

# The role of HIV integration in viral persistence: no more whistling past the proviral graveyard

Frank Maldarelli

Clinical Retrovirology Section, HIV Dynamics and Replication Program, National Cancer Institute, NIH, Frederick, Maryland, USA.

**A substantial research effort has been directed to identifying strategies to eradicate or control HIV infection without a requirement for combination antiretroviral therapy (cART). A number of obstacles prevent HIV eradication, including low-level viral persistence during cART, long-term persistence of HIV-infected cells, and latent infection of resting CD4<sup>+</sup> T cells. Mechanisms of persistence remain uncertain, but integration of the provirus into the host genome represents a central event in replication and pathogenesis of all retroviruses, including HIV. Analysis of HIV proviruses in CD4<sup>+</sup> lymphocytes from individuals after prolonged cART revealed that a substantial proportion of the infected cells that persist have undergone clonal expansion and frequently have proviruses integrated in genes associated with regulation of cell growth. These data suggest that integration may influence persistence and clonal expansion of HIV-infected cells after cART is introduced, and these processes may represent key mechanisms for HIV persistence. Determining the diversity of host genes with integrants in HIV-infected cells that persist for prolonged periods may yield useful information regarding pathways by which infected cells persist for prolonged periods. Moreover, many integrants are defective, and new studies are required to characterize the role of clonal expansion in the persistence of replication-competent HIV.**

## Introduction

Combination antiretroviral therapy (cART) has revolutionized the care of HIV-infected individuals (1). Although cART effectively suppresses virus replication and prevents morbidity and mortality from HIV-related complications, cART regimens do not eradicate infection, and continuous therapy is required to prevent rebound viremia and disease progression. Prolonged cART is associated with a number of systemic complications (2), and regimen nonadherence during therapy can be complicated by the development of drug resistance, resulting in progressively complex regimens and potential associated toxicity (3, 4). Consequently, a substantial research commitment has been made to eradicate HIV infection. “Cure” strategies may include those therapies that completely eradicate infection as well as approaches that result in control of viremia and arrest of disease progression without ART (5). One instance of HIV eradication (6, 7) in the setting of bone marrow transplant spurred research into pharmacologic and immunologic approaches, but gaps in understanding of HIV infection during cART prevent precisely targeted and highly effective curative therapies.

HIV persists during therapy, and populations of infected cells that give rise to infectious HIV (termed HIV reservoirs) have been recovered, particularly in resting CD4<sup>+</sup> T cells (8–14); however, the mechanisms of viral persistence remain incompletely understood. Recently, analysis of HIV in plasma and in infected cells revealed the emergence of clonal sequences after prolonged cART (15, 16). Advances in molecular technology and next-generation sequencing have provided the first detailed look at proviruses and the host genes into which they integrate. A number of reports have identi-

fied clonally expanded, HIV-infected CD4<sup>+</sup> T lymphocytes in individuals undergoing long-term cART (17–19), suggesting that clonal expansion may represent one mechanism of HIV persistence. The replication competence of HIV in these clonally expanded populations has not been extensively investigated, and the precise role of clonal expansion in maintaining a true HIV reservoir is under investigation. Sequence analysis of HIV in peripheral blood monocytes (PBMCs) is complex, as many infected cells harbor defective or deleted HIV genomes, and the complement of HIV proviruses in PBMCs is commonly described as a “graveyard” of integration events (20–22). Nevertheless, replication-competent HIV reservoirs are present among the defective clones, and new research indicates integration sites may drive clonal expansion and persistence of these cells during therapy (18, 19). This review describes HIV populations prior to and following introduction of cART, the central role of HIV integration in long-term persistence, the evidence for clonal expansion of HIV-infected cells, and current concepts regarding the role of clonal expansion in the persistence of replication-competent HIV.

## HIV infection establishes genetically diverse populations of infected cells

Rapid and error-prone HIV replication *in vivo* results in virus populations that are large, genetically diverse, and slow to undergo genetic change (23–30). In chronically infected individuals, genetic diversity, characterized as the average pairwise nucleotide differences of HIV variants, is substantial, in the range of 1.5% to 2%. At this level of diversity, essentially no two single genomes in plasma are identical. Introduction of cART interrupts HIV replication and results in a marked decline in viral RNA levels in plasma; decreases in viremia occur with multiphase decay kinetics that correspond to half-lives of virus-producing cells. Prior to cART initiation, more

**Conflict of interest:** The author has declared that no conflict of interest exists.

**Reference information:** *J Clin Invest.* 2016;126(2):438–447. doi:10.1172/JCI180564.

than 99% of virus is produced by cells that have a relatively short half-life (~1 day) (31, 32); this population corresponds to the number of activated CD4<sup>+</sup> T cells. As these cells are eliminated, viremia declines more slowly, and additional phases of decay can be detected. After 3 to 4 years of continuous cART, viral RNA levels remain relatively stable, although Mellors and colleagues recently detected ongoing decay of HIV viremia over time (33).

A variety of cell types support HIV infection, including cells in lymphocyte and monocyte lineages. In tissues such as the central nervous system, microglia and astrocytes have been reported to be infected with HIV. Macrophages, by virtue of their longevity despite infection, may contribute to long-lived cells and persistence following introduction of cART (34–37).

In contrast to the marked decline in HIV RNA, the decay of HIV-infected cells is less dramatic; only 8- to 30-fold declines in HIV DNA-positive cells are detected after initiating cART, and a stable level of cell-associated HIV DNA has been reported after the first 2 to 3 years of cART (38). Sensitive viral outgrowth assays (VOAs) to detect replication-competent HIV reveal very few cells with infectious HIV, and around one in a million resting cells is capable of producing infectious HIV after prolonged cART (9–12, 39). Many VOA studies have been performed with peripheral blood lymphocytes, but subtypes other than T cells may represent substantial sources of long-lived infected cells, which are not typically prevalent in PBMCs. Primate studies have directly identified these cells as critical for viral persistence (37).

The origin of low-level viremia detected during prolonged cART remains uncertain and is the subject of ongoing study; support for a number of potential sources has been described (39–45). Long-lived infected cells that produce HIV either on a chronic constitutive basis or episodically in response to an immune activation event can sustain low-level viremia. Alternatively, ongoing cycles of HIV replication with infection of new host cells may occur. The inability of cART to prevent such cycles of infection may be due to inadequate drug potency, insufficient drug penetration into tissues, replication in sanctuary sites, or chronic production from long-lived cells. In antiretroviral intensification studies, evidence of ongoing replication could be detected as a transient increase in 2-long terminal repeat (2-LTR) circular forms in some patient PBMCs (46, 47). No significant decreases in plasma HIV RNA were detected in intensification studies (46–50), suggesting that chronically infected cells rather than ongoing replication was the source of plasma viremia. Genetic analyses of HIV nucleic acids obtained from end-point dilution of plasma virions or cell-associated virus have yielded substantial data regarding the source and nature of persistence (15, 16, 51). Despite the profound decline in viremia, HIV genetic diversity persists at pretherapy levels, demonstrating that the population of HIV-producing cells remains large after cART initiation.

