**REVIEW**

# The benefits of integration

**A. Ciuffi**

*Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland*

## Abstract

Retroviruses, including the human immunodeficiency virus (HIV), are notorious for two essential steps of their viral replication: reverse transcription and integration. This latter property is considered to be essential for productive replication and ensures the stable long-term insertion of the viral genome sequence in the host chromatin, thereby leading to the life-long association of the virus with the infected cell. Using HIV as a prototypic example, the present review aims to provide an overview of how and where integration occurs, as well as presenting general consequences for both the virus and the infected host.

Clinical Microbiology and Infection © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

**Keywords:** Endogenous retrovirus, human immunodeficiency virus, integrase, integration, lens-epithelium derived growth factor/p75, nuclear import, nuclear pore proteins, persistence, retrovirus

**Corresponding author:** A. Ciuffi, Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Bugnon 48, 1011 Lausanne, Switzerland
**E-mail:** Angela.Ciuffi@chuv.ch

## Introduction

Retroviruses are enveloped RNA viruses, containing two copies of single-stranded, non-segmented, positive RNA as genome. Like all retroviruses, the human immunodeficiency virus type 1 (HIV-1, abbreviated as HIV throughout the text) encodes three major open reading frames: (i) *gag*, coding for the internal structural proteins, which in the case of HIV are matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7) and p6; (ii) *pol*, coding for the virus enzymes, which are reverse transcriptase (RT, p66/p51), integrase (IN, p32) and protease (PR, p11); and (ii) *env*, coding for the envelope external structural proteins, which are the surface glycoprotein (SU, gp120) and the transmembrane glycoprotein (TM, gp41) for HIV (Fig. 1a). Moreover, HIV-1 encodes regulatory and accessory genes—*tat, rev, vif, vpr, vpu* and *nef*—and is therefore considered to be a complex retrovirus.

In order to establish a productive replication, HIV needs first to deliver its genome-containing viral core to the cytoplasm of the infected cell (Fig. 1b). Subsequently, the viral RNA genome

is reverse transcribed in a double-stranded DNA linear copy, hence the name, retrovirus. The viral DNA genome is complexed with the viral integrase enzyme (also referred to as the intasome), as well as with additional viral and cellular proteins in a ribonucleoprotein complex called the pre-integration complex (PIC). The exact PIC composition is still controversial but additional proteins may include the viral proteins RT, MA, Vpr and CA, as well as the cellular proteins lens-epithelium derived growth factor (LEDGF/p75), barrier-to-autointegration factor (BAF) and high-mobility group AT-hook 1 (HMGA1) (reviewed in ref. [1]). Components of PIC are necessary for the successful nuclear import and final stable insertion of the viral genome in the host DNA. As discussed below, these two steps have been shown to impact integration efficiency and site location, and are essential to ensure the life-long persistence of the provirus in the infected cell.

## How? Mechanism of Integration

The viral integrase enzyme is the only protein determinant required to successfully join and so insert a DNA fragment into a heterologous target DNA sequence (reviewed in refs [2,3]).

