	Illumina MiSeq	IS recurrence
Wang et al., 2025[22]	China	14.5	HIV-1B	N/A	nr-LAM PCR	Illumina MiSeq	N/A
Kok et al., 2016[23]	Switzerland	4.7	HIV-1B	7	1. LM-PCR 2. nr-LAM PCR 3. ISLA	Illumina MiSeq	IS recurrence
Einkauf et al., 2019[24]	USA	N/A	HIV-1B	3	LM-PCR	Illumina MiSeq	N/A
Sun et al., 2022[25]	USA	N/A	HIV-1B	4	LM-PCR	1. Illumina Nextera XT 2. Sanger	IS recurrence
Rozera et al., 2022[26]	Italy	N/A	1. HIV-1 B 2. HIV-1 C 3. CRF02_AG 4. CRF63_02A1	8	LM-PCR	Ion Torrent S5	N/A
Joy et al., 2024[27]	USA	10.1	HIV-1B	11	ISLA	Illumina MiSeq	N/A
Coffin et al., 2021[16]	USA	N/A	HIV-1B	2	LM-PCR	1. Illumina MiSeq 2. Sanger	N/A
Joseph et al., 2022[28]	USA	10	HIV-1B	5	LM-PCR	Illumina MiSeq	1. Proviral sequencing 2. IS recurrence
Einkauf et al., 2022[34]	USA	N/A	HIV-1B	7	ISLA	Illumina MiSeq	N/A
Wu et al., 2020[30]	USA	N/A	HIV-1B	9	LM-PCR	Illumina MiSeq	IS recurrence
Cole et al., 2022[31]	Belgium	N/A	HIV-1B	4	ISLA	Illumina MiSeq	1. Proviral sequencing 2. IS recurrence

N/A: not available; IS: integration site; USA: United States of America; ISLA: integration site loop amplification; nr-LAM-PCR: non-restrictive linear amplification mediated polymerase chain reaction; LM-PCR: ligand mediated polymerase chain reaction; LTA-PCR: linker-tagged amplification polymerase chain reaction; ART: antiretroviral therapy; TCR: T-cell receptor.

did not specify the number of unique IS identified. Overall, studies with smaller sample sizes and restriction enzyme-based PCR resulted in low unique IS yield, whereas those with relatively large sample sizes and untreated participants for comparison had high unique IS (Table 3).

Table 2. Distribution of CD4+ T-cells by isolation method

Isolation method category	Frequency (%)	Examples
Negative magnetic selection	9 (50.0)	EasySep, RosetteSep
$CD8^+$ depletion–based enrichment	1 (5.6)	Dynabeads, RosetteSep CD8
FACS sorting	4 (22.2)	FACSAria II/III (BD)
Not reported	4 (22.2)	

Integration patterns

Most HIV integration events occurred within genic regions, with reported proportions ranging from 77.0% to 100.0% of all sites. Intergenic integration was less frequent (9-20.0%) and generally appeared as random insertions relative to genome size. Within genes, proviruses were mainly enriched within introns, which accounted for 71.0-100.0% of genic insertions across studies. Exonic integration was uncommon, ranging from 0.0% to 8.0%. Ten out of the twelve studies that commented on integration orientation indicated that the main preference was for the opposite orientation. However, two studies found equal orientation preference.[25,26] In one of these studies,[26] it was observed that integration into certain genes (NRG2, B4GALT5, CNGB1, and FLT1) was always in the same orientation, whereas in other genes, such as TXNDC16, STUM, and GATAD2A, integration was always convergent.

Gene ontology analysis from various studies indicated that the ISs were broadly distributed. Multiple studies identified recurring hotspots, including genes such as BACH2, MKL1/2, STAT5B, and zinc-finger families. One study[23] identified three gene hotspots (MECP2, KANSL1, and UBE2G1) that appeared four or more times in both CD4+ T-cells and MDMs. ISs within chromatin modification regions consistently showed higher significant values in the Acute ART group compared to the acquired immune deficiency syndrome (AIDS)-ART group, especially in areas marked by H3K4ac, H3K9me3, H3K4me3, and CpG.[22] Hi-C analysis indicated that intact proviruses are less frequently located in active chromatin compared with defective proviruses.[18]