The composition of plasma virus populations undergoes detectable shifts over the course of therapy. Bailey and colleagues identified groups of identical sequences in these viruses, with some of these sequences accounting for over half of all single genomes detected (15). Such sequences were termed predominant plasma clones (PPCs) (49) and could also be detected in plasma (16, 52), cerebrospinal fluid (53), PBMCs (54–57), including resting CD4<sup>+</sup> cells (52, 58), various CD4<sup>+</sup> subsets (59, 60), and episomal HIV (61) as well as in various HIV-associated compartments, including

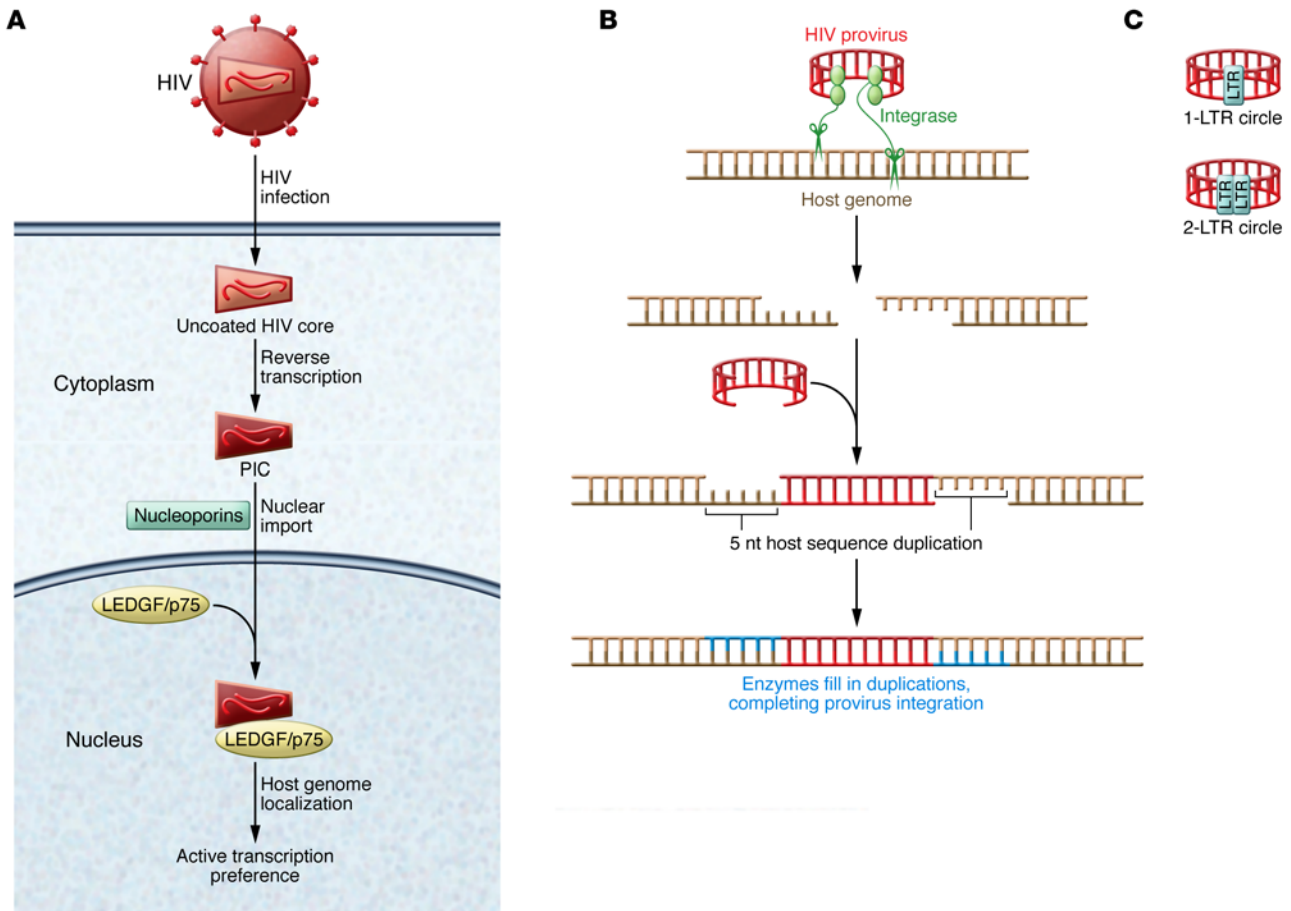
the gut (60, 62), genital tract (63), bronchial-alveolar lavage fluid (64), and breast milk (65). The source of these variants was uncertain, but PPCs were detected over long periods and were present in infected children (66) and in patients with low-level viremia (67). Comparisons with pretherapy virus populations revealed no evidence of molecular evolution in these PPC variants, even over 9 or more years of cART (16). Detection of multiple copies of identical sequences in individuals who initially had wide genetic diversity in plasma was unexpected and suggested that the individual lineages producing these viruses had undergone cell division. In all of these studies, it was unclear whether the identical sequences could be the result of long-lived cells from early in the course of HIV infection when the viral diversity is quite restricted and identical sequences are common. Studies of rebound viremia after treatment interruption revealed multiple populations with no evidence of molecular evolution compared with pretherapy virus (68). Many of these individual populations were present in multiple copies, but in these cases, it is likely that the identical sequences were the result of new and ongoing replication. In general, the replication competence of PPCs or other variants present in persistent viremia is unknown.

Further evidence supporting expansion of HIV-infected cells was reported in studies of cell-associated HIV DNA from individuals undergoing long-term cART. Sequence analysis of PBMC-derived HIV can be complex, as many proviral sequences are defective. Longitudinal studies of HIV population genetics often avoided analysis of these graveyards, as they included cells that may have been recently or remotely infected with HIV, precluding precise measurements of genetic change over time. Nevertheless, studies of HIV sequences in patients undergoing long-term therapy revealed examples in which multiple, identical copies of an HIV hypermutant were repeatedly recovered over prolonged periods (60, 69). As these HIV variants were clearly defective, they could not have persisted by ongoing cycles of infection and the only explanation for their persistence and recovery was that cells with these proviruses had undergone cell division. Further investigation of clonal expansion required new approaches to identifying and quantifying HIV integration events. The majority of these studies have analyzed HIV in plasma and in peripheral blood-derived lymphocytes. Tissue sources of HIV, which harbor the majority of HIV-infected cells (70), have not been extensively investigated.

## Integration: a central event in retrovirus replication

Like all retroviruses, HIV uses a virion-encoded reverse transcriptase (RT) enzyme to synthesize a double-stranded copy of DNA (denoted cDNA) from virion RNA templates (71), and integrates the viral DNA into the host genome using viral integrase (IN) (72). Retroviruses infect diverse vertebrate hosts and cause a wide spectrum of diseases (73); in all of these conditions, viral integration represents a critical step that is key for both virus replication and disease pathogenesis.