The prototypical HIV-1 IN is a 288-amino-acid protein, and is divided into three major domains: an N-terminal domain, a catalytic core domain and a C-terminal domain (Fig. 2a)
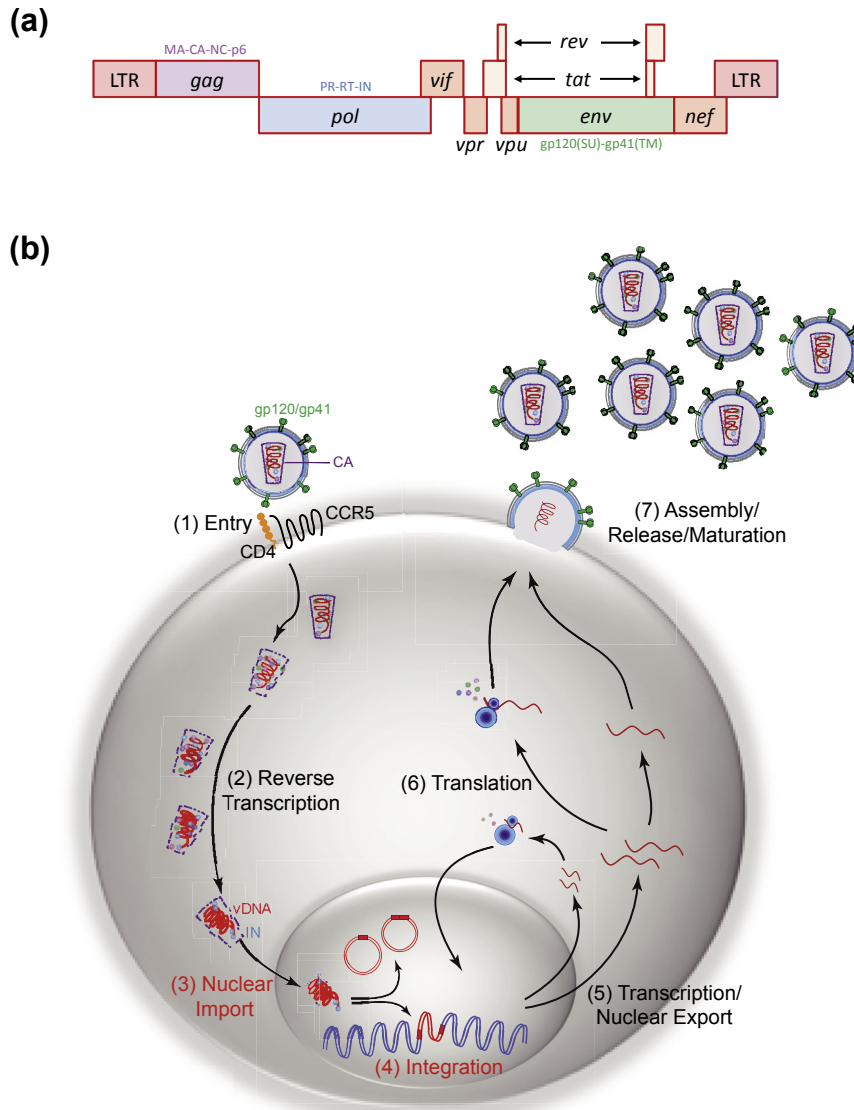
FIG. 1. Overview of the human immunodeficiency virus type 1 (HIV-1) genome organization and replication cycle. (a) HIV genome contains nine open reading frames (ORF), coding for 15 proteins. The proviral genome is flanked by direct long-terminal repeats (LTR) that contain all transcriptional regulatory sequences. The *pol* ORF encodes the three viral enzymes necessary for replication that are protease (PR), reverse transcriptase (RT) and integrase (IN). (b) The HIV replication cycle can be divided into seven steps: (1) HIV binds to its target cell through the interaction of gp120 to the cell CD4 molecule, which is mostly expressed at the surface of lymphoid and myeloid cells. This first interaction allows the subsequent binding of gp120 with a chemokine receptor, CCR5 or CXCR4, followed by the fusion between the viral membrane and the cellular membrane triggered by gp41. This ensures the release of the viral core in the cytoplasm of the host cell. (2) The viral core disassembles (uncoating process) and the viral RNA genome is reverse transcribed in a linear double-stranded DNA copy through the action of the viral reverse transcriptase enzyme, giving rise to the pre-integration complex (PIC). (3) The PIC, minimally containing the viral DNA (vDNA) genome and the viral IN enzyme, is translocated to the nucleus through the nuclear pore. This nuclear import step requires multiple interactions between viral and cellular proteins, including capsid (CA) binding to nuclear pore proteins (NUPs). (4) Once in the nucleus, the viral IN catalyses the stable insertion of the viral DNA genome into the host chromatin, tethered mainly by the cellular lens-epithelium derived growth factor (LEDGF)/p75 protein. LTR circles, 1-LTR and 2-LTR circles, are considered as dead-end by-products produced by the cellular DNA repair machinery, the homology repair (HR) or non-homologous end-joining (NHEJ) pathway, respectively. (5) Once integrated, the provirus is transcribed by the cellular RNA polymerase II machinery as most cellular coding genes. Viral transcripts (with different levels of splicing) are exported from the nucleus to the cytoplasm where they are translated (6). (7) Two copies of full-length (unspliced) viral RNA and viral proteins assemble, thereby producing new particles that are released from the cellular membrane. Finally, the viral protease cleaves viral polyproteins leading to mature and infectious viral particles.
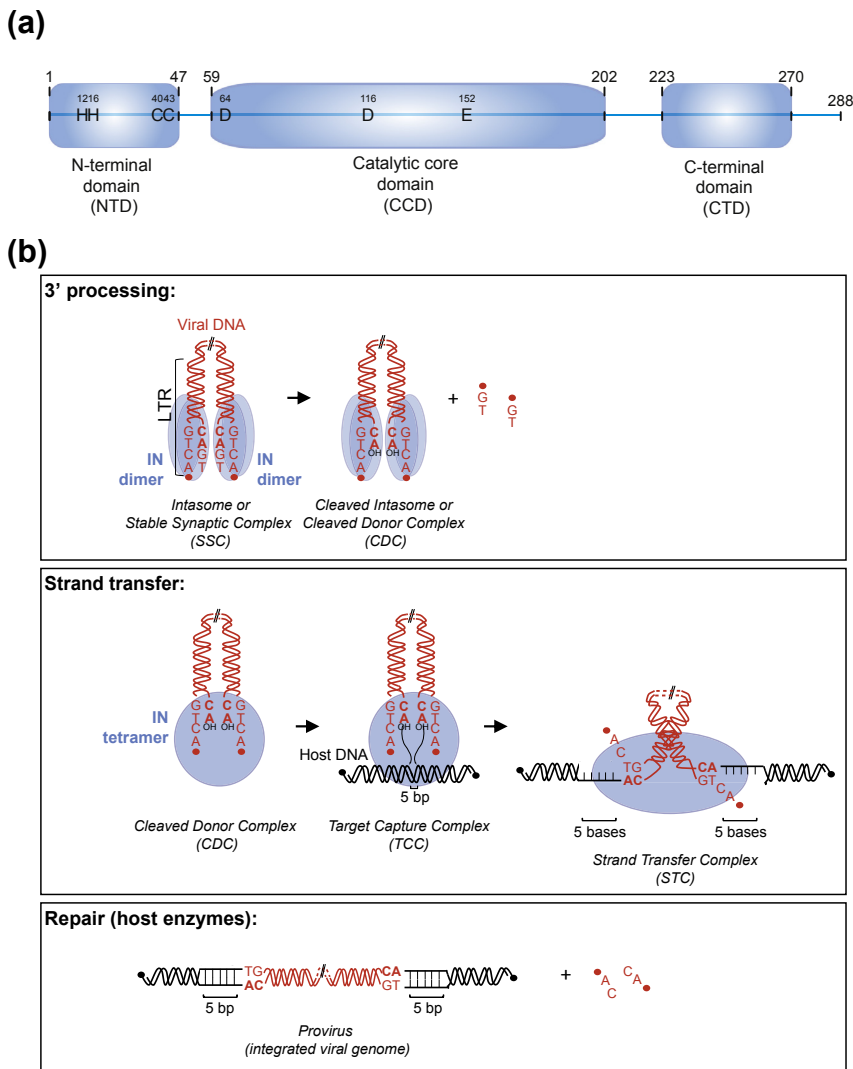
**(a)**