Prevalence and determinants of clonal expansion

Twelve of the included studies reported on clonal expansion (Table 1). Methods for reporting clonality varied among studies, including the use of clonal frequency, oligoclonality indices, or qualitative descriptions of expanded clones. Investigators also employed different metrics and definitions to characterize clonal expansion. Simonetti et al.[11] combined provirus sequencing and variable, diversity, and joining (VDJ) rearrangement of the T-cell receptor beta (TCRβ) to identify CMV-specific CD4+ clones. They found that in 3 clones from 3 participants, the proportion of clones containing the provirus $\left(\frac{provirus\,copies}{VDJ\,copies}\right)$ ranged from 14% to 72%.

Two of the included studies reported the clonal frequency, which refers to clones containing a particular provirus.[15,19] Using IS analysis, Halvas et al.[19] found the clonal frequency to range from 0.03% to 1.1%, estimating the total clonal burden in each participant to be between 50 million and 350 million cells. Brandt et al.[15] also observed a similar clonal frequency of about 0.03-1.08% using IS analysis. However, employing another technique called IS-quantitative polymerase chain reaction, they detected a relatively higher clonal frequency, ranging from 0.68% to 2.98%.

Coffin et al.[16] computed the amplification ratio (AR) $\left(\frac{total\,integration}{unique\,integration\,sites}\right)$ as a metric measurement for clonality. AR describes the average number of times an IS appears, indicating the degree of clonal expansion. They found that the overall AR across all sites in the entire genome was 1.6 for on-ART participants and 1.2 for pre-ART participants. The pre-ART value > 1 also indicates that cells can form clones before ART initiation.[16]

Some investigators described clonality before ART initiation (pre-ART) and during ART therapy (on-ART) for the same participants.[21] They defined a clone as ISs detected with > 1 breakpoint as well as those with > 2 breakpoints. They found the pre-ART median clonal sizes when defined as > 1 breakpoint and > 2 breakpoints to be 12 (range: 5-50) and 3 (range: 0-27), respectively. The on-ART median clonal size was 9 (range: 3-32), reported as > 1 breakpoint. They also reported on the oligoclonal index (OCI), which reflects the level of skewness of the entire clonal distribution in each sample. The median OCI pre-ART was 0.027 (range: 0.006-0.085), and that for on-ART was 0.161 (range: 0.055-0.403). This indicates that, upon ART initiation, the CD4+ T-cell population shows relatively greater skew toward dominant clones.

Table 3. Distribution of HIV IS yield across published studies

IS yield category	Range of unique IS	Number of studies	Typical study characteristics	Studies
Low yield	< 500	8	Small cohorts, restriction enzyme-based PCR	Cole et al., 2021[17]
				Huang et al., 2021[18]
				Simonetti et al., 2021[11]
				Einkauf et al., 2019[24]
				Sun et al., 2022[25]
				Rozera et al., 2022[26]
				Joseph et al., 2022[28]
				Cole et al., 2022[31]
Moderate yield	500-1000	3	Intermediate cohort size	Kok et al., 2021[20]
				Joy et al., 2024[27]
				Einkauf et al., 2022[29]
High yield	> 1000	4	Large cohorts and inclusion of untreated participants	Bale et al., 2021[21]
				Kok et al., 2016[23]
				Coffin et al., 2021[16]
				Wu et al., 2020[30]
N/A		3		Halvas et al., 2020[19]
				Wang et al., 2025[24]
				Brandt et al., 2021[15]

N/A: not available; PCR: polymerase chain reaction; IS: integration site.

Effects of co-infections on clonality and IS distribution

Two of the studies reported on the effects of CMV, MTB and/or influenza on either clonality or the distribution of ISs.[11,17] Simonetti et al.[11] found that the proviral populations in CMV-responding CD4+ T cells were dominated by a few expanded clones, compared with the control (polyclonal stimulation with CD3/CD28). They also reported a positive correlation between TCR clonality and HIV DNA ($r = 0.52$, $p = 0.01$), indicating that individuals with more clonally expanded T cells tend to have a more uneven HIV reservoir. This suggests that clonality increases with chronic antigen exposure, and the same biological force that drives T cell expansion may also shape how HIV-1 persists. They retrieved ISs from 22 expanded clones (16 CMV-responding and 6 gag-responding clones). They found that the features of ISs identified from the antigen-stimulated clones were similar to those they retrieved from participants on ART. Most of the ISs were within introns, were in opposite orientation, and occurred in cancer-associated genes.