Reverse transcription takes place early after infection and generates a DNA copy of viral RNA with LTR sequences at the 5' and 3' ends. Upon completion of viral DNA synthesis, the nascent viral DNA is part of a large nucleoprotein complex known as the preintegration complex (PIC) (Figure 1 and refs. 72,



**Figure 1. Establishing the HIV provirus.** (A) HIV infects susceptible cells, and reverse transcription takes place early after infection, following attachment and fusion. Both RNA copies are required for replication, and multiple strand transfers between the two molecules occur, resulting in frequent recombination. The terminal repeats are duplicated at each end of the integration, and the product of reverse transcription is a full-length molecule containing LTRs. As a consequence of complex interactions and strand transfers, defective copies with large internal deletions (but intact 5' and 3' ends) are frequent products of reverse transcription. Reverse transcripts with HIV Gag p24 and IN compose the viral PIC, which is transported to the nucleus, a process requiring specific host proteins, including nucleoporins. In the nucleus, PIC structures interact with host protein LEDGF/p75, which preferentially facilitates integration in areas of transcriptional activity. (B) IN is a dimer of dimers that creates 5-bp staggered cuts in each DNA strand of the integration site, resulting in a 5-nt duplication of the host sequence for each integrant. Integration is completed by host enzymes, which fill in the staggered cuts. Integration sites are frequently mutated or internally deleted, but have precise integration sites with duplications of the host sequence at each end of the provirus. (C) Not all of the newly synthesized viral DNA molecules are integrated into the host genome; a subset of DNA molecules form circular molecules that include one or both LTRs (1- and 2-LTR circles).

74, 75), which includes an IN molecule bound at each of the 5' and 3' DNA ends. The DNA sequence specificity for IN binding to HIV DNA is conferred by nucleotides at each end, and not by size or internal sequence; consequently, full-length, mutant, and defective genomes are all substrates for integration (67). HIV infects nondividing cells, and nuclear import by the PIC is specific, requiring interactions between PIC components, including the HIV capsid, and cellular proteins, such as nucleoporins (76–82). Not all of the newly synthesized viral DNA molecules are integrated into the host genome; a subset of DNA molecules form circular molecules that include one or both LTRs (Figure 1). These LTR circles are the products of cellular ligase reactions and are likely to have long half-lives in infected cells, although they are not sources of transcription and cannot support productive or ongoing replication.

In general, integration site targets in the host genome are relatively nonspecific, but the virus exhibits site preferences in

vitro and in vivo (83). The first level of preference is mediated by cellular cofactors binding to IN as part of the PIC (Figure 1A). In HIV, the lens epithelial-derived growth factor (LEDGF; also known as p75) binds specifically to HIV IN, resulting in integration preferences for genes with transcriptional activity by tethering the DNA/IN/LEDGF complex (called the intasome) within areas of active transcription (84). Data from samples obtained from tissue culture infections as well as from infected animals do not reveal specific integration “hot spots,” but do identify general preferences for actively transcribed genes (85–88). A second level of preference is entirely localized within selection sequences and is based on where the IN enzyme itself binds to cut the host sequence. In HIV, 5'-AAT, 5'-TAA, and 5'-AAA sequences are enriched at the site of integration. This preference becomes part of the 5-bp repeat at each end of the integration (89). Other retroviruses exhibit distinct preferences (90), suggesting such preferences are IN specific.

**Table 1. Potential sources of HIV persistence**

	Decay kinetics/half-life	HIV replication competence	After long-term cART	
			Contribution to HIV reservoirs	Contribution to HIV-infected cell population
<b>HIV-producing</b>				
Activated	Phase 1/~1 d	+	No	No
Activated	Phase 1/~1 d	-	No	No
Intermediate	Phase 2/~14 d	+	Unlikely	Unlikely
Intermediate	Phase 2/~14 d	-	Unlikely	Unlikely
Long-lived	Phase 3,4/>39 wk	+	Yes	Yes
Long-lived	Phase 3,4/>39 wk	-	No	Yes
<b>Quiescent</b>				
Resting	yr	+	Yes	Yes
Resting	yr	-	No	Yes

A number of different cell types are infected with HIV, having variable half-life after infection. The majority (>99%) are short-lived, highly activated cells that do not directly represent a source of HIV persistence. Longer-lived, HIV-producing cells are eliminated slowly (intermediate) or only after prolonged periods and may contribute to long-lived reservoirs, but only if they are infected with replication-competent virus. Resting cells with infectious HIV represent a critical source of persistence. All long-lived cells with replication-competent or -incompetent virus will contribute to measures of cell-associated HIV.

## Integration site: a determining factor in viral pathogenesis

Integration site has important consequences for clonal expansion in retroviral pathogenesis. For example, human T-lymphotrophic virus-1 (HTLV-1) integrates into upstream regions, and several thousand distinct HTLV-1-infected cells are present in infected individuals. Many of these cells undergo clonal expansion and have viral integrations in genes associated with growth potential, but only a small subset of infected cells develop into oligoclonal T cell leukemias (91–93). In HTLV-2 infection, which does not cause leukemia or lymphoma, clonally expanded populations persist for prolonged periods without the development of neoplasia (94), indicating that clonal expansion events driven by integration do not universally result in the development of neoplasias. Similarly, MLV integration occurs near transcriptional start sites (TSS) of many genes and results in clonal expansion, but does not invariably result in neoplasia (95). In avian retroviral systems, viral integration is widespread, but expansions occur in tumor and nontumor tissues (96).

Integration sites also affect retroviral vectors used in gene therapy. Use of murine leukemia virus-based (MLV-based) vectors in gene therapy for SCID resulted in integration events near TSS of human genes. Integration events occurring near enhancers resulted in clonally expanded populations (97). Analysis of the leukemic cells revealed a selected population of genes with MLV integrations, suggesting leukemias were favored when these integrations occurred.

These findings reveal that clonal expansion is a characteristic of retroviral infections. As insertion sites with precise junction sequences accumulate, they have been catalogued for investigations of insertional mutagenesis and cancer (95, 98, 99). Such “retroviral-tagged genes” can be used to map specific metabolic and processing pathways that promote growth, thereby identifying potential points where pathways can be interrupted to prevent or treat neoplasia (95, 98, 100, 101).

## HIV integration in persistence and clonal expansion of infected cells

In the systems described above, where retrovirus infection does not typically result in cell death, retroviral integration commonly drives clonal expansion; however, it is not known whether similar events occur in HIV infection, where infection is frequently cytopathic. Many infected cells are eliminated after infection, but reservoirs of infected cells persist during therapy; therefore, it is critical to identify and quantify HIV integration sites, especially during cART, to determine whether specific sites are associated with persistence or clonal expansion. Initial studies (102–104) of integration sites in PBMCs from HIV-infected individuals undergoing cART were consistent with *in vitro* studies, and over 80% of integrants were identified in transcriptional units. Several individuals studied had integrations in a single gene (*BACH2*), but the mechanism of recurrent integration in this gene was uncertain, and it was not known whether this gene represents a true hot spot for integration or whether cells with *BACH2* integrations have a selective advantage favoring persistence (104). These early studies that identified integrants were technically superb and accurately identified junction sequences, but the experimental design was unable to determine whether any of the integrants were present multiple times and did not permit quantification of clonal expansion. Nevertheless, these studies confirmed *in vitro* studies, demonstrating that HIV integration was widespread in transcriptional units and was orientation independent.