**(b)**



**FIG. 2.** Mechanism of viral integration. (a) Domain organization of the human immunodeficiency virus type 1 (HIV-1) integrase enzyme, common to all retroviral integrases. Indicated residues constitute motifs that are conserved throughout integrases. (b) Details of the integration reaction. The integration reaction can be divided in three steps: (i) the 3′ processing or terminal cleavage reaction, (ii) the strand transfer reaction or DNA joining reaction, and (iii) the DNA repair. The first two steps are carried out by the viral integrase, whereas the last one is thought to be performed by host cellular enzymes. The intasome or stable synaptic complex (SSC) contains viral DNA ends (red; red circles mark the 5′ end of viral DNA), each bound to an integrase (IN) dimer (blue ovoids) that multimerize, thereby resulting in a tetramer of IN. The viral DNA attachment (att) sites contain minimally 16 bp, including the invariant CANN 3′-terminus, essential for efficient IN binding, processing and subsequent integration in target DNA. First, the IN dimer removes a dinucleotide (GT) on each 3′ viral DNA end, leaving an hydroxyl group on the A nucleotide of the conserved CA sequence (bold), and resulting in the cleaved intasome or cleaved donor complex (CDC). The second step occurs in the nucleus as IN from the CDC has to bind to the host target DNA (black), thereby forming the target capture complex (TCC). A one-step transesterification reaction is carried out by the viral IN, catalysing the simultaneous breaking of the target DNA 5 bp apart and the covalent joining of the 3′-OH recessed viral DNA end, thereby leading to the strand transfer complex (STC). Of note, the 5-bp stagger in the target DNA varies between 4 and 6 bp depending on the retroviral IN (4 bp for gammaretroviruses and spumaviruses, 5 bp for lentiviruses and 6 bp for alpha-, beta- and deltaretroviruses). Finally, enzymes of the host DNA repair machinery remove the 5′ protruding viral DNA ends (5′-AC) and repair the unpaired 5-base gap flanking both ends of the inserted viral DNA sequence, resulting in the duplication of the 5-bp stagger on each side of the viral sequence, and leading to the stable integration of the viral DNA genome (provirus) in the host target DNA.

(reviewed in refs [2,3]). The N-terminal domain contains a zinc-binding HHCC motif, and is involved in IN multimerization and viral DNA binding. The catalytic core domain contains the D,D-35-E triad motif, which constitutes the catalytic active site essential to coordinate a pair of $Mg^{2+}$ ions and carry out the IN enzymatic function. This motif is typical of polynucleotidyl transferases, including retrotransposon integrases, bacterial transposases and RAG1/2 recombinases [2]. The catalytic core domain also contains key residues involved in target and viral DNA binding. Finally, the C-terminal domain is mostly involved in IN multimerization and DNA binding. Of note, some retroviral integrases (including primate foamy virus, the spumaretrovirus prototype that provided most of the structural information to date) display an additional domain of ~50 amino acids called the N-terminal extension domain [4].

Successful viral integration occurs in three major steps, which are the 3′ processing reaction, the strand transfer reaction and the DNA repair (Fig. 2b).

## Where? Integration Site Selection

*In vitro*, IN is sufficient to promote insertion of a donor DNA into a target DNA, at any phosphodiester bond. However, *in vivo*, retroviral integration is not random but rather favours chromosomal features in a retrovirus-specific manner (reviewed in refs [1,5−8]). This suggests that additional proteins may play a role in integration, impacting both integration efficiency and integration site selection [6,9].

Upon the publication of the human genome, pioneering studies using HIV and murine leukaemia virus (MLV, a gammaretrovirus) revealed specific retroviral genomic locations that were preferentially selected for viral integration [10,11]. Since then, multiple studies further investigated these insertion preferences and this for most retroviral genera (reviewed in refs [1,3,6−8,12]). It is now well established that HIV, and other lentiviruses, favour active transcription units (i.e. genes for which high amounts of corresponding transcripts were detected) that are located in gene-dense regions, thereby correlating with high GC content, high Alu elements, low LINE elements, light Giemsa bands (so less condensed chromatin), and epigenetic marks of active transcription (H3K4me, H3K4me2, H3K9me, H3K27me, H3K36me, DNaseI hypersensitive sites and acetylated histones) [6−8]. Lentiviruses favour integration throughout the transcription units rather than promoter regions or transcription start sites (Fig. 3, red arrows in the inset). In contrast, MLV favours 5′ ends of genes, i.e. close to transcription start sites, as well as strong promoters and active enhancer regions, typically marked by H3K4me1 and H3K27ac (Fig. 3, black arrows in the inset) [13,14]. Other retroviral genera display alternative

integration site distributions, with milder preferences for transcription units or transcription start sites [7,8]. Of note, no specific integration site preference for betaretroviruses (with the mouse mammary tumour virus prototype) has been revealed, so far making them unique among retroviruses because they show a random integration site pattern.