Cole et al.[17] also found that TCRβ sequencing revealed clones with predicted pathogen-specificity. They identified a total of 43 distinct TCRβ sequences, with 8/43 (19%) of HIV-infected cells having specificity that could be predicted. When restricted to the clonal population, this proportion was 5/19 (26%). The predicted TCRβ were confined to CMV, MTB, and influenza. This also suggests that these co-infections increase clonality. They also found that these clones, with predicted TCRβ toward these pathogens, harbored proviruses that were integrated into genes associated with cancer.

Discussion

This systematic review synthesizes evidence on genomic features of HIV ISs, clonal expansion,

methodological approaches, and the influence of co-infections among individuals on long-term ART. We found that investigators employed several techniques to amplify the ISs from host genomic DNA. This usually reflected differences in sensitivity, bias, and workflow preferences. Each of these methods may have its advantages and disadvantages. The methods can be broadly categorized into three: Linker-mediated PCR-based approaches, non-restrictive LAM-PCR, and IS loop amplification.

Unique ISs varied across different studies. Variations in sample size, study participants, methods for amplifying ISs, and bioinformatic pipelines may explain these differences. This pattern indicates HIV's preference for large intronic regions and its tendency to avoid coding exons, potentially reducing disruption of vital host functions. The relative orientations of the provirus compared to the host transcription direction were also identified. Integration events were generally found in both orientations. However, the HIV provirus preferred to integrate in a direction opposite to host transcription. The HIV provirus prefers to insert into genes that regulate cell division and T-cell activation, and pro-cancer genes. The integration hotspots were found to frequently correspond to regions exhibiting clonal expansion, suggesting selective survival or proliferation of cells.[16] This indicated that the most common functional enrichments of genes from monoclonal unique IS included regulation of lymphocyte activation, regulation of cell processes, regulation of chromosome segregation, and organization. Significant differences in functional enrichment between the acute HIV and AIDS groups included regulation of lymphocyte activation, regulation of chromosome segregation and organization, and responses to biotic stimuli.[22] Cancer-related genes were overrepresented in HIV-1 integration across all clinical stages of infection.[17] Einkauf et al.,[24] also found that genes harboring proviral sequences frequently encoded T cell transcription factors or were involved in regulating T-cell behavior. However, there were no clear differences in the predicted functional profiles between genes hosting intact proviruses and defective proviruses. Although HIV is not considered a classic oncovirus, it is associated with oncogenesis. Integration of HIV into cancer-related genes can also contribute to oncogenesis by several mechanisms. Integration into cancer-related genes influences clonal expansion, which could result in accumulation of mutations over time. Other systemic effects of HIV infection, such as chronic infection and immunosuppression, could facilitate oncogenesis from such mutations.[32,33] Furthermore, proto-oncogenes may gain function, and tumor suppressor genes may also lose function upon HIV integration into those genes, contributing to oncogenesis. A study by Smith et al.[34] showed that HIV integration occurred commonly in the phosphatase and tensin homolog (PTEN) tumor suppressor gene in a patient with high-grade *de novo* neuroendocrine small cell lung cancer. HIV infected cells showed reduced PTEN expression and subsequent impaired antitumor immune expression, which may promote cancer progression. Yeh et al.[8] highlighted that HIV-1 integrated associated cancers are rare because malignant transformation occurs if the insertion occurs in a dominant gene. They further indicated that since malignant transformations require multiple hits, HIV-1 integration into cancer genes may not be sufficient to cause cancer.[8]