Direct evidence for clonal expansion of HIV-infected cells was obtained in several studies of individuals on long-term cART using newer techniques, including adaptations of high-throughput, linker-mediated PCR assays (18, 92) and integration site loop amplification (19), both of which recover substantial numbers of integrants. These methods, reviewed in ref. 105, identify junction sequences and the host gene into which HIV has integrated, determine the orientation of integration, and readily distinguish



**Table 2. Mechanisms of persistence**

Immune mediated	Integration site mediated
Homeostatic proliferation	Host genes involved in cell growth regulation (orientation independent)
Antigenic stimulation	Specific, orientation-dependent sites
Generalized immune activation	

Immune mechanisms, viral integration site, or both may influence persistence and clonal expansion of HIV-infected cells. Homeostatic proliferation, antigen-directed expansion, or generalized immune activation may result in cell persistence and expansion. Some integrations occurring in host genes associated with cell growth will result in persistence and expansion; specific integrations, e.g., *BACH2* or *MKL2*, result in persistence and clonal expansion that is orientation specific.

between true clonal expansion and artifacts arising from PCR amplification. In these studies, repeat recovery of the identical junction sequence reflects clonal expansion, as the possibility of independent integrations into the identical nucleotide position in the human genome is highly statistically unfavorable. These studies, which generated over 3,100 integration sites for study, again demonstrated a bias of HIV for transcriptional units (around 79% of all integrants in either study), but no orientation preference with respect to host transcription (18, 19). The recovery of substantial numbers of integrants in combination with integrant copy number quantification revealed a number of new findings. (a) After prolonged cART, approximately 40% of all integrants had undergone clonal expansion; as the sample size is relatively modest compared with the infected cell population, this initial measurement is likely to be an underestimate. (b) The cells with HIV proviruses remaining after prolonged cART had integrations in diverse genes, but they were heavily biased toward genes involved in cellular growth or oncogenesis. Such biases were not present in analyses of HIV integration sites in *in vitro* infections, but were detected after prolonged cART as the likely result of cells with integrants in sites conferring a selective advantage for persistence, clonal expansion, or both. (c) Cells with integrants in particular genes, notably *BACH2* and *MKL2*, had a surprising pattern of integration. After prolonged cART, the remaining integrants, some of which were clonally expanded, were only detected in intron 5 of *BACH2* and intron 4 or 6 of *MKL2* upstream of the translational start site. In all cases, integrants were oriented in the same direction of transcription as the corresponding host gene. Analysis of prior studies of patients undergoing cART showed that integrants were similarly restricted within *BACH2* and were also oriented in the same direction of transcription as the host gene. Very recently, Cohn and colleagues used a modified linker-mediated approach and identified *BACH2* among the integration sites (17). In contrast to previous studies, their results implicated hot spots for integration near human *Alu* sequences, with less bias for integration in transcriptionally active genes.

*BACH2* is a basic region, leucine zipper family transcription factor that regulates transcription of both B and T lymphocytes through enhancer binding sequences. In T cells, *BACH2* contributes to T cell homeostasis (106, 107). In transgenic mouse models, loss of *BACH2* was associated with increases in effector T cell (Teff) function and loss of regulatory T lymphocyte (Treg) function (108). *BACH2*

has also been implicated in tumorigenesis and *BACH2-BCL2L1* and *BACH2-IGHD* fusions have been reported in lymphoma (109, 110). In mouse models of lymphoma, MLV integrated in *BACH2*, resulting in upregulation of alternatively spliced variants of *BACH2* in lymphomas (111). Finally, a series of observations suggest that loss of *BACH2* may be linked to cellular persistence. The EBV-transformed cell line Raji is a well-described B cell line with substantial growth potential. DNA analyses revealed that Raji cells have an integrated copy of EBV in intron 1 of *BACH2*, and transcriptome analyses detected a loss of *BACH2* expression in this cell line (112, 113). Reintroduction of *BACH2* expression in these cells results in a loss of clonogenic activity (114), suggesting downregulation of *BACH2* facilitated persistence, clonal expansion, or both.

*MKL2* is also a transcription factor that is broadly expressed in human tissues. Although less studied than *BACH2*, it has been identified as a serum response factor (SRF) coactivator and is therefore involved in cell growth (115–117). *MKL2* fusion oncogenes have also been identified, especially in chondroid lipomas, where *C11orf95-MKL2* fusions are often present (118, 119).

Taken together, these data suggest that the long-term fate of infected cells after cART is introduced is determined by intrinsic cell properties (e.g., activated vs. resting) and immune responses (homeostatic proliferation, antigenic response, or nonspecific immune activation mechanisms) as well as viral integration site (Table 1). The majority of infected cells have short half-lives after infection, are rapidly eliminated, and do not contribute to long-lived HIV reservoirs. Resting cells, or those responsible for low-level viremia, may persist for prolonged periods. The subset infected with replication-competent virus contributes to the population of infected cells representing the true HIV reservoir, but all proviruses contribute to the HIV-infected cell population. As previously described (21, 39, 120), long-lived cells infected with replication-competent HIV represent a very small proportion of the total infected cell population. Mechanisms driving persistence and clonal expansion include both immune functions and viral integration site (Table 2). The substantial bias that emerges over prolonged cART for cells with integrants in genes associated with cell growth indicates that integration is an important mechanism for persistence and clonal expansion of HIV-infected cells.

### New approaches to analyze HIV integrants

To date, next generation sequencing has provided only a short sequence of HIV with the junction sequence and host genome sequence; matching integrants to viral variants in plasma is possible, but is not readily accomplished with high-throughput sequencing approaches. Amplification with sequencing of longer HIV sequences will aid the tracking of HIV populations expressed during HIV infection.

With the advent of new methods to quantify the diversity of HIV-infected cells, additional statistical approaches will be needed to analyze alterations (or stability) in these populations over time, particularly with regard to integrant diversity and clonality during therapy. Bangham and colleagues developed an oligoclonality index (92, 94) to quantify clonal expansion of HTLV-1/2-infected populations, which will also likely be useful in analyzing

clonality of HIV-infected cells. With evidence of an increase in the number of integrants in host genes involved in cell growth and persistence, bioinformatic approaches that classify host cell gene function with integrants will be quite useful.

Further characterization of specific integrants, such as those integrated into *BACH2* and *MKL2*, is essential. Expanded populations with proviruses in *BACH2* or *MKL2* are striking, but they represent a relatively small proportion of all the proviruses in infected individuals and will likely be difficult to identify and study directly. Site-specific recombination approaches to introduce individual proviruses within genes of interest will provide tractable *in vitro* systems to study cellular functions. As integration sites are identified, it will be necessary to assemble these data into common, publicly available databases. Much groundwork for this process has already been provided for other retroviruses (121, 122), and Bushman, Shao, and coworkers have established interfaces (123, 124) that will allow investigators to take advantage of the increasing number of integration sites for *in silico* analyses with statistical power.