In the past decade, these integration preferences have been shown to be due to both nuclear import and tethering mechanisms (recently reviewed in [12]). To integrate in the host chromatin, the pre-integration complex has first to enter the nucleus [15]. MLV PIC succeeds in reaching the host genome only during cell mitosis, that is, when the nuclear membrane is disrupted, and requires the viral p12 protein [12]. In contrast, HIV can also infect non-dividing cells because HIV PIC can interact with the nuclear import machinery to be actively translocated through the nuclear membrane into the nucleus. This first difference in accessing the host genome also suggests that the chromatin landscape encountered by the retroviral PIC is different at the moment of tethering.

The first hint of a link between nuclear import and integration site selection was the observation that knock-down of TNPO3/transportin-3, RanBP2/NUP358 and NUP153, all involved in nuclear import, could bias integration preferences to regions with lower gene density (Fig. 3) [16−20]. Since then, additional nuclear pore proteins were shown to impact integration, including CPSF6, Tpr and NUP98 [17,19,21−25]. Finally, Marini et al. investigated the nuclear architecture and showed that the open chromatin landscape in the vicinity of the nuclear pore complex guided at first HIV integration site preferences [26]. Indeed, chromatin is differentially distributed within the nucleus, with condensed heterochromatin at the nuclear envelope and more open chromatin within the nucleus [27]. Genes that are actively transcribed tend to be located close to the nuclear pore complex basket tip, in euchromatin, whereas inactive genes are associated with lamin at the nuclear envelope, thereby coupling nuclear architecture and gene regulation [18,27−30]. Using three-dimensional immune-DNA fluorescence *in situ* hybridization and recurrent integrated genes, Marini et al. demonstrated that HIV integrated preferentially in the nuclear periphery, within 1 μm of the nuclear envelope (considering that on average the cell diameter of CD4+ T cells is 7 μm). There transcription was active, as shown by association with typical epigenetic marks (H3K9ac, H3K36me3, H4K16ac, H4K20me). In contrast, proviral insertions were disfavoured in lamin-associated domains, containing inactive genes and located at the nuclear membrane, as well as in the centre of the nucleus. These data further suggest that chromatin located in the vicinity of the nuclear pore complex is a major determinant of HIV integration site selection [26].

Once the viral intasome is within the nucleus and can access the host chromatin, additional players further fine tune final
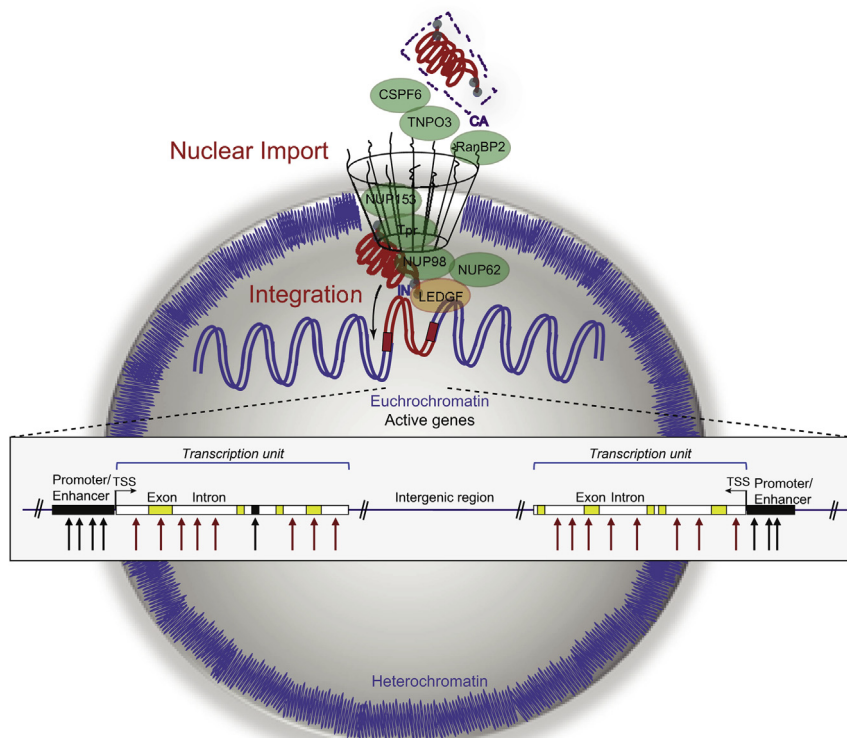
**FIG. 3.** Mechanisms of integration site selection. Integration site selection varies according to the retrovirus and depends on nuclear entry, chromatin organization, host cell tethering proteins and viral integrase (IN). Gag-encoded proteins are viral determinants required for nuclear entry, whereas the viral IN is required for the final stable insertion of the viral genome into host chromatin. Human immunodeficiency virus (HIV) pre-integration complex (PIC) enters the nucleus through multiple interactions with nuclear pore complex components (green ovals) involving mostly HIV capsid (CA) protein. Upon entry to the nucleus, HIV PIC is close to gene-dense regions that are located at the basket tip of the nuclear pore complex, where are located genes that are expressed (euchromatin). Lens-epithelium derived growth factor (LEDGF/p75) protein (orange oval) then binds to HIV IN and to H3K36me3 histones that mark elongating transcription units, guiding viral genome insertion in active transcription units (red arrows in the inset), with no bias for exons or introns if normalized upon their respective lengths. Murine leukemia virus (MLV) PIC enters the nucleus during mitosis upon nuclear envelope breakdown and binds first to chromatin via p12 (not shown). Subsequently, BET proteins bind to MLV IN, recruiting the intasome to acetylated histone tails on the chromatin (not shown), thereby guiding MLV genome insertion in promoter and enhancer regions (black arrows in the inset).