Six of the studies examined chromatin features of ISs. Integration was highly enriched in open chromatin, promoter-proximal regions, and areas marked by histone modifications associated with active transcription. There was a preferential enrichment of intact HIV-1 proviruses near active transcriptional units and accessible chromatin sites of the host.[24] They suggested that the latency of such cells could result from transcriptional interference between host and proviral gene expression, giving a selection advantage for intact proviruses during long-term ART.[24] Joy et al.[27] found that IS in the ART-chronic-HIV group were enriched in "NUYTTEN_NIPP1_TARGETS_UP," which is associated with histone methylation by EZH2. EZH2 is an H3K27 methyltransferase that leads to epigenetic modification that may maintain latency.[35,36] One study found that the progressive accumulation of genic proviruses integrated into hypermethylated chromosomal regions suggests silencing of proviral transcription during prolonged ART.[29] They also found that chromosomal distance between proviral IS and the nearest "same-oriented" transcription start site (TSS) increased over time; the trend was less apparent for the closest "opposite-oriented" TSS.[29] This implies that over time, cells with proviruses inserted closer to, and in the same orientation as, the TSS are selected against. This indicates a preferential selection for cells with proviruses inserted in opposite directions relative to the TSS, or in the same direction but farther from the TSS. This ensures the provirus remains latent and transcriptionally inactive. In addition, using transposase-accessible chromatin sequencing and chromatin immunoprecipitation followed by sequencing, the level of transcription for proviruses with the same orientation IS was found to decrease significantly over time.[29]

ISs were found to influence clonal expansion.[15,19,21] One study defined gene hotspots as genes identified more than once as sites of HIV integration.[23] The investigators found that enrichment of HIV integration within clonal expansion-associated gene hotspots was significant in latently infected CD4+ T cells but absent in MDMs, suggesting cell-type-specific differences in integration and persistence.[23] This suggests that the adaptive immune responses influence the expansion of HIV-infected clones. They investigated whether clones of those sizes could result from homeostatic proliferation by calculating the likelihood that a clone could reach that size. They found that the probability approaches zero, supporting the idea that the major driver of clonal expansion is a response to antigen rather than to independent homeostatic or IS -related proliferation.

The T-cell specificities of HIV-infected cells were limited to CMV, MTB, and/or influenza, suggesting that infection with or immunization against these pathogens (antigens) helped maintain the clones and contributed to HIV persistence. TCR specificity toward pathogens, with proviruses integrated into genes associated with cellular proliferation (STAT5, KCNA3), works together to support the continued existence of the translation-competent reservoir.[11,37]

Advances in gene targeting techniques are revolutionizing HIV cure research, further highlighting the need for HIV integration and genomics research. There are several ongoing trials on gene therapy for HIV cure.[38-41] Targeted gene editing uses the following strategies: Zinc finger nucleases, transcription activator-like effector nucleases, and the RNA-guided Clustered Regularly Interspaced Short Palindromic Repeats-/Caspases (g-RNA CRISPR/Cas).[42,43] CRISPR/Cas has proven to be the most effective gene editing method currently available.[43] It consists of the Caspase enzymes (molecular scissors), which are guided to the genomic target by guide RNA (about 20 nucleotides long).[44] Currently, the CRISPR/Ca9 gene editing is being used to achieve effective CCR5 ablation in addition to hematopoietic stem cell transplantation to achieve HIV cure.[40,41,44] Clinical translation is expected to be met with challenges such as editing efficiency, gene knockdown, and safety issues. Wang et al.[40] evaluated the efficacy and safety of several caspase systems and found that saCas9 demonstrated high antiviral activity. These systems inactivate the HIV provirus when designed with a single g-RNA or excise the proviral DNA with two g-RNAs. These g-RNAs are carefully designed to match certain sequences in viral DNA and thus guide the CRISPR/Cas enzyme to attach to the viral DNA, either inactivating or excising it.[41] These disrupt the virus's ability to replicate and can potentially lead to a cure. CRISPR/Ca13 has also been developed for HIV-1 RNA targeting.[38] Gene therapy for HIV cure should be personalized, using patients' proviral sequences to design the CRISPR-Cas systems.

