

Interactions between Retroviruses and the Host Cell Genome

Valentina Poletti¹ and Fulvio Mavilio²¹Genethon, 1bis rue de l'Internationale, 91002 Evry, France; ²Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy

Replication-defective retroviral vectors have been used for more than 25 years as a tool for efficient and stable insertion of therapeutic transgenes in human cells. Patients suffering from severe genetic diseases have been successfully treated by transplantation of autologous hematopoietic stem-progenitor cells (HSPCs) transduced with retroviral vectors, and the first of this class of therapies, Strimvelis, has recently received market authorization in Europe. Some clinical trials, however, resulted in severe adverse events caused by vector-induced proto-oncogene activation, which showed that retroviral vectors may retain a genotoxic potential associated to proviral integration in the human genome. The adverse events sparked a renewed interest in the biology of retroviruses, which led in a few years to a remarkable understanding of the molecular mechanisms underlying retroviral integration site selection within mammalian genomes. This review summarizes the current knowledge on retrovirus-host interactions at the genomic level, and the peculiar mechanisms by which different retroviruses, and their related gene transfer vectors, integrate in, and interact with, the human genome. This knowledge provides the basis for the development of safer and more efficacious retroviral vectors for human gene therapy.

Integration in the genome of a host cell is a key step in the life cycle of a retrovirus, allowing stable transmission of the viral genome to the host cell progeny and persistent viral gene expression. Retroviral integration is a non-random process whereby the viral RNA genome, reverse transcribed into double-stranded DNA and assembled in a pre-integration complex (PIC), associates to the host cell chromatin and integrates in its proviral form in the genome through the activity of the viral integrase (IN), a specialized protein encoded by the viral *pol* gene.¹ Early studies on viral integration were based on in vitro models, which identified several physical factors playing a role in the process, such as nucleosome-induced DNA bending or steric hindrance by DNA-binding proteins. In vivo, however, viral integration is an active process involving mainly cellular and viral players that have been, and still are, extensively investigated. When the complete sequence of the genomes of several mammals, including humans, became available, PCR-based methods were developed to amplify, clone, and sequence the junctions between proviral and host genome,^{2,3} with the scope of mapping proviruses on genomic DNA and understanding the rules governing the integration process. Thanks to the spectacular development of DNA sequencing technology, ligation-mediated (LM) or linear amplification-mediated (LAM) PCR have been progressively refined,

allowing extensive mapping of viral integration sites (ISs) in mammalian genomes in a quantitative fashion.

The first pioneering studies showed that retroviruses select their genomic target sites in a sequence-independent manner, with preferences for transcribed genes in the case of the HIV³ or gene promoters in the case of the Moloney murine leukemia virus (MLV).⁴ More recently, massive parallel sequencing has greatly increased the resolution of integration maps, while genome-wide association with genetic and epigenetic annotations of the human genome provided crucial clues to the understanding of the molecular determinants of target site selection. In summary, these studies revealed that each retrovirus has a unique and peculiar pattern of integration within mammalian genomes, which is faithfully reproduced by the gene transfer vectors derived thereof. The non-randomness of viral integration patterns has significant implications in terms of biosafety for gene therapy applications.

Retroviruses Have Different Integration Preferences

Retroviruses are divided into seven genera (*alpha*-, *beta*-, *gamma*-, *delta*-, and *epsilon*-*Retroviridae*, *Spumaviridae*, and *Lentiviridae*), and integration characteristics and preferences are known for at least one member of each family, except for *epsilon*-retroviruses. Integration profiles have been defined for members of the lentivirus, spumaretrovirus, *alpha*-retrovirus, and *gamma*-retrovirus genera. Not surprisingly, the extent of available scientific knowledge on target site selection and its determinants directly reflects the clinical relevance of each retrovirus type, thus the HIV and MLV integration profiles are the most extensively characterized.

Based on IS preferences, retroviruses can be classified in roughly three groups: MLV, foamy virus (FV), and human T cell leukemia virus (HTLV), which integrate preferentially around transcription start sites (TSSs) and in transcriptional regulatory elements; HIV and simian immunodeficiency virus (SIV), which preferentially target transcribed sequences; and avian sarcoma-leucosis virus (ASLV), which shows little preference for any genomic feature. The different integration patterns and propensity for specific genomic characteristics indicate the involvement of specific viral and cellular factors in the

<https://doi.org/10.1016/j.omtm.2017.10.001>

Correspondence: Fulvio Mavilio, Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 287, 41125 Modena, Italy.