chromosomal preferences (Fig. 3). To date, LEDGF/p75 protein (a splice variant product of *PSIP1* gene) has been identified as the major cellular tethering protein for HIV, recruiting the viral integrase and further guiding HIV PIC along active transcription units, to regions associated with epigenetic marks of gene expression, including H3K36me3 [9,31–37]. Recently, cellular proteins tethering MLV PIC to promoter and enhancer regions associated with acetylated histone were identified as Bromodomain and Extraterminal domain (BET) proteins, including Brd2, Brd3 and Brd4 [9,37–41]. Mutating the chromatin-binding domain of these tethering proteins resulted in modifying integration site preferences, further highlighting their role

in driving fine integration site selection [33,42–45]. Finally, the viral integrase also affects the final integration step, favouring DNA sequences that are preferentially wrapped around nucleosomes, facing outwards, and that are severely bent [12,46–49].

In summary, PIC nuclear import and integration are tightly coupled. HIV CA protein is required for nuclear import and dictates integration in gene-dense regions that are located in the nuclear pore periphery [12,50,51]. Once there, finer chromosomal features for integration site selection are dictated by cellular tethering proteins and the viral IN [9,12]. To date, LEDGF/p75 has been shown to be the major protein guiding HIV PIC to active

transcription units [7,8], whereas BET proteins are involved in MLV PIC binding to promoter and enhancer regions [9].

## So What? Consequences for the Virus and the Host

The stable insertion of the viral genome in the host chromatin implies a life-long association of the virus with the infected cell, which may have several implications for both the virus and the host (Fig. 4).

To propagate, retroviruses need to replicate efficiently, i.e. to infect/enter, to express and to release viral particle progeny. The site of viral genome integration may impact viral propagation as it may impact the level of viral gene expression (Fig. 4) [52]. Integration in heterochromatin may repress viral expression, and so inhibit particle production and spreading infection. In contrast, integration in a very active chromatin environment might result in viral gene over-expression that can be toxic for the infected cell, thereby leading to its premature death and a reduced release of infectious particles. The orientation of viral integration with respect to the hosting gene can also affect viral expression because RNA interference may occur between viral transcription and cellular gene transcription. In conclusion, there is a balance between integration site location and viral gene expression, which in turn affects viral particle production, and hence propagation.

Viral genome insertion in the host DNA, known as insertional mutagenesis, also impacts the infected cell as it disrupts the host genome integrity, with a more or less dramatic outcome (Fig. 4) [53,54]. This latter depends on three parameters: (i) the nature of the gene hosting or close to the integration event (i.e. genes involved in cell proliferation, essential genes), (ii) the type of gene structure modification (i.e. dominant negative splice variant, gene truncation) and (iii) the impact on gene expression (i.e. over-expression or knock-down).