Future directions

While this review offers a comprehensive synthesis of HIV IS biology, significant geographic gaps, particularly in sub-Saharan Africa, highlight the need for regionally representative, longitudinal studies to fully understand global HIV reservoir dynamics. There is therefore an urgent need for well-designed IS studies in the African population, especially in sub-Saharan Africa, where the HIV burden is high, incorporating locally prevalent subtypes.[45] Most published studies have relied on relatively small sample sizes, limiting statistical power and generalizability of their findings. Therefore, future primary studies should use larger cohort sizes to provide robust estimates and to allow for patient-level analysis. Future reviews should consider the depth and comprehensiveness of IS and reservoir analyses, such as sequencing depth, bioinformatic pipelines, and criteria for identifying and quantifying clones. Comparative evaluation and standardization of protocols for IS and reservoir methodologies will be important with respect to reproducibility, sensitivity, specificity, cost, and cross-study comparability. In addition, research should extend beyond circulating CD4+ T-cells to investigate other cellular and anatomical reservoirs that contribute to HIV persistence. Finally, future research should include more diverse clinical populations, including recent seroconverts, treatment-naïve individuals, elite controllers, virally suppressed individuals on ART, and individuals with a history of multiple virological failures. Including these diverse individuals will help to determine how disease stage, treatment history, and viral suppression influence HIV reservoir characteristics and IS patterns.

Conclusion

This systematic review synthesizes current evidence on HIV IS dynamics in individuals receiving long-term ART. The genomic distribution of ISs demonstrates a consistent preference for transcriptionally active genes, intronic regions, and specific genomic hotspots associated with cell survival pathways. Expanded HIV-infected

cell clones were common across studies and were frequently linked to integration into genes regulating proliferation and immune signaling. Although few studies evaluated co-infections, available evidence suggests that viral and immune co-factors may modify IS patterns and clonal persistence. Methodological differences across studies were substantial, with variation in PCR enrichment strategies, sequencing depth, and bioinformatic pipelines influencing resolution and comparability. Overall, these findings highlight both the biological complexity of the HIV reservoir and the need for standardized, high-coverage IS profiling methods. Understanding these patterns will be critical for reservoir-targeted cure strategies, particularly in high-burden settings.

Funding

This article is part of the H-Cure-RT project supported by the Fogarty International Center of the National Institutes of Health under Award Number D43TW012774. EYB was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under the Award Number R21AI174880. This work was also supported in part by the European Union/EDCTP (Grant No. TMA2017SF-1955, H-CRIS) awarded to GBK. The content is solely the responsibility of the authors and does not represent the views of the funders.

Conflicts of interest

None.

Ethical considerations

Protection of human subjects and animals. The authors declare that no experiments on humans or animals were performed for this research.

Confidentiality, informed consent, and ethical approval. This study does not involve personal patient data, medical records, or biological samples, and does not require ethical approval. SAGER guidelines do not apply.

Declaration on the use of artificial intelligence. The authors declare that no generative artificial intelligence was used in the writing or creation of the content of this manuscript.

Supplementary data

Supplementary data are available at 10.24875/AIDSRev.26000005. These data are provided by the corresponding author and published online for the benefit of the reader. The contents of supplementary data are the sole responsibility of the authors.