E-mail: fulvio.mavilio@unimore.it

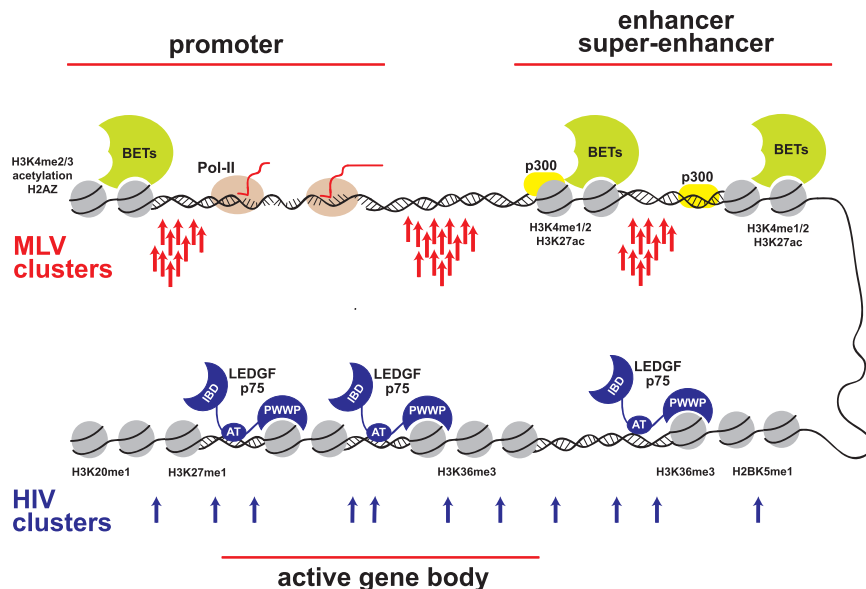


Figure 1. Differential Distribution of Gamma-Retroviral and Lentiviral Integration Sites in the Human Genome

MLV and HIV integration sites are differentially distributed within the genome of a target cell. MLV preferred sites are tightly clustered in and around active regulatory elements (red arrows), whereas HIV preferred sites are spread along transcription units, grouped in larger clusters (blue arrows).

Lentiviruses Target Transcribed Genes

HIV-1 is the best-characterized representative of the *Lentiviridae* family, and the etiological agent of AIDS. As soon as LM-PCR mapping technology became available, the HIV integration pattern in the human genome was characterized,³ showing a strong preference for gene bodies, with up to 80% of the proviruses located inside an active transcriptional unit. As a consequence, HIV integration is influenced by the

transcription pattern of the target cell.^{3,15,16} Higher resolution maps (up to 150,000 ISs) of the integration of HIV-1-derived lentiviral vectors (LVs) were determined on human primary T cells¹⁴ and CD34⁺ hematopoietic stem-progenitor cells (HSPCs).¹⁷ These studies confirmed the HIV-1 preference for transcribed gene bodies, and revealed a negative correlation between integration frequency and TSSs and other typical feature of gene promoters, such as CpG islands, G/C-rich sequences, DNase I hypersensitive sites, and clusters of transcription factor binding sites (TFBSs).^{17,18} The association of ISs with genome-wide maps of histone modifications revealed another interesting feature: HIV-1 integration is strongly associated to epigenetic marks of transcribed gene bodies such as H4K20me1, H3K36me3, H2BK5me1, and H3K27me1.^{14,19,20} Interestingly, HIV-1 has apparently evolved to avoid integrating into active transcriptional regulatory regions, such as promoters and enhancers, as defined by mono-, di-, and tri-methylation of H3K4 and histone hyperacetylation, or in genes controlling cell development and differentiation, including proto-oncogenes¹⁷ (Figure 1). Genes targeted by HIV-1 and LVs are mostly located in euchromatic regions in the outer, membrane-proximal portion of the cell nucleus in close correspondence with the nuclear pore, indicating a strong association between nuclear entry and integration mechanisms.²¹ On the contrary, HIV and LVs strongly disfavor genes located in heterochromatic regions marked by H3K27me3 and H3K9me3,^{14,17,20} including the nuclear lamin-associated heterochromatin.²¹ Transcriptionally active genes are also disfavored if located centrally within the nucleus.²¹

Mechanisms of Lentiviral IS Selection: The Tethering Factors

HIV integration has been the first comprehensively described model of target site selection based on protein-mediated tethering of the retroviral PIC to the host cell chromatin. Several cellular proteins have been isolated as physically bound to lentiviral PICs, and for some of them the association occurs via direct interaction with the

integration process. While several host cell factors play important roles in target site selection, genetic and biochemical evidence indicate that the viral IN is the major viral determinant for most retroviruses. Unsupervised clustering of IS patterns and phylogenetic analysis of INs of six different retroviruses showed that global and local IS preferences correlate with the sequence and/or structure of viral INs,⁵ suggesting a strong correlation between target site selection and viral evolution. The choice of a particular set of ISs is essential for the viral fitness, because it influences not only the transcriptional regulation the provirus will be subjected to, but also any interference of the provirus with the host genome transcription.