FIG. 4. Overview of consequences of viral genome insertion in the host chromatin. (a) Gene structure without viral genome insertion. (b) Integration in intergenic regions should have few consequences on host cell gene expression, at least at the structural level. However, integration in regulatory elements (enhancers) could disrupt their activity and so affect the level of cellular gene expression. If these regions are in heterochromatin, then proviral genome expression will be reduced. (c) Provirus is located within enhancer or promoter region. This should result in efficient viral gene expression. The 3′ long terminal repeats may act as a strong promoter driving higher expression level of the downstream flanking cellular gene (increased black arrow). (d, e) Provirus is located within the transcription unit, in the antisense orientation (d) or in the same orientation (e), and disrupts cellular gene expression. Viral gene expression might be impaired: (i) by collision of opposite transcriptional machineries, (ii) by a stronger cellular promoter or (iii) by viral gene out-splicing. Aberrant cellular gene products can arise due to truncations mediated by the viral genome insertion or by chimeras with viral genes. TSS: Transcription start sites.
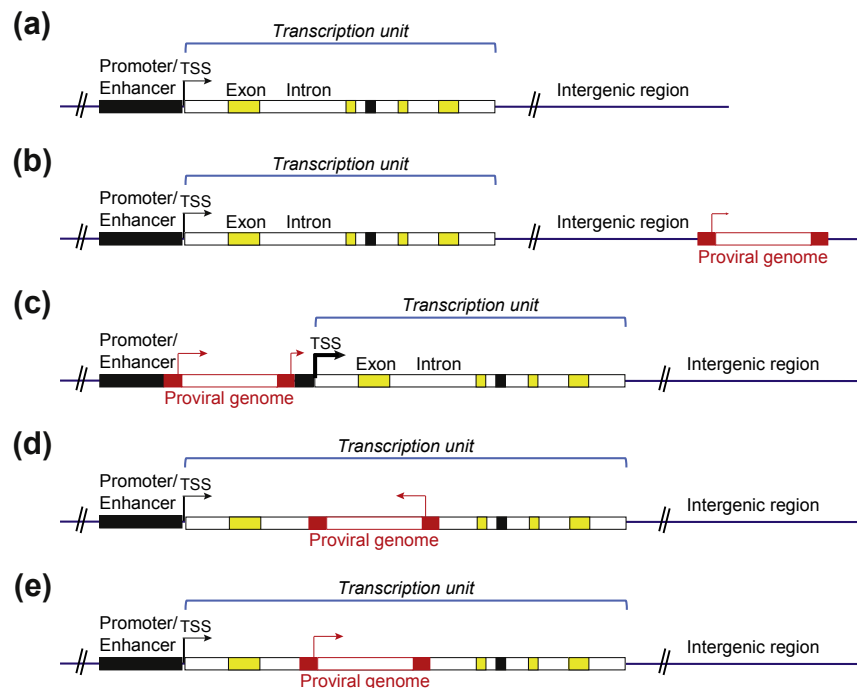
The type of gene affected by viral insertion will dictate the outcome. Indeed it is easy to imagine that a cellular gene involved in driving apoptosis, or in contrast in promoting cell proliferation, eventually results in cell death or cancer. The direct impact of integration site location on apoptosis has not been thoroughly investigated because of the difficulty in discriminating such cell death from viral gene-mediated toxicity. In contrast multiple examples of leukaemia due to insertional mutagenesis in the *LIM-domain only 2* (*LMO2*) proto-oncogene (and not due to virus-encoded oncogenes) have been identified in gene therapy clinical trials using MLV-based vectors (reviewed in [53,54]). Other genes over-expressed upon insertional mutagenesis and resulting in clonal dominance or in leukaemia include *cyclin D2* (*CCND2*), *MDS1 and EVI1 complex locus* (*MECOM*), *PR domain containing 16* (*PRDM16*), *SET binding protein 1* (*SETBP1*), *BMI1 proto-oncogene, polycomb ring finger* (*BMI1*) and *High mobility group AT-hook 2* (*HMGA2*) [53]. In some cases, insertional mutagenesis does not lead to cellular transformation but provides a growth advantage (by promoting cell proliferation or by promoting cell survival) finally resulting in clonal expansion and dominance. This has been recently exemplified in antiretroviral therapy-suppressed HIV-positive individuals, where most cells displayed the exact same proviral insertion site within the same gene, such as *BTB and CNC homology 1, basic leucine zipper transcription factor 2* (*BACH2*) [55−57]. The mechanism by which such integration sites promote clonal dominance, i.e. through enhanced proliferation or through enhanced survival, remains to be clarified. These studies would be helpful because they may impact the use of safe viral vectors in gene therapy.

Finally, the stable integration of viral DNA in the host genome may have further consequences for the infected host, including long-term persistent infection and species evolution. Indeed, in the case of HIV for example, infected cells can harbour a silent provirus, i.e. in a state of transcriptional latency, and so are not eliminated by the immune system [58]. These latently infected cells have a long half-life (memory CD4$^+$ T cells with $t_{1/2}$ ~44 months), thereby allowing the long-term persistence of HIV within the infected organism. By mechanisms that are not yet fully understood, latently infected cells can be reactivated *in vivo* and then produce again viral particles, re-kindling the infection. This long-term HIV persistence is considered to represent the major obstacle for the complete elimination of HIV, and thus for curing HIV. Integration therefore ensures a life-long stable association of the virus with the infected host, representing a challenge for treatment strategies.

Modern retroviruses currently challenging humans (HIV, human T-lymphotropic virus) have a CD4 tropism, i.e. they infect CD4$^+$ cells, implying horizontal transmission between hosts. In contrast, retroviruses able to infect germ-line cells, and so stably inserting retroviral DNA sequences in the genome of germ-line cells, would transmit the virus vertically, to descendants, implying genetic inheritance. This process is called endogenization of the retrovirus and may contribute to shaping the human species. Although there is no such retrovirus infecting humans currently, the human genome contains multiple human endogenous retrovirus (HERV) sequences, providing the proof-of-concept that such infections occurred in the past (recently reviewed in [59,60]). These remnants of ancient retroviral infections currently represent ~8% of the human DNA genome [59]. More than 500 HERV families have been identified, most of which have accumulated mutations and are therefore defective to propagate infection [60,61]. However, some retroviral sequences, displaying different degrees of evolution, are still active, thereby contributing to host life and mammalian evolution. One prominent example of endogenization is syncytin [62]. *Syncytin* genes are derived from HERV-encoded *envelope* genes (HERV-W, HERV-FRD), which have fusogenic properties, and they are expressed in trophoblasts and are essential for placenta morphogenesis in mammals. Other examples also include the activity of HERV long terminal repeat enhancer and promoter sequences in modulating or driving expression of human genes, such as *bile acid-CoA:amino acid N-acyltransferase* (*BAAT*), a liver-specific enzyme present in bile and involved in lipid absorption [60]. Hence, retroviral DNA sequences present in the host genome that survived natural selection may offer benefits to the infected host.

