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+ 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+ 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+ 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 identified clonally expanded, HIV-infected CD4+ 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 mononuclear cells (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
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+ 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 VOAs 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 term infection 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 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+ cells (52, 58), various CD4+ 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,
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 provirus (89). Other retroviruses exhibit distinct preferences (90), suggesting such preferences are IN specific.

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 provirus (89). Other retroviruses exhibit distinct preferences (90), suggesting such preferences are IN specific.
Integration site: a determining factor in viral pathogenesis

Integration site has important consequences for clonal expansion in retroviral pathogenesis. For example, human T-lymphotropic 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 of infected cells undergo clonal expansion (97). Analysis of the leukemic cells revealed a selected population of genes with MLV integrations, suggesting leukemias were favored when these integrations occurred.

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 cytotoxic. 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 1. Potential sources of HIV persistence

<table>
<thead>
<tr>
<th>Decay kinetics/half-life</th>
<th>HIV replication competence</th>
<th>After long-term cART</th>
<th>Contribution to HIV reservoirs</th>
<th>Contribution to HIV-infected cell population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activated</td>
<td>Phase 1/~1 d</td>
<td>+</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Activated</td>
<td>Phase 1/~1 d</td>
<td>-</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Phase 2/~14 d</td>
<td>+</td>
<td>Unlikely</td>
<td>Unlikely</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Phase 2/~14 d</td>
<td>-</td>
<td>Unlikely</td>
<td>Unlikely</td>
</tr>
<tr>
<td>Long-lived</td>
<td>Phase 3.4/&gt;39 wk</td>
<td>+</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Long-lived</td>
<td>Phase 3.4/&gt;39 wk</td>
<td>-</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

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.
Table 2. Mechanisms of persistence

<table>
<thead>
<tr>
<th>Immune mediated</th>
<th>Integration site mediated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeostatic proliferation</td>
<td>Host genes involved in cell growth regulation (orientation independent)</td>
</tr>
<tr>
<td>Antigenic stimulation</td>
<td>Specific, orientation-dependent sites</td>
</tr>
<tr>
<td>Generalized immune activation</td>
<td></td>
</tr>
</tbody>
</table>

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 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 oligogoclony 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 sites 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 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 Bohles St., Frederick National Cancer Laboratory, Frederick, Maryland 21702, USA. Phone: 301.846.5611; E-mail: fmaili@mail.nih.gov.

---


24. Igarashi T, et al. Macrophage are the principal reservoir and sustain high virus loads in rhesus macaques after the depletion of CD4+ T cells by a highly pathogenic simian immunodeficiency virus (SIV)/SHIV strain that results in persistent low-level viremia in patients with HIV-1 infection during receipt of combination antiretroviral therapy. *J. Leuk. Biol.* 2003;74(5):635–641.

REVIEW SERIES: HIV


71. Hu WS, Hughes SH. HIV-1 reverse transcrip-

72. Craige R, Bushman FD. HIV DNA inte-

vania, USA: Lippincott Williams and Wilkins; 2013:1424–1473.

74. Engelman A. Isolation and analysis of HIV-1 inte-

75. Li M, Craige R. Nucleoprotein complex

76. Krishnan L, et al. The requirement for cellular transpor-
tin 3 (TNPO3 or TRN-SR2) during infect-
dion maps to human immunodeficiency


78. Matreyek KA, Engelman A. The requirement for nucleo-
porin NUP153 during human immunode-

79. De Iaco A, Luban J. Inhibition of HIV-1 infection and nuclear architecture on HIV integration site


82. Engelman A, Cherepanov P. The lentiviral inte-


85. Serra O, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral inta-


87. Serra O, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral inta-

88. Demeulemeester J, De Rijck J, Gijbers R, Debayer Z. Retroviral integration: site matters: mecha-


90. Gillet NA, et al. The host genomic environ-
ment of the provirus determines the abun-

91. Niederer HA, Bangham CR. Integration site and clonal expansion in human chronic ret-


93. Du Y, Jenkins NA, Copeland NL. Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow pro-

94. Justice JT, et al. The MET gene is a common integration target in avian leukosis virus sub-

95. Dave UP, et al. Murine leukemias with retro-
viral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients fol-


97. Kustikova OS, et al. Retroviral vector insertion sites associated with dominant hematopo-


100. 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 macro-


102. Ikeda T, Shibata J, Yoshimura K, Koito A, Matsumoto J. Retroviral integration: site matters: mecha-

103. Kobayashi Y, et al. Identification of IGHc-Dealta-

104. Roychoudhuri R, et al. BACH2 represses effector pro-


110. Huang D, et al. Cl10orf95–MKL2 is the result-

111. Brady T, et al. Quantitation of HIV DNA inte-


114. Serra O, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral inta-


The Journal of Clinical Investigation

2014;135:5355.

2015;23(12):1461–1471.