### Characterizing the role of clonal expansion in HIV infection

Data accumulated to date indicate that HIV-infected cells persist for prolonged periods and undergo clonal expansion; it remains uncertain to what degree clonally expanded populations contribute to the population of replication-competent cells, which represent the relevant reservoir that requires eradication or long-term control. In an exhaustive analysis of 213 uninduced proviruses, Ho and colleagues found that only 11.3% were full length (21). It is not known whether any of these proviruses were members of a clonally expanded population, but the data indicate that few integrants, expanded or not, are replication competent. Cohn and colleagues (17) analyzed 75 proviruses from expanded populations and did not identify any full-length HIV proviruses, suggesting the frequency of clonally expanded populations with replication-competent virus is low and may not contribute significantly to the HIV reservoir.

In contrast to the findings described above, direct demonstration of replication-competent HIV from a clonally expanded population has been reported (125). Persistent low-level PPC was identified, sequenced, and matched to an identical full-length provirus with all HIV reading frames intact in PBMCs; an unequivocal chromosomal assignment was not possible because the provirus was in an ambiguous region of the genome. The provirus was amplified from patient-derived PBMC DNA and found to be infectious in *in vitro* infection experiments. Moreover, infectious virus could be recovered from patient-derived PBMCs stimulated *in vitro*. These results indicate that clonally expanded HIV-infected cells can produce infectious HIV. The frequency of such infectious, clonally expanded cells is unknown, and the relative contribution of clonally expanded populations to the HIV reservoir remains to be determined.

Further analysis of proviral structures in highly expanded populations will be necessary to characterize the HIV reservoir. Initial studies detected complex deletions in expanded populations and many defective proviruses with internal defects (17, 21, 120). Integrants with intact promoter or other regulatory sequences (e.g., splice donor sequences) can have profound effects on host

genes, facilitating persistence and expansion without production of virus. The roles of HIV genes other than IN remain uncertain, and additional studies will determine contributions of HIV genes to persistence or expansion. Expression of HIV regulatory genes such as *Tat* and *Rev* does not require intact genomes and may affect expression of HIV and chimeric host:HIV transcripts. The HIV *Vpr* gene has been implicated in cell cycle arrest. While it may seem paradoxical that proviruses with *Vpr* may undergo clonal expansion, cell activation can occur without HIV transcription, and clonal expansion may occur in the absence of HIV transcription. Clonal expansion itself can affect proviral structure. During DNA replication, LTR sequences at each end of a provirus can undergo homologous recombination, resulting in deletion of internal sequences that yield highly deleted proviruses, including those consisting of a single LTR sequence. Analogous to numerous animal systems and endogenous retroviruses, where such “solo LTRs” have become common over millions of years of evolution (126, 127), all of these deleted proviral structures may influence persistence and expansion.

The studies reported thus far represent very early steps in the investigation of a role for clonal expansion in HIV persistence. Recent studies have identified integrants in a cross section of patients with HIV infection of variable duration. As HIV integrates broadly in transcribed genes, many more junction sequences will be necessary for comprehensive analysis of integration sites during prolonged therapy. The number of integrants recovered from patients (in the few thousands) is low relative to the number of integrants obtained from model systems or *in vitro* HIV infection, in which many thousands of junction sequences can be unequivocally identified. A larger compendium of integration sets will be essential to identifying whether the persisting populations have integrants in specific growth or apoptotic pathways after prolonged cART.

Further longitudinal data will be essential in understanding the contribution of clonally expanded populations to the total HIV integrant population. Interestingly, a number of studies have indicated gradual shifts in the genetic composition of HIV populations during prolonged periods on cART. Moreover, precise quantitative analysis by Besson (38) revealed that the relative levels of HIV proviruses in PBMCs (quantified using an internal IN sequence) were stable over time. These data suggest that the HIV population is dynamic, undergoing population shifts and clonal expansion, while the total amount of PBMC-associated HIV DNA remains relatively constant. Longitudinal studies of integrant distribution combined with quantification of relative levels of HIV integrants will help to determine whether maintaining a stable population of proviral DNA represents a steady state of expansion and elimination of variants. Investigations into the roles of specific immune responses and homeostatic proliferation will help elucidate the mechanisms of persistence of infected cells.

Analyses of a number of natural history cohorts will provide useful insights into the emergence of clonally expanded populations. For example, investigation of pre- and on-therapy PBMCs will help determine the degree of clonal expansion prior to introduction of antiretrovirals. Similarly, longitudinal analysis of acute infection, early infection, and perinatally acquired HIV-infected patients undergoing cART will yield critical information regard-

ing establishment of HIV-infected cells early in infection and their persistence during therapy. HIV-infected individuals with low CD4 cell numbers who initiate cART may experience immune reconstitution syndrome. Analysis of these individuals may help determine whether the exuberant immune response includes marked clonal expansion of HIV-infected cells. Individuals undergoing cART, but with incomplete reconstitution, represent an important subset of treated patients; mechanisms preventing robust CD4 reconstitution in these immune nonresponders is unknown, but early studies of a few individuals (18) indicate that immune nonresponders have potential for extensive clonal expansion. Comprehensive analyses of immune responders and nonresponders may yield useful insights into the mechanism of incomplete reconstitution.

In addition to natural history studies, specific interventional studies will yield critical and strategically useful information regarding clonal expansion during suppressive cART. A number of ongoing studies investigating the effects of latency reversing agents (LRAs) and other modalities to eliminate HIV reservoirs (128–133) may also benefit from analysis of clonal populations. In shock and kill strategies, LRAs induce HIV by activating infected, quiescent T cells. These cells presumably undergo cell death, but may undergo expansion as well, preventing eradication. Specific integration site data will be useful in identifying integrants that are activated, eliminated, or expanded during latency reversal. Sensitive integration assays are likely to represent useful adjunct analyses to characterize proviral populations in eradication strategies, especially as these approaches begin to incorporate multiple agents to activate and eliminate infected cells. The effects of preventative or therapeutic vaccination and other biologically active agents on clonal expansion will also be of interest in understanding immune function during cART.

Ultimately, data describing the persistence and expansion of cells containing HIV proviruses should be analyzed in the context of overall CD4 cell persistence and expansion. Studies analyzing

T cell repertoire will provide useful comparative analyses. It is not known whether extensive clonal expansions occur in the setting of a diverse or restricted repertoire; thus, analysis of integration sites may represent a tool for understanding the immunologic response as much as for characterizing the virology of the infection itself.

## Summary

Recent studies of HIV integration have been made possible by advances in the identification of HIV-host junction sequences and determination of the relative extent of clonal expansion of cells containing HIV proviruses. Initial analyses of clonally expanded populations demonstrate a substantial role for the integration site in shaping the population of cells persisting after prolonged cART. Detailed analysis of HIV in clonally expanded populations has revealed complex proviral structures with many deletions and insertions, but also replication-competent virus. Thus, the population of cells with integrated HIV is no longer just a “graveyard” of proviruses, analysis of which can be naively or conveniently avoided. Rather than whistling past this graveyard, analysis of proviruses, their integration sites, and their clonal expansion represents a new approach to characterizing HIV infection during cART, which will be useful in evaluating the effects of strategies to eradicate the infection in humans.

## Acknowledgments

I am grateful to the study participants who contributed to research described and to F.R. Simonetti, S.H. Hughes, J. Coffin, J. Mellors, M.F. Kearney, W. Shao, X. Wu, T. Uldrick, R. Yarchoan, C. Lane, M. Polis, and J. Kovacs for insightful discussions. This work was supported in part by the NIH Bench-to-Bedside Program.