During the past decade, the impact of ERV expression in human diseases has also been under scrutiny (reviewed in [59,60,63]). These studies identified a role for ERV expression in the development of some cancers, in neurological disorders as well as in autoimmune disorders. Indeed expression of HERV-encoded *env* (HERV-K, HERV-W, HERV-E) was shown to be immunosuppressive and to promote cancer development, including leukaemia, prostate cancer and breast cancer. Similar correlation was identified between high expression levels of HERV-W-encoded *env* and some neurological diseases, including multiple sclerosis and schizophrenia [64]. HERV expression can activate the innate immune response, thereby leading to chronic inflammation and potentially contributing to autoimmune diseases. Hence, deregulation of HERV expression can be associated with human pathologies, using mechanisms that will need to be further addressed and investigated.

## Conclusion

Retroviruses are fascinating organisms because they can be detrimental to the infected host as well as providing benefits, shaping human evolution. They also provide fantastic tools and opportunities to further understand human biology as well as improve human health with gene therapy.

A better understanding of all the players involved in guiding integration site selection and identifying the required determinants for safe and targeted integration events are still considered as the Holy Grail for gene therapy purposes. Similarly, the advent of novel technologies in the last decade such as high-throughput sequencing just started to unravel the impact of HERV in human biology.

## Acknowledgements

## Transparency Declaration

The author has declared that there are no conflicts of interest in connection with this article.

## References

[1] Craigie R, Bushman FD. Host factors in retroviral integration and the selection of integration target sites. Microbiol Spectrum 2014;2(6).

[2] Engelman A, Cherepanov P. Retroviral integrase structure and DNA recombination mechanism. Microbiol Spectrum 2014;2:1–22.

[3] Grandgenett DP, Pandey KK, Bera S, Aihara H. Multifunctional facets of retrovirus integrase. World J Biol Chem 2015;6:83–94.

[4] Krishnan L, Engelman A. Retroviral integrase proteins and HIV-1 DNA integration. J Biol Chem 2012;287:40858–66.

[5] Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, et al. Genome-wide analysis of retroviral DNA integration. Nature Rev 2005;3:848–58.

[6] Ciuffi A. Mechanisms governing lentivirus integration site selection. Curr Gene Ther 2008;8:419–29.

[7] Desfarges S, Ciuffi A. Retroviral integration site selection. Viruses 2010;2:111–30.

[8] Ciuffi A, Bushman F. Retroviral integration target site selection. In: Neamati N, editor. HIV-1 integrase: mechanism of action and inhibitor design wiley series in drug discovery and development. 1 ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011. p. 51–65.

[9] Debyser Z, Christ F, De Rijck J, Gijsbers R. Host factors for retroviral integration site selection. Trends Biochem Sci 2015;40:108–16.

[10] Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 2002;110:521–9.

[11] Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. Science (New York, NY) 2003;300(5626):1749–51.

[12] Demeulemeester J, De Rijck J, Gijsbers R, Debyser Z. Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. Bioessays 2015;37:1202–14.

[13] De Ravin SS, Su L, Theobald N, Choi U, Macpherson JL, Poidinger M, et al. Enhancers are major targets for murine leukemia virus vector integration. J Virol 2014;88:4504–13.

[14] LaFave MC, Varshney GK, Gildea DE, Wolfsberg TG, Baxevanis AD, Burgess SM. MLV integration site selection is driven by strong enhancers and active promoters. Nucleic Acids Res 2014;42:4257–69.

[15] Hennig T, O'Hare P. Viruses and the nuclear envelope. Curr Opin Cell Biol 2015;34:113–21.

[16] Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, et al. HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. PLoS Pathogens 2011;7:e1001313.

[17] Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, et al. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS Pathogens 2011;7:e1002439.

[18] Ptak C, Aitchison JD, Wozniak RW. The multifunctional nuclear pore complex: a platform for controlling gene expression. Curr Opin Cell Biol 2014;28:46–53.

[19] Koh Y, Wu X, Ferris AL, Matreyek KA, Smith SJ, Lee K, et al. Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. J Virol 2013;87:648–58.

[20] Di Nunzio F, Fricke T, Miccio A, Valle-Casuso JC, Perez P, Souque P, et al. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. Virology 2013;440:8–18.

[21] Lelek M, Casartelli N, Pellin D, Rizzi E, Souque P, Severgnini M, et al. Chromatin organization at the nuclear pore favours HIV replication. Nature Communications 2015;6:6483.

[22] Matreyek KA, Engelman A. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. Viruses 2013;5:2483–511.

[23] Matreyek KA, Yucel SS, Li X, Engelman A. Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. PLoS Pathogens 2013;9(10):e1003693.

[24] Chin CR, Perreira JM, Savidis G, Portmann JM, Aker AM, Feeley EM, et al. Direct visualization of HIV-1 replication intermediates shows that capsid and CPSF6 modulate HIV-1 intra-nuclear invasion and integration. Cell Reports 2015;13:1717–31.

[25] Di Nunzio F. New insights in the role of nucleoporins: a bridge leading to concerted steps from HIV-1 nuclear entry until integration. Virus Res 2013;178:187–96.

[26] Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, et al. Nuclear architecture dictates HIV-1 integration site selection. Nature 2015;521(7551):227–31.

[27] Eagen KP, Hartl TA, Kornberg RD. Stable chromosome condensation revealed by chromosome conformation capture. Cell 2015;163: 934–46.