Early studies demonstrated that DNA wrapped in nucleosomes is a favored target for retroviral integration compared to naked DNA,^{6–8} with the outward-facing DNA major groove being a preferred target region.⁶ Much is known about the sequence constraints for efficient integration of retroviruses. A dinucleotide CA is invariably positioned exactly 2 bp from both ends of the viral termini. The sequences extending up to 15 bp into the host DNA have significant influence on integration efficiency, although retroviral integration is not a sequence-specific process.^{4,9–12} The first analysis of hundreds of HIV-1, MLV, ASLV, and SIV ISs in mammalian cells showed only a statistically weak palindromic consensus centered on the virus-specific duplicated target site sequence at the insertion sites.¹³ Later analysis of larger datasets confirmed the occurrence of a weakly conserved consensus at the ISs, not recurrent among different retroviruses and enriched when integration is studied on naked DNA in vitro, suggesting that the consensus is mainly selected by the integration machinery itself, most likely reflecting the spatial or energy requirements of the integration complex.^{9,14} These studies indicated that the linear DNA sequence plays a minor role in the integration preferences of retroviruses, which depends essentially on viral and cellular determinants and on their cell-specific availability and interaction.

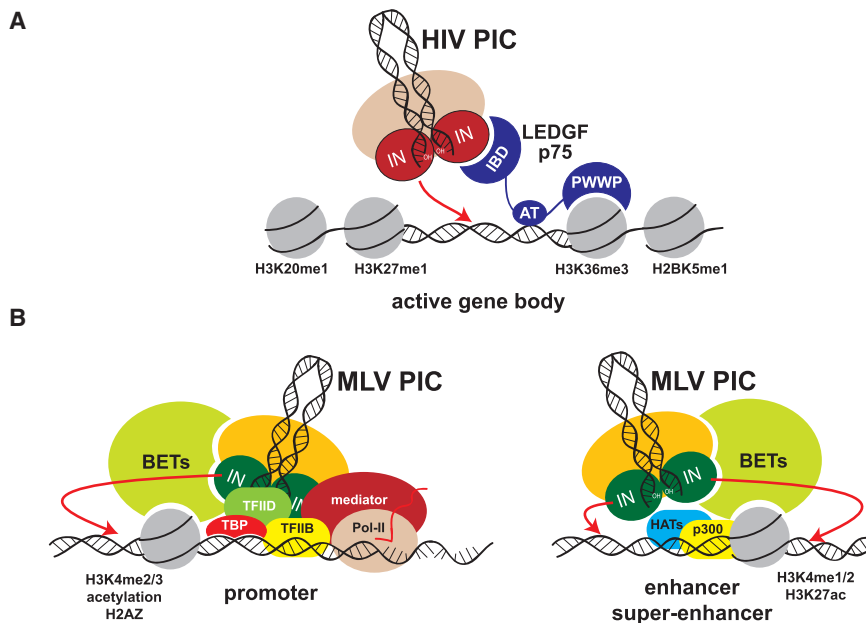


Figure 2. HIV and MLV Pre-integration Complexes Are Tethered to Chromatin by Different Mechanisms

(A) The HIV pre-integration complex (PIC) is tethered to transcribed gene regions (active gene body), marked by specific histone modifications (H3K20me1, H3K27me1, H3K36me3), by the LEDGF-P75 protein, which interacts with the HIV integrase (IN) through an integrase-binding domain (IBD) and with histones through its PWWP domain. An AT-hook domain (AT) mediates interaction with AT-rich DNA sequences on genomic DNA. (B) The MLV PIC is tethered to transcriptionally active promoters (left), enhancers, and super-enhancers (right) through interaction of IN with bromodomain/extraterminal domain proteins (BETs) bound to acetylated histones. A simplified transcription initiation complex is shown on the promoter, including the TATA-binding protein (TBP), the basal transcription factors TFIIIB and TFIIIB, the Mediator complex, and an elongating RNA polymerase II (Pol-II). A simplified histone acetylation complex is shown on the enhancer, including histone acetyl transferases (HATs) and p300. Promoters (left) are marked by H3K4me2/3 histone modifications, histone acetylation, and the H2AZ histone variant. Enhancers (right) are marked by H3K4me1/2 and H3K27ac histone modifications.

HIV-1 IN. These include members of the DNA repair machinery such as hRad1,²² components of chromatin remodeling complexes such as INI1²³ and EED,²⁴ the constitutive chromatin components HMGI(Y),²⁵ and most importantly, the lens epithelium-derived growth factor (PSIP1/LEDGF/p75),²⁶ a ubiquitously expressed nuclear protein tightly associated with chromatin throughout the cell cycle. The role of LEDGF/p75 in mediating HIV infectivity has been deeply investigated because of its tight interaction with the viral IN and its role in stimulating the IN catalytic activity *in vitro*.^{27–29} LEDGF/p75 knockdown significantly reduces HIV-1 infectivity, showing the functional role of this protein as the major cellular binding partner of HIV-1 IN.^{30,31}