Address correspondence to: Frank Maldarelli, Clinical Retrovirology Section, HIV Dynamics and Replication Program, National Cancer Institute, NIH, Building 535, Room 108E, 1050 Boyles St., Frederick National Cancer Laboratory, Frederick, Maryland 21702, USA. Phone: 301.846.5611; E-mail: fmalli@mail.nih.gov.

- Piot P, Legido-Quigley H. Global perspectives on Human Immunodeficiency Virus infection and Acquired Immunodeficiency Syndrome. In: Bennett JE, Dolin R, Blaser MJ, eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. 8th ed. Philadelphia, Pennsylvania, USA: Elsevier; 2015:1469–1482.
- Morse CG, Kovacs JA. Metabolic and skeletal complications of HIV infection: the price of success. *JAMA*. 2006;296(7):844–854.
- Calvo KR, Daar ES. Antiretroviral therapy: treatment-experienced individuals. *Infect Dis Clin North Am*. 2014;28(3):439–456.
- Solomon DA, Sax PE. Current state and limitations of daily oral therapy for treatment. *Curr Opin HIV AIDS*. 2015;10(4):219–225.
- Blankson JN, Siliciano JD, Siliciano RF. Finding a cure for human immunodeficiency virus-1 infection. *Infect Dis Clin North Am*. 2014;28(4):633–650.
- Yukl SA, et al. Challenges in detecting HIV persistence during potentially curative interventions: a study of the Berlin patient. *PLoS Pathog*. 2013;9(5):e1003347.
- Hutter G, et al. Long-term control of HIV by CCR5  $\Delta 32/\Delta 32$  stem-cell transplantation. *N Engl J Med*. 2009;360(7):692–698.
- Chun TW. Tracking replication-competent HIV reservoirs in infected individuals. *Curr Opin HIV AIDS*. 2013;8(2):111–116.
- Chun TW, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997;387(6629):183–188.
- Chun TW, Finzi D, Margolick J, Chadwick K, Schwartz D, Siliciano RF. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nat Med*. 1995;1(12):1284–1290.
- Finzi D, et al. Latent infection of CD4<sup>+</sup> T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med*. 1999;5(5):512–517.
- Finzi D, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997;278(5341):1295–1300.
- Siliciano JD, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4<sup>+</sup> T cells. *Nat Med*. 2003;9(6):727–728.
- Siliciano JD, Siliciano RF. A long-term latent reservoir for HIV-1: discovery and clinical implications. *J Antimicrob Chemother*. 2004;54(1):6–9.
- Bailey JR, et al. Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4<sup>+</sup> T cells. *J Virol*. 2006;80(13):6441–6457.
- Kearney MF, et al. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog*. 2014;10(3):e1004010.
- Cohn LB, et al. HIV-1 integration landscape during latent and active infection. *Cell*. 2015;160(3):420–432.
- Maldarelli F, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*. 2014;345(6193):179–183.
- Wagner TA, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014;345(6196):570–573.