References

1. Anderson EM, Maldarelli F. The role of integration and clonal expansion in HIV infection: live long and prosper ben berkhout, alexander pasternak. Retrovirology. 2018;15:71.
2. Jacobs JL, Halvas EK, Tosiano MA, Mellors JW. Persistent HIV-1 viremia on antiretroviral therapy: measurement and mechanisms. Front Microbiol. 2019;10:2383.
3. Murray AJ, Kwon KJ, Farber DL, Siliciano RF. The latent reservoir for HIV-1: how immunologic memory and clonal expansion contribute to HIV-1 persistence. J Immunol. 2016;197:407-17.
4. Coffin JM, Hughes SH. Clonal expansion of infected CD4+ T cells in people living with HIV. Viruses. 2021;13:2078.
5. Murray JM, Zaunders JJ, McBride KL, Xu Y, Bailey M, Suzuki K, et al. HIV DNA subspecies persist in both activated and resting memory CD4+ T cells during antiretroviral therapy. J Virol. 2014;88:3516-26.
6. Cantero-Pérez J, Grau-Expósito J, Serra-Peinado C, Rosero DA, Luque-Ballesteros L, Astorga-Gamaza A, et al. Resident memory T cells are a cellular reservoir for HIV in the cervical mucosa. Nat Commun. 2019;10:4739.
7. Liu R, Simonetti FR, Ho YC. The forces driving clonal expansion of the HIV-1 latent reservoir. Virol J. 2020;17:4.
8. Yeh YJ, Yang K, Razmi A, Ho YC. The clonal expansion dynamics of the HIV-1 reservoir: mechanisms of integration site-dependent proliferation and HIV-1 persistence. Viruses. 2021;13(9):1858
9. Lau CY, Adan MA, Maldarelli F. Why the HIV reservoir never runs dry: clonal expansion and the characteristics of HIV-infected cells challenge strategies to cure and control HIV infection. Viruses. 2021;13(12):2512.
10. Mendoza P, Jackson JR, Oliveira TY, Gaebler C, Ramos V, Caskey M, et al. Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. J Exp Med. 2020;217:e20200051.
11. Simonetti FR, Zhang H, Soroosh GP, Duan J, Rhodehouse K, Hill AL, et al. Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. J Clin Invest. 2021;131:145254.
12. Singh PK, Bedwell GJ, Engelman AN. Spatial and genomic correlates of HIV-1 integration site targeting. Cells. 2022;11:655.
13. Bedwell GJ, Engelman AN. Factors that mold the nuclear landscape of HIV-1 integration. Nucleic Acids Res. 2021;49:621-35.
14. Vansant G, Chen HC, Zorita E, Trejbalová K, Miklík D, Filion G, et al. The chromatin landscape at the HIV-1 provirus integration site determines viral expression. Nucleic Acids Res. 2020;48:7801-17.
15. Brandt LD, Guo S, Joseph KW, Jacobs JL, Naqvi A, Coffin JM, et al. Tracking HIV-1-infected cell clones using integration site-specific qPCR. Viruses. 2021;13:1235.
16. Coffin JM, Bale MJ, Wells D, Guo S, Luke B, Zerbato JM, et al. Integration in oncogenes plays only a minor role in determining the in vivo distribution of HIV integration sites before or during suppressive antiretroviral therapy. PLoS Pathog. 2021;17:e1009141.
17. Cole B, Lambrechts L, Gantner P, Noppe Y, Bonine N, Witkowski W, et al. In-depth single-cell analysis of translation-competent HIV-1 reservoirs identifies cellular sources of plasma viremia. Nat Commun. 2021;12:3727.
18. Huang AS, Ramos V, Oliveira TY, Gaebler C, Janković M, Nussenzweig MC, et al. Integration features of intact latent hiv-1 in cd4+ T cell clones contribute to viral persistence. J Exp Med. 2021;218:e20211427.
19. Halvas EK, Joseph KW, Brandt LD, Guo S, Sobolewski MD, Jacobs JL, et al. HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. J Clin Invest. 2020;130:5847-57.
20. Kok YL, Vongrad V, Chaudron SE, Shilaih M, Leemann C, Neumann K, et al. HIV-1 integration sites in CD4+ T cells during primary, chronic, and late presentation of HIV-1 infection. JCI Insight. 2021;6:e143940.
21. Bale MJ, Katusiime MG, Wells D, Wu X, Spindler J, Halvas EK, et al. Early emergence and long-term persistence of HIV-infected T-cell clones in children. mBio. 2021;12:e00568-21.
22. Wang J, Xiao N, Zhu Z, Qiao H, Zhao F, Zhang L, et al. Comparing acute versus AIDS ART initiation on HIV-1 integration sites and clonal expansion. Signal Transduct Target Ther. 2025;10:23.
23. Kok YL, Vongrad V, Shilaih M, Di Giallonardo F, Kuster H, Kouyos R, et al. Monocyte-derived macrophages exhibit distinct and more restricted HIV-1 integration site repertoire than CD4(+) T cells. Sci Rep. 2016;6:24157.
24. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, et al. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J Clin Invest. 2019;129:988-98.