LEDGF/p75 was initially discovered as a human transcriptional coactivator³² interacting with a number of cellular proteins, including JPO2,^{31,33} Cdc7-activator of S-phase kinase (ASK),³⁴ the transposase *pogZ*,³⁵ and menin, which links LEDGF/p75 with mixed-lineage leukemia (MLL) histone methyltransferase, causing MLL-dependent transcription and leukemic transformation.³⁶ LEDGF/p75 belongs to the hepatoma-derived growth factor-related protein (HRP) family, which comprises five additional members—HDGF, HRP1-3, and LEDGF/p52, an alternatively spliced, smaller isoform of LEDGF—that, except for HRP2, lack the domains necessary to HIV-1 IN interaction.³⁷ LEDGF/p75 is characterized by a conserved N-terminal PWWP domain, a basic-type nuclear localization signal, two AT-hook DNA binding motifs, and three highly charged regions (CR1-3) that allow to tightly engage chromatin throughout the cell cycle.^{29,38} The C-terminal region contains the IN-binding domain (IBD), which directly interacts with HIV-1 IN.³⁹ The N-terminal PWWP domain is the key determinant for the site-selective association of LEDGF/p75 with chromatin: HIV-1 ISs generated in the pres-

ence of PWWP domain-defective LEDGF/p75 mutants differ substantially from those generated in the presence of wild-type LEDGF/p75.⁴⁰

NMR structures of the LEDGF/p75 PWWP domain revealed two distinct functional interfaces: a hydrophobic pocket that interacts with the H3K36me3 histone tail and an adjacent basic interface that non-specifically engages DNA.^{41,42} Interestingly, the LEDGF/p75 PWWP domain exhibits low binding affinity for either an H3K36me3 peptide or naked DNA, whereas it interacts tightly with mononucleosomes containing an H3K36me3 analog, indicating that cooperative binding of LEDGF/p75 with both the H3K36me3 tail and nucleosomal DNA is essential for its tight and site-selective chromatin association.⁴¹ Indeed, mutations introduced in either the hydrophobic pocket or the basic surface significantly compromised the ability of LEDGF/p75 to both associate with chromatin and stimulate HIV-1 integration.⁴³ These findings collectively indicate that LEDGF/p75-mediated tethering of lentiviral PICs to actively transcribed genes provides IN with increased access to nucleosomal DNA, which are the favored sites for integration both *in vitro* and in infected cells^{6,14,44} (Figure 2A).

Interestingly, LEDGF/p75 fails to interact with INs from other retroviral genera.^{28,45,46} *In vitro* assays with purified INs have revealed that LEDGF/p75 significantly stimulates the strand transfer activity of lentiviral but no other retroviral INs.^{27,29,39,47} Studies using LEDGF/p75 knockout cells revealed 5- to 80-fold defects in HIV-1 infectivity, associated with ~2- to 12-fold reduction in integration.^{48–51} Significant inhibitory effects on HIV-1 replication were also observed in cells engineered to express the LEDGF/p75 IBD domain only.^{52–54} LEDGF/p75 depletion and overexpression of dominant-interfering



IBD constructs do not significantly affect HIV-1 reverse transcription while selectively impairing provirus integration.

Genome-wide IS mapping provided additional evidence for the role of LEDGF/p75 in the selectivity of HIV-1 integration. LEDGF/p75 knockdown significantly reduces integration into active genes,^{46,55} while complete knockout kills infectivity and dramatically changes integration preferences, with a significant percentage of proviruses aberrantly located near TSSs.^{48,49} Interestingly, chimeric LEDGF/p75 proteins can retarget lentiviral integration: replacing the N-terminal PWWP domain and AT hooks with a plant homeodomain redirects HIV-1 integration to TSSs,⁵⁶ while the use of the chromobox homolog 1 (CBX1) and heterochromatin protein 1 (HP1) alpha chromatin-binding modules randomizes the integration pattern.^{56,57} Mapping of the LEDGF/p75 chromatin-binding profile has revealed a preference for binding active transcription units, which paralleled the enhanced HIV-1 integration frequencies at these locations.⁵⁸ Collectively, these findings provide strong evidence that LEDGF/p75 tethers PICs to active transcription units during HIV-1 integration.

Although LEDGF/p75 can potently stimulate HIV-1 IN catalytic function *in vitro*,^{27,29,39,59,60} it is unclear whether it provides this function during natural virus infection. Normal levels of HIV-1 PIC activity are maintained in LEDGF/p75 knockout cells,⁴⁹ indicating that LEDGF/p75 may provide chromatin-tethering functions without contributing to the formation of a catalytically active PIC. HIV-1 integration in transcription units is over-represented even in LEDGF/p75 knockout cells,^{48–50} suggesting a potential role of other cellular proteins in the IS selection. In particular, HRP2 was closely investigated because of its structural similarity with LEDGF/p75: *in vitro* assays with purified proteins demonstrated that HRP2 tightly binds HIV-1 IN and significantly stimulates its catalytic function,³⁹ although, unlike LEDGF/p75, it does not bind chromatin throughout the cell cycle.⁶¹ HRP2 depletion in cells containing normal levels of LEDGF/p75 has no effect on HIV-1 infectivity or integration preferences,^{50,52,62,63} while depletion of both HRP2 and LEDGF/p75 reduces integration into active genes.^{63,64} The preference of HIV-1 for active genes remains greater than random even in LEDGF/HRP2 double-knockout cells, suggesting that additional host factors may play a role in HIV target site selection.⁶³