20. Grant RM, Abrams DI. Not all is dead in HIV-1 graveyard. *Lancet*. 1998;351(9099):308–309.
21. Ho YC, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013;155(3):540–551.
22. Kesic M, Green PL. Retroviruses and insights into cancer: retroviral regulatory/accessory genes and cancer. In: Dudley J, ed. *Retroviruses And Insights Into Cancer*. New York, New York, USA: Springer; 2011:163–190.
23. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev*. 2012;76(2):159–216.
24. Maldarelli F, et al. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *J Virol*. 2013;87(18):10313–10323.
25. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet*. 2004;5(1):52–61.
26. Rouzine IM, Coffin JM. Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc Natl Acad Sci U S A*. 1999;96(19):10758–10763.
27. Rouzine IM, Coffin JM, Weinberger LS. Fifteen years later: Hard and soft selection sweeps confirm a large population number for HIV in vivo. *PLoS Genet*. 2014;10(2):e1004179.
28. Rouzine IM, Rodrigo A, Coffin JM. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol Mol Biol Rev*. 2001;65(1):151–185.
29. Coffin JM. Genetic diversity and evolution of retroviruses. *Curr Top Microbiol Immunol*. 1992;176:143–164.
30. Coffin JM. HIV viral dynamics. *AIDS*. 1996;10(suppl 3):S75–S84.
31. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995;373(6510):123–126.
32. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996;271(5255):1582–1586.
33. Riddler SA, et al. Continued slow decay of the residual plasma viremia level in HIV-1-infected adults receiving long-term antiretroviral therapy [published online ahead of print September 2, 2015]. *J Infect Dis*. doi:jiv433.
34. Verollet C, Le Cabec V, Maridonneau-Parini I. HIV-1 infection of T lymphocytes and macrophages affects their migration via Nef. *Front Immunol*. 2015;6:514.
35. Abbas W, Tariq M, Iqbal M, Kumar A, Herbein G. Eradication of HIV-1 from the macrophage reservoir: an uncertain goal? *Viruses*. 2015;7(4):1578–1598.
36. Crowe S, Zhu T, Muller WA. The contribution of monocyte infection and trafficking to viral persistence, and maintenance of the viral reservoir in HIV infection. *J Leuk Biol*. 2003;74(5):635–641.
37. Igarashi T, et al. Macrophage are the principal reservoir and sustain high virus loads in rhesus macaques after the depletion of CD4<sup>+</sup> T cells by a highly pathogenic simian immunodeficiency virus/HIV type 1 chimera (SHIV): Implications for HIV-1 infections of humans. *Proc Natl Acad Sci U S A*. 2001;98(2):658–663.
38. Besson GJ, et al. HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. *Clin Infect Dis*. 2014;59(9):1312–1321.
39. Siliciano JM, Siliciano RF. The remarkable stability of the latent reservoir for HIV-1 in resting memory CD4<sup>+</sup> T cells. *J Infect Dis*. 2015;212(9):1345–1347.
40. Maldarelli F. Targeting viral reservoirs: ability of antiretroviral therapy to stop viral replication. *Curr Opin HIV AIDS*. 2011;6(1):49–56.
41. Palmer S, Josefsson L, Coffin JM. HIV reservoirs and the possibility of a cure for HIV infection. *J Intern Med*. 2011;270(6):550–560.
42. Siliciano RF, Siliciano RF. Recent developments in the search for a cure for HIV-1 infection: targeting the latent reservoir for HIV-1. *J Allergy Clin Immunol*. 2014;134(1):12–19.
43. Siliciano RF. What do we need to do to cure HIV infection. *Top HIV Med*. 2010;18(3):104–108.
44. Siliciano RF, Greene WC. HIV latency. *Cold Spring Harb Perspect Med*. 2011;1(1):a007096.
45. Archin NM, Sung JM, Garrido C, Soriano-Sarabia N, Margolis DM. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat Rev Microbiol*. 2014;12(11):750–764.
46. Buzon MJ, et al. HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nat Med*. 2010;16(4):460–465.
47. Hatano H, et al. A randomized controlled trial assessing the effects of raltegravir intensification on endothelial function in treated HIV infection. *J Acquir Immune Defic Syndr*. 2012;61(3):317–325.
48. Dinoso JB, et al. Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proc Natl Acad Sci U S A*. 2009;106(23):9403–9408.
49. Gandhi RT, et al. No effect of raltegravir intensification on viral replication markers in the blood of HIV-1-infected patients receiving antiretroviral therapy. *J Acquir Immune Defic Syndr*. 2012;59(3):229–235.
50. McMahon D, et al. Short-course raltegravir intensification does not reduce persistent low-level viremia in patients with HIV-1 suppression during receipt of combination antiretroviral therapy. *Clin Infect Dis*. 2010;50(6):912–919.
51. Simonetti FR, Kearney MF. Review: Influence of ART on HIV genetics. *Curr Opin HIV AIDS*. 2015;10(1):49–54.
52. Anderson JA, et al. Clonal sequences recovered from plasma from patients with residual HIV-1 viremia and on intensified antiretroviral therapy are identical to replicating viral RNAs recovered from circulating resting CD4<sup>+</sup> T cells. *J Virol*. 2011;85(10):5220–5223.
53. Evering TH, Kamau E, St Bernard L, Farmer CB, Kong XP, Markowitz M. Single genome analysis reveals genetic characteristics of Neuroadaptation across HIV-1 envelope. *Retrovirology*. 2014;11:65.
54. Chun TW, et al. HIV-infected individuals receiving effective antiviral therapy for extended periods of time continually replenish their viral reservoir. *J Clin Invest*. 2005;115(11):3250–3255.
55. Lopez CA, et al. Characterization of HIV-1 RNA forms in the plasma of patients undergoing successful HAART. *Arch Virol*. 2010;155(6):895–903.
56. Wagner TA, McKernan JL, Tobin NH, Tapia KA, Mullins JI, Frenkel LM. An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J Virol*. 2013;87(3):1770–1778.
57. Buzon MJ, et al. Deep molecular characterization of HIV-1 dynamics under suppressive HAART. *PLoS Pathog*. 2011;7(10):e1002314.
58. Brennan TP, Woods JO, Sedaghat AR, Siliciano JD, Siliciano RF, Wilke CO. Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4<sup>+</sup> T cells reveals a novel source of residual viremia in patients on antiretroviral therapy. *J Virol*. 2009;83(17):8470–8481.
59. Chomont N, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med*. 2009;15(8):893–900.
60. Josefsson L, et al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc Natl Acad Sci U S A*. 2013;110(51):E4987–E4996.
61. Sharkey M, Babic DZ, Greenough T, Gulick R, Kuritzkes DR, Stevenson M. Episomal viral cDNAs identify a reservoir that fuels viral rebound after treatment interruption and that contributes to treatment failure. *PLoS Pathog*. 2011;7(2):e1001303.
62. Imamichi H, et al. Lack of compartmentalization of HIV-1 quasispecies between the gut and peripheral blood compartments. *J Infect Dis*. 2011;204(2):309–314.
63. Bull M, et al. Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations. *PLoS One*. 2009;4(9):e7122.
64. Heath L, et al. Evidence for limited genetic compartmentalization of HIV-1 between lung and blood. *PLoS One*. 2009;4(9):e6949.
65. Gantt S, et al. Genetic analyses of HIV-1 env sequences demonstrate limited compartmentalization in breast milk and suggest viral replication within the breast that increases with mastitis. *J Virol*. 2010;84(20):10812–10819.
66. Shiu C, et al. Identification of ongoing human immunodeficiency virus type 1 (HIV-1) replication in residual viremia during recombinant HIV-1 poxvirus immunizations in patients with clinically undetectable viral loads on durable suppressive highly active antiretroviral therapy. *J Virol*. 2009;83(19):9731–9742.
67. Tobin NH, et al. Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *J Virol*. 2005;79(15):9625–9634.
68. Joos B, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci U S A*. 2008;105(43):16725–16730.
69. Imamichi H, et al. Lifespan of effector memory CD4<sup>+</sup> T cells determined by replication-competent integrated HIV-1 provirus. *AIDS*. 2014;28(8):1091–1099.
70. Svicher V, Ceccherini-Silberstein F, Antinori A, Aquaro S, Perno CF. Understanding HIV compartments and reservoirs. *Curr HIV/AIDS Rep*.