[28] Pascual-Garcia P, Capelson M. Nuclear pores as versatile platforms for gene regulation. Curr Opin Genetics Dev 2014;25:110–7.

[29] Capelson M, Doucet C, Hetzer MW. Nuclear pore complexes: guardians of the nuclear genome. Cold Spring Harbor Symp Quant Biol 2010;75:585–97.

[30] Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW. Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell 2010;140:372–83.

[31] Ciuffi A, Bushman FD. Retroviral DNA integration: HIV and the role of LEDGF/p75. Trends Genet 2006;22:388–95.

[32] Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, et al. A role for LEDGF/p75 in targeting HIV DNA integration. Nature Med 2005;11:1287–9.

[33] Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M, et al. LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther 2010;18:552–60.

[34] Qin S, Min J. Structure and function of the nucleosome-binding PWWP domain. Trends Biochem Sci 2014;39:536–47.

[35] Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, et al. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS ONE 2007;2:e1340.

[36] Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 2007;21:1767–78.

[37] Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. Molecular mechanisms of retroviral integration site selection. Nucleic Acids Res 2014;42:10209–25.

[38] Gupta SS, Maetzig T, Maertens GN, Sharif A, Rothe M, Weidner-Glunde M, et al. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. J Virol 2013;87:12721–36.

[39] De Rijck J, de Kogel C, Demeulemeester J, Vets S, El Ashkar S, Malani N, et al. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. Cell Reports 2013;5:886–94.

[40] Sharma A, Larue RC, Plumb MR, Malani N, Male F, Slaughter A, et al. BET proteins promote efficient murine leukemia virus integration at transcription start sites. Proc Natl Acad Sci USA 2013;110:12036–41.

[41] Larue RC, Plumb MR, Crowe BL, Shkriabai N, Sharma A, DiFiore J, et al. Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. Nucleic Acids Res 2014;42: 4868–81.

[42] Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD. Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Hum Gene Ther 2006;17:960–7.

[43] Silvers RM, Smith JA, Schowalter M, Litwin S, Liang Z, Geary K, et al. Modification of integration site preferences of an HIV-1-based vector by expression of a novel synthetic protein. Hum Gene Ther 2010;21: 337–49.

[44] Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, et al. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. Proc Natl Acad Sci USA 2010;107:3135–40.

[45] Aiyer S, Swapna GV, Malani N, Aramini JM, Schneider WM, Plumb MR, et al. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. Nucleic Acids Res 2014;42:5917–28.

[46] Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 2007;17: 1186–94.

[47] Benleulmi MS, Matysiak J, Henriquez DR, Vaillant C, Lesbats P, Calmels C, et al. Intasome architecture and chromatin density modulate retroviral integration into nucleosome. Retrovirology 2015;12:13.

[48] Naughtin M, Haftek-Terreau Z, Xavier J, Meyer S, Silvain M, Jaszczyszyn Y, et al. DNA Physical properties and nucleosome positions are major determinants of HIV-1 integrase selectivity. PLoS ONE 2015;10:e0129427.

[49] Serrao E, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral intasomes. Retrovirology 2015;12:39.

[50] Wong RW, Mamede JI, Hope TJ. Impact of nucleoporin-mediated chromatin localization and nuclear architecture on HIV integration site selection. J Virol 2015;89:9702–5.

[51] Campbell EM, Hope TJ. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nature Rev 2015;13:471–83.

[52] Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, et al. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 2013;154:914–27.

[53] Knight S, Collins M, Takeuchi Y. Insertional mutagenesis by retroviral vectors: current concepts and methods of analysis. Curr Gene Ther 2013;13:211–27.

[54] Desfarges S, Ciuffi A. Viral integration and consequences on host gene expression. In: Witzany G, editor. Viruses: essential agents of life. Dordrecht, Netherlands: Springer Science+Business Media; 2012.

[55] Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science 2014;345(6193):179–83.

[56] Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science 2014;345(6196): 570–3.

[57] Ikeda T, Shibata J, Yoshimura K, Koito A, Matsushita S. Recurrent HIV-1 integration at the BACH2 locus in resting CD4[+] T cell populations during effective highly active antiretroviral therapy. J Infect Dis 2007;195:716–25.

[58] Ciuffi A, Mohammadi P, Golumbeanu M, di Iulio J, Telenti A. Bioinformatics and HIV latency. Curr HIV/AIDS Rep 2015;12:97–106.

[59] Mager DL, Stoye JP. Mammalian endogenous retroviruses. Microbiol Spectrum 2015;3. MDNA3-0009-2014.

[60] Suntsova M, Garazha A, Ivanova A, Kaminsky D, Zhavoronkov A, Buzdin A. Molecular functions of human endogenous retroviruses in health and disease. Cell Mol Life Sci 2015;72:3653–75.

[61] Garazha A, Ivanova A, Suntsova M, Malakhova G, Roumiantsev S, Zhavoronkov A, et al. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. Cell Cycle (Georgetown, TX) 2015;14:1476–84.

[62] Imakawa K, Nakagawa S, Miyazawa T. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. Genes Cells 2015;20: 771–88.

[63] Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. J Gen Virol 2015;96:1207–18.

[64] Aftab A, Shah AA, Hashmi AM. Pathophysiological role of HERV-W in schizophrenia. J Neuropsychiatry Clin Neurosci 2016;28:17–25.