Lentiviral Integration and Nuclear Import

A striking feature of HIV-1 integration is the preference for large chromosomal domains, or hotspots, rich in active genes,^{3,17,65,66} which is not explained by the chromatin-binding characteristics of LEDGF/p75 or other chromatin interactors. Recent research on the mechanisms of HIV active entry into the cell nucleus showed that the interaction between the HIV PIC and components of the nuclear import machinery plays a key role in tethering HIV integration to its target chromatin, and pointed to nuclear topology as a major determinant of target site selection. The nuclear pore complex (NPC) mediates docking and entry of the HIV into the nucleus through direct interaction of the PIC with nucleoporins, thereby indirectly targeting HIV integration to the pore-proximal chromatin regions.^{21,67} The

nuclear pore is associated with 10- to 500-kb macrogenomic domains of open, actively transcribed chromatin. Knockdown of the Nup358/RanBP2 or Nup153 components of the NPC impair HIV-1 nuclear entry and bias integration toward a non-canonical pattern, indicating that nuclear translocation and IS selection are coupled processes.^{68–70} Interestingly, HIV ISs in cells depleted of the chromatin-proximal Tpr nucleoporin, which is not required for HIV nuclear entry, are less associated with epigenetic marks of transcribed genes (H3K36me3), as revealed by super-resolution microscopy.⁶⁷ Interaction with the NPC is mediated by the HIV-1 capsid protein (CA) encoded by the *Gag* gene. Accordingly, chimeric HIV-1 viruses containing the MLV *Gag* showed a reduced integration frequency into gene-rich regions.^{69–71}

To account for the overall integration preferences of HIV-1, a two-step model has been proposed whereby NPC components direct HIV-1 PICs toward membrane-proximal euchromatic regions of high gene density, whereas LEDGF/p75 tethers the PIC to the transcribed gene bodies within these regions.^{69,70} A recent study showed that the integration hotspots are in fact located in close proximity to the nuclear pore, and that the likelihood of integration into an active gene is a function of its radial distance from the nuclear membrane²¹ (Figure 3). Because the topological organization of chromatin is cell specific, integration hotspots vary from one cell type to another,²¹ reflecting their position with respect to the NPC. Hotspots are therefore essentially generated by the topological organization of chromatin in the cell nucleus, while the local preference for transcribed regions within hotspots is actively determined by the PIC through its tethering interactions primarily with LEDGF/p75 (Figure 3). Because all the integration preferences are mediated by the protein components of the PIC, LVs faithfully reproduce the integration characteristics of parental HIV.

Gamma-Retroviruses Target Transcriptionally Active Promoters and Enhancers

MLV belongs to the gamma-retrovirus family, historically known as *Oncoretroviridae* since the discovery that viruses of this family can transmit leukemia to newborn mice.⁷² Because of their relatively simple structure and efficiency in infecting hematopoietic cells, replication-defective MLV-based vectors were the first gene transfer tools used in gene therapy for hematological disorders. The MLV integration profile has been extensively analyzed in HSPCs and other cell types for many years, particularly after the occurrence of insertional leukemia in patients treated for inherited immunodeficiencies.^{73,74} High-throughput sequencing technology recently extended the number of analyzed MLV ISs into the millions, providing deep insight into the mechanism of MLV target site selection.¹⁸

Initial, low-resolution integration studies showed that MLV and MLV-derived vectors had a modest bias for integration into active genes with a peculiar distribution around TSSs, with ~20% of insertions landing 2.5 kb upstream or downstream of the +1 position of target genes.^{4,15,16,65} Thus, gene promoters were considered for some time as the major target of MLV integration. Later studies