25. Sun C, Liu L, Pérez L, Li X, Liu Y, Xu P, et al. Droplet-microfluidics-assisted sequencing of HIV proviruses and their integration sites in cells from people on antiretroviral therapy. Nat Biomed Eng. 2022;6:1004-12.
26. Rozera G, Sberna G, Berno G, Gruber CE, Giombini E, Spezia PG, et al. Intact provirus and integration sites analysis in acute HIV-1 infection and changes after one year of early antiviral therapy. J Virus Erad. 2022;8:100306.
27. Joy J, Gervassi A, Chen L, Kirshenbaum B, Styrchak S, Ko D, et al. Antigen specificities and proviral integration sites differ in HIV-infected cells by timing of antiretroviral treatment initiation. J Clin Invest. 2024;134:e159569
28. Joseph KW, Halvas EK, Brandt LD, Patro SC, Rausch JW, Chopra A, Mallal S, Kearney MF, Coffin JM, Mellors JW. Deep sequencing analysis of individual HIV-1 proviruses reveals frequent asymmetric long terminal repeats. Journal of Virology. 2022 Jul 13;96(13):e00122-22.
29. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, et al. Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell. 2022;185:266-82.e15.
30. Wu VH, Nobles CL, Kuri-Cervantes L, McCormick K, Everett JK, Nguyen S, et al. Assessment of HIV-1 integration in tissues and subsets across infection stages. JCI Insight. 2020;5:e139783.
31. Cole B, Lambrechts L, Boyer Z, Noppe Y, De Scheerder MA, Eden JS, Vrancken B, Schlub TE, McLaughlin S, Frenkel LM, Palmer S. Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia before and during analytical treatment interruption. Cell reports. 2022 Apr 26;39(4).
32. Omar A, Marques N, Crawford N. Cancer and HIV: the molecular mechanisms of the deadly duo. Cancers (Basel). 2024;16:546.
33. Dandachi D, Morón F. Effects of HIV on the tumor microenvironment. Adv Exp Med Biol. 2020;1263:45-54.
34. Smith D, Rowland EF, Gianella S, Patel SP, Solso S, Dullano C, et al. HIV integration into the PTEN gene and its tumor microenvironment implications for lung cancer. Curr Oncol. 2025;32:389.
35. Friedman J, Cho WK, Chu CK, Keedy KS, Archin NM, Margolis DM, et al. Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of zeste 2. J Virol. 2011;85:9078-89.
36. Gramatica A, Miller IG, Ward AR, Khan F, Kemmer TJ, Weiler J, et al. EZH2 inhibition mitigates HIV immune evasion, reduces reservoir formation, and promotes skewing of CD8+ T cells toward less-exhausted phenotypes. Cell Rep. 2025;44:115652.
37. Coffin JM, Wells DW, Zerbato JM, Kuruc JD, Guo S, Luke B, et al. Clones of infected cells arise early in HIV-infected individuals. JCI Insight. 2019;4:e128432.
38. Ashraf A, Dahri AS, Ashraf MS, Kainat N, Haseeb M, Kiran I, et al. CRISPR-Cas13 for HIV-1 RNA targeting: advances, challenges, and NGS-driven insights. J Health Nutr. 2026;2:88-98.
39. Fattahi M, Pirbadami F, Motallebirad T, Beheshti Z, Azadi D. TMJ of I and, 2026 undefined. Next-generation gene therapy for infectious disease: advances, challenges, and future directions. J Infect Public Health. 2026;19:103164.
40. Wang JW, Liu JH, Xun JJ. CCR5 gene editing and HIV immunotherapy: current understandings, challenges, and future directions. Front Immunol. 2025;16:1590690.
41. Xu L, Yang H, Gao Y, Chen Z, Xie L, Liu Y, et al. CRISPR/Cas9- mediated CCR5 ablation in human hematopoietic stem/progenitor cells confers HIV-1 resistance in vivo. Mol Ther. 2017;25:1782-9.
42. Cetin B, Erendor F, Eksi YE, Sanlioglu AD, Sanlioglu S. Advancing CRISPR genome editing into gene therapy clinical trials: progress and future prospects. Expert Rev Mol Med. 2025;27:e16.
43. Zubair A, Sujan A, Ali M, Hussain SM. Current challenges with highly active antiretroviral therapy and new hope and horizon with CRISPR-CAS9 technology for HIV treatment. Chem Biol Drug Des. 2025;105:e70121.
44. Rehman SU. Revolutionizing HIV/aids treatment: advances in crispr/cas9. Braz J Pharm Sci. 2026;62:e24485.
45. Kharsany AB, Karim QA. HIV Infection and AIDS in sub-saharan Africa: current status, challenges and opportunities. Open AIDS J. 2016;10:l 34-48