- 2014;11(2):186–194.
71. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*. 2012;2(10):a006882.
  72. Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harb Perspect Med*. 2012;2(7):a006890.
  73. Goff SP. Retroviridae. In: Knipe DM, Howley P, eds. *Fields Virology*. 6th ed. Philadelphia, Pennsylvania, USA: Lippincott Williams and Wilkins; 2013:1424–1473.
  74. Engelman A. Isolation and analysis of HIV-1 preintegration complexes. *Methods Mol Biol*. 2009;485:135–149.
  75. Li M, Craigie R. Nucleoprotein complex intermediates in HIV-1 integration. *Methods*. 2009;47(4):237–242.
  76. Krishnan L, et al. The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *J Virol*. 2010;84(1):397–406.
  77. Lee K, et al. Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*. 2010;7(3):221–233.
  78. Matreyek KA, Engelman A. The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *J Virol*. 2011;85(15):7818–7827.
  79. De Iaco A, Luban J. Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus. *Retrovirology*. 2011;8:98.
  80. Di Nunzio F, et al. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology*. 2013;440(1):8–18.
  81. Hulme AE, Kelley Z, Foley D, Hope TJ. Complementary assays reveal a low level of CA associated with viral complexes in the nuclei of HIV-1-infected cells. *J Virol*. 2015;89(10):5350–5361.
  82. Wong RW, Mamede JI, Hope TJ. Impact of nucleoporin-mediated chromatin localization and nuclear architecture on HIV integration site selection. *J Virol*. 2015;89(19):9702–9705.
  83. Mitchell RS, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. 2004;2(8):E234.
  84. Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res*. 2014;42(16):10209–10225.
  85. Craigie R, Bushman FD. Host factors in retroviral integration and the selection of integration target sites. *Microbiol Spectr*. 2014;2(6):MDNA3-0026-2014.
  86. Engelman A, Cherepanov P. The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog*. 2008;4(3):e1000046.
  87. Ferris AL, et al. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci U S A*. 2010;107(7):3135–3140.
  88. Marshall HM, et al. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One*. 2007;2(12):e1340.
  89. Serrao E, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology*. 2015;12:39.
  90. Demeulemeester J, De Rijck J, Gijsbers R, Debyser Z. Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. *Bioessays*. 2015;37(11):1202–1214.
  91. Cook LB, Rowan AG, Melamed A, Taylor GP, Bangham CR. HTLV-1-infected T cells contain a single integrated provirus in natural infection. *Blood*. 2012;120(17):3488–3490.
  92. Gillet NA, et al. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood*. 2011;117(11):3113–3122.
  93. Niederer HA, Bangham CR. Integration site and clonal expansion in human chronic retroviral infection and gene therapy. *Viruses*. 2014;6(11):4140–4164.
  94. Melamed A, et al. Clonality of HTLV-2 in natural infection. *PLoS Pathog*. 2014;10(3):e1004006.
  95. Du Y, Jenkins NA, Copeland NG. Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. *Blood*. 2005;106(12):3932–3939.
  96. Justice JT, et al. The MET gene is a common integration target in avian leukosis virus subgroup J-induced chicken hemangiomas. *J Virol*. 2015;89(9):4712–4719.
  97. Dave UP, et al. Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy. *PLoS Genet*. 2009;5(5):e1000491.
  98. Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. RTDGD: retroviral tagged cancer gene database. *Nucleic Acids Res*. 2004;32(Database issue):D523–D527.
  99. Kustikova OS, et al. Retroviral vector insertion sites associated with dominant hematopoietic clones mark “stemness” pathways. *Blood*. 2007;109(5):1897–1907.
  100. Dave UP, Jenkins NA, Copeland NG. Gene therapy insertional mutagenesis insights. *Science*. 2004;303(5656):333.
  101. Du Y, Spence SE, Jenkins NA, Copeland NG. Cooperating cancer-gene identification through oncogenic-retrovirus-induced insertional mutagenesis. *Blood*. 2005;106(7):2498–2505.
  102. Mack KD, et al. HIV insertions within and proximal to host cell genes are a common finding in tissues containing high levels of HIV DNA and macrophage-associated p24 antigen expression. *J Acquir Immune Defic Syndr*. 2003;33(3):308–320.
  103. Han Y, et al. Resting CD4<sup>+</sup> T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J Virol*. 2004;78(12):6122–6133.
  104. Ikeda T, Shibata J, Yoshimura K, Koito A, Matsushita S. Recurrent HIV-1 integration at the BACH2 locus in resting CD4<sup>+</sup> T cell populations during effective highly active antiretroviral therapy. *J Infect Dis*. 2007;195(5):716–725.
  105. Maldarelli F. HIV-infected cells are frequently clonally expanded after prolonged antiretroviral therapy. *J Virus Erad*. 2015;1:237–244.
  106. Uittenboogaard LM, et al. BACH2: a marker of DNA damage and ageing. *DNA Repair (Amst)*. 2013;12(11):982–992.
  107. Kuwahara M, et al. The Menin-Bach2 axis is critical for regulating CD4 T-cell senescence and cytokine homeostasis. *Nat Commun*. 2014;5:3555.
  108. Roychoudhuri R, et al. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature*. 2013;498(7455):506–510.
  109. Turkmen S, Riehn M, Klopocki E, Molkenin M, Reinhardt R, Burmeister T. A BACH2-BCL2L1 fusion gene resulting from a t(6;20)(q15;q11.2) chromosomal translocation in the lymphoma cell line BLUE-1. *Genes Chromosomes Cancer*. 2011;50(6):389–396.
  110. Kobayashi S, et al. Identification of IGHCdelta-BACH2 fusion transcripts resulting from cryptic chromosomal rearrangements of 14q32 with 6q15 in aggressive B-cell lymphoma/leukemia. *Genes Chromosomes Cancer*. 2011;50(4):207–216.
  111. Liu J, Sorensen AB, Wang B, Wabl M, Nielsen AL, Pedersen FS. Identification of novel Bach2 transcripts and protein isoforms through tagging analysis of retroviral integrations in B-cell lymphomas. *BMC Mol Biol*. 2009;10:2.
  112. Cao S, et al. High-throughput RNA sequencing-based virome analysis of 50 lymphoma cell lines from the Cancer Cell Line Encyclopedia project. *J Virol*. 2015;89(1):713–729.
  113. Wang X, et al. Integration of Epstein-Barr virus into chromosome 6q15 of Burkitt lymphoma cell line (Raji) induces loss of BACH2 expression. *J Virol*. 2004;164(3):967–974.
  114. Sasaki S, et al. Cloning and expression of human B cell-specific transcription factor BACH2 mapped to chromosome 6q15. *Oncogene*. 2000;19(33):3739–3749.
  115. Lee SM, Vasishttha M, Prywes R. Activation and repression of cellular immediate early genes by serum response factor cofactors. *J Biol Chem*. 2010;285(29):22036–22049.
  116. Cen B, Selvaraj A, Prywes R. Myocardin/MKL family of SRF coactivators: key regulators of immediate early and muscle specific gene expression. *J Cell Biochem*. 2004;93(1):74–82.
  117. Cen B, et al. Megakaryoblastic leukemia 1, a potent transcriptional coactivator for serum response factor (SRF), is required for serum induction of SRF target genes. *Mol Cell Biol*. 2003;23(18):6597–6608.
  118. Huang D, et al. C11orf95-MKL2 is the resulting fusion oncogene of t(11;16)(q13;p13) in chondroid lipoma. *Genes Chromosomes Cancer*. 2010;49(9):810–818.
  119. Flucke U, et al. Presence of C11orf95-MKL2 fusion is a consistent finding in chondroid lipomas: a study of eight cases. *Histopathology*. 2013;62(6):925–930.
  120. Ho Y-C. Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure [dissertation]. Baltimore MD: Johns Hopkins University. 2013. <https://jscholarship.library.jhu.edu/handle/1774.2/37077>.
  121. Brady T, Agosto LM, Malani N, Berry CC, O’Doherty U, Bushman F. HIV integration site distributions in resting and activated CD4<sup>+</sup> T cells infected in culture. *AIDS*. 2009;23(12):1461–1471.
  122. Brady T, et al. Quantitation of HIV DNA integration: effects of differential integration site distributions on Alu-PCR assays. *J Virol Methods*. 2013;189(1):53–57.

123. Bushman FD. Integration Site Pipeline and Database. UPenn Web site. <http://www.bushmanlab.org>. Accessed December 14, 2015.
124. Shao W, et al. NCI Retrovirus Integration Database (RID). DRP-RID Web site. <https://fr-s-bsg-hgi-0.ncifcrf.gov>. Accessed December 14, 2015.
125. Simonetti FR, et al. Clonally expanded CD4<sup>+</sup> T-cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci U S A*. In press.
126. Hughes JF, Coffin JM. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci U S A*. 2004;101(6):1668–1672.
127. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*. 2011;8:90.
128. Badley AD, Sainski A, Wightman F, Lewin SR. Altering cell death pathways as an approach to cure HIV infection. *Cell Death Dis*. 2013;4:e718.
129. Evans VA, Khoury G, Saleh S, Cameron PU, Lewin SR. HIV persistence: chemokines and their signalling pathways. *Cytokine Growth Factor Rev*. 2012;23(4–5):151–157.
130. Kent SJ, et al. The search for an HIV cure: tackling latent infection. *Lancet Infect Dis*. 2013;13(7):614–621.
131. Rajasuriar R, Wright E, Lewin SR. Impact of antiretroviral therapy (ART) timing on chronic immune activation/inflammation and end-organ damage. *Curr Opin HIV AIDS*. 2015;10(1):35–42.
132. Archin NM, Margolis DM. Emerging strategies to deplete the HIV reservoir. *Curr Opin Infect Dis*. 2014;27(1):29–35.
133. Margolis DM, Hazuda DJ. Combined approaches for HIV cure. *Curr Opin HIV AIDS*. 2013;8(3):230–235.