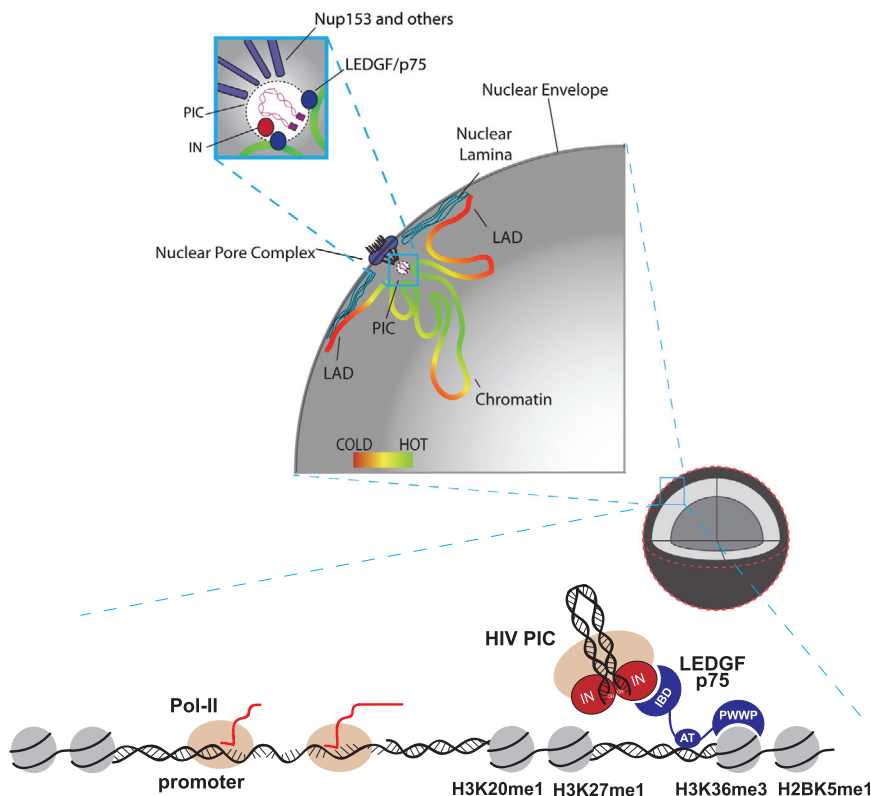


Figure 3. Overview of Lentiviral Nuclear Entry and Integration

The lentiviral pre-integration complex (PIC) docks to and passes through the nuclear pore via the interaction with Nup153 and other nucleoporins, and integrates preferentially in the euchromatic regions (in green) near the nucleopores. The lentiviral PIC is then targeted to regions marked by histone modifications specific of transcribed gene (H3K36me3) through tethering operated by LEDGFp75 (see Figure 2). The hot-cold color scale indicates low (cold) and high (hot) probability of lentiviral integration. IN, HIV-1 integrase; LAD, lamina-associated domains; Nup153, nucleoporin 153. (Modified from Marini et al., 2015.²¹)

gration characteristics suggest that oncoretroviruses may have developed a unique integration strategy that, by coupling target site selection to cell-specific gene regulation, maximizes the chances of maintaining proviral expression in the target cells. In addition, integration around promoters and regulatory elements of cell growth and differentiation may increase the chances of inducing clonal expansion or transformation, and ultimately favor viral propagation. This hypothesis is supported by recent chromatic conformation capture studies showing that MLV ISs co-localize in tridimensional clusters enriched for known cancer genes in the nucleus of target cells,⁸² suggesting that MLV proviruses engage in long-range chromatin interactions that may favor oncogene deregulation.

showed that MLV ISs are enriched in RNA Pol-II binding sites, CpG islands, DNase I hypersensitive sites, TFBSs, phylogenetically conserved non-coding sequences, and binding sites for the p300/CBP histone acetyl transferase, often predictive of *cis*-acting regulatory elements.^{20,71,75} High-definition integration maps showed that regions bound by the Pol-II basal transcriptional machinery, such as core promoters, are protected from the MLV insertion,^{17,20} confirming that integration is directed to TF-bound transcriptionally active elements and not simply to open chromatin regions. Genome-wide integration and association studies showed that the MLV bias for TSSs is just a consequence of a more general preference of MLV PICs for active regulatory elements, particularly enhancers and super-enhancers.^{17,18,76–78} High-resolution studies carried out in human HSPCs and committed progenitors,^{17,18,78} T cells,^{20,79} and keratinocytes⁷⁷ indicated that MLV ISs are highly clustered and occur almost exclusively in regions carrying epigenetic signatures of transcriptionally active regulatory regions, such as acetylation of H3K27 and H3K9, and mono-, di- and tri-methylation of H3K4 (Figure 1). Symmetrically, typical heterochromatic marks, such as H3K27me3 and H3K9me3, are significantly under-represented at these genomic loci.^{14,17,20,76} As a consequence, the MLV integration pattern is strongly cell specific and depends on the enhancer and promoter usage of any given cell type: all studies invariably showed a correlation between MLV integration and gene expression levels,^{3,16,17,20,65,80} and a significant bias toward functional categories such as developmental regulation, cell growth, and cell differentiation.^{20,65,76,80,81} These inte-

Mechanism of Gamma-Retroviral Target Site Selection

As for HIV, at least two determinants are involved in MLV target site selection, the viral PIC and the host cell factors that tether it to chromatin. The viral determinants of MLV target site selection have been extensively investigated by genetic and biochemical analysis. Early studies indicated that an HIV-1 vector packaged with an MLV IN acquires the MLV-specific bias for TSSs, CpG islands and TFBS-rich regions,^{71,75} identifying the IN protein as the main viral determinant of MLV target site selection. Given the preference for active transcriptional regulatory elements, components of the basal RNA Pol-II transcriptional machinery were obvious candidates as cell-derived tethering factors. Early data indicated that the protein BAF (barrier to autointegration factor) is physically associated to the MLV PICs.⁸³ BAF was originally identified as an inhibitor of suicide integration of the MLV provirus, which promotes efficient intermolecular DNA recombination.⁸⁴ Although essential for PIC integration activity, interaction with BAF did not explain the MLV integration preferences. A yeast two-hybrid screen of proteins potentially interacting with the MLV IN later provided a number of potential binding targets, many of which were indeed components of active chromatin and transcription complexes.⁸⁵ More recent mass spectrometry-based proteomic analysis of human cellular proteins co-purifying with

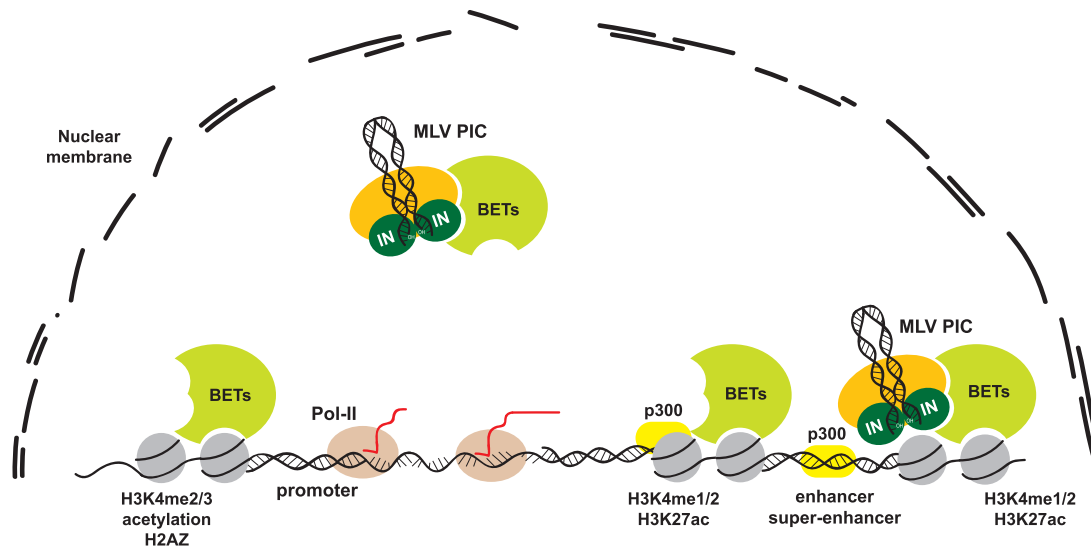


Figure 4. Overview of the Genomic Characteristics Favoring Gamma-Retroviral Integration

Retroviral integration occurs preferentially in active transcriptional regulatory elements (promoters, enhancers, and super-enhancers), associated to specific histone modifications (H3K4me1-3, H3K27ac) and bound by histone acetyltransferases and bromodomain/extraterminal domain-containing proteins (CBP/p300, BETs) (see Figure 2). The dotted line indicates the disassembled nuclear membrane, necessary for the MLV PIC to access chromatin. IN, MLV integrase; MLV, Moloney leukemia virus.

recombinant MLV IN identified the bromodomain-containing BET proteins BRD2, BRD3, and BRD4 as main binding partners of the viral IN protein⁸⁶ (Figure 2B). BRD2 was in fact one of the interactors identified by the yeast two-hybrid screening.⁸⁵ Enhancers are the major source of BRD4-dependent transcriptional activation,⁸⁷ and genes under the control of super-enhancers are particularly sensitive to BET inhibition.⁸⁸

BRD2, BRD3, BRD4, and BRDT belong to the extended BET protein family that includes BRD1, BRD7, BRD8, and BRD9. BRD2, BRD3, and BRD4 are ubiquitously expressed, whereas BRDT is only expressed in the testis. BET proteins have been implicated in transcription, DNA replication, and cell-cycle control.⁸⁹ They exhibit two N-terminal bromodomains (BD-I and BD-II) that bind acetylated H3 and H4 tails,⁹⁰ but not with their unmodified counterparts,⁹¹ and two conserved motifs, A and B, which are positively charged and could contribute to DNA binding.⁹² The simultaneous binding to acetylated histone tails and nucleosomal DNA could be a generic mechanism used by chromatin tethers to achieve tighter and more specific interactions with chromatin. The conserved C-terminal extra-terminal (ET) and SEED, Ser/Glu/Asp-rich, domains directly bind several cellular proteins including transcription factors, chromatin-modifying proteins, and histone modification enzymes.⁸⁹ BET proteins directly bind acetylated histone tails, so their association with strong enhancers is mediated by direct interactions with H3K27ac and H3K9ac and/or with other chromatin readers.⁸⁸ Chromatin binding sites of BET proteins have been mapped by chromatin immunoprecipitation sequencing (ChIP-seq) experiments⁹³ and show a positive correlation with MLV, but not with HIV-1 or ASLV ISs.⁹⁴

Binding of BRD2, BRD3, and BRD4 to the MLV IN is mediated by the ET domain.^{86,94,95} Solution NMR and protein interaction studies showed that the unstructured C-terminal polypeptide (TP) domain of the MLV IN directly binds the BET ET domain and acquires a secondary structure upon complex formation.⁹⁶ Deletion of the TP domain does not disrupt the catalytic properties of IN but widely alters the MLV targeting profile in human 293T cells, with significantly reduced association to TSSs, CpG islands, and BRD2, BRD3, and BRD4 binding sites.⁹⁶ Downregulation of BET protein expression by small interfering RNAs (siRNAs), or inhibition of their binding to chromatin by the small-molecule inhibitor JQ-1, dramatically alters the canonical MLV integration profile in HEK293T cells, with a dose-dependent, 4-fold reduction of the integrations around TSSs.^{94,95} Complementary, genetic evidence for the role of BET association in directing MLV integration was provided by expressing an artificial LEDGFp75-BET fusion protein in murine NIH 3T3 and human SupT1 cells, which retargeted MLV integration into actively transcribed gene bodies, mimicking the HIV-1 integration pattern.⁸⁶

Overall, these studies demonstrate that BET proteins are the MLV counterpart of LEDGF/p75 for HIV-1, i.e., specific chromatin tethers that interact with the PIC by binding the IN and possibly stimulating its enzymatic function (Figures 2 and 4). Although the interaction of MLV IN with BET proteins is probably the main determinant of gamma-retroviral integration target site selection, it is remarkable that even in the presence of potent inhibitors of BET activity such as JQ-1, integration at active regulatory elements is still substantially higher compared to random controls or to HIV-1.⁹⁵ This suggests that additional host and/or viral factors may contribute to the integration preferences of MLV. Early studies with chimeric HIV viruses



packaged with MLV Gag proteins had suggested that other components of the PIC, in particular the p12 Gag protein, may have a role in MLV target site selection.⁷¹ The MLV p12 is a constituent of the PIC, but its function in the complex remains ill-defined. Imaging of the MLV PIC traffic in live cells allowed the visualization of its docking to mitotic chromosomes and its release upon exit from mitosis. Docking occurs concomitantly with the breakdown of the nuclear envelope and is impaired in PICs carrying lethal p12 mutations. The insertion of a heterologous chromatin binding module into p12 restores PIC attachment to chromosomes, confirming the role of p12 as a tethering factor to mitotic chromosomes.⁹⁷ Later studies, however, demonstrated that mutations altering the p12 interaction with chromatin have no detectable effects on the MLV integration pattern,⁹⁸ indicating that p12 plays a role in MLV integration by allowing chromosome association but has no role in the target site selection process.

Alpha-Retroviruses Integrate Almost Randomly in Mammalian Cells

ASLV is a member of the alpha-retrovirus family whose natural host is chicken. The natural viral host tropism can be altered in ASLV-derived vectors by envelope pseudotyping, in order to efficiently infect mammalian cells.⁹⁹ Genome-wide analysis of the integration pattern of alpha-retroviral vectors in human CD34⁺ HSPCs revealed a weak preference for open chromatin regions, as in the case of gamma-retroviral and LVs, but no bias for transcribed genes or transcriptional regulatory elements, as defined by the epigenetic marks H3K4me1 and H3K4me3.¹⁰⁰ The nearly random ASLV insertion profile in mammals has encouraged the development of optimized vectors as gene transfer tools for gene therapy applications.¹⁰¹ However, the biology of ASLV-host interaction is much less studied compared to HIV or MLV, and nothing is known about the viral and cellular determinants of its association to chromatin.

Integration Preferences of Other Retroviruses

The mouse mammary tumor virus (MMTV) is a representative of the beta-retrovirus family. The only MMTV integration profile so far described was obtained by a very limited (<500) set of ISs in the human and murine genomes, essentially from infected mammary (murine NmuMG and human Hs587T) and non-mammary (HeLa and HEK293) cell lines. In all cases, MMTV displayed the most random IS distribution among all known retroviruses, with no preference for active genes, TSSs, gene-dense regions, CpG islands, or DNase hypersensitive sites.^{102,103}

The human T cell leukemia virus type 1 (HTLV-1) is the only component of the delta-retroviral family for which an integration profile has been determined to date. Initially, 541 unique HTLV-1 ISs were cloned and mapped in HeLa cells,⁵ showing that the virus integrates into the human genome with a modest though significant preference for TSSs, transcription units, promoters, and gene-dense regions. More recent data, obtained from >2,100 ISs of HTLV-1 in the human Jurkat T cell line, revealed a strong bias for genes, promoters (identified by CpG islands), and epigenetic marks associated to gene expres-

sion control.¹⁰⁴ The same pattern was observed in vivo in peripheral blood mononuclear cells of HTLV⁺ patients, but mitigated by a strong selection against clones harboring a highly expressed provirus operated by the host immune response against the HTLV-1 infection.¹⁰⁴

The human FV is the prototype of the *Spumaviridae* genus of retroviruses. First isolated from a human cell line, it was later reported to have a chimpanzee origin.¹⁰⁵ FV vectors have been designed and pseudotyped to have a broad host range, large packaging capacity, and high transduction efficiency in human hematopoietic cells, and developed as an alternative to MLV vectors for gene therapy of hematological disorders, particularly AIDS.¹⁰⁶ Low-resolution profiling of FV ISs showed integration preferences similar to, though weaker than, those of MLV for CpG islands and TSSs,^{107,108} confirmed by more recent analysis on human HSPCs after transplantation in immunodeficient mice.¹⁰⁹ A study analyzed 139 FV-derived vector insertion sites recovered ex vivo from human HSPCs after transplantation in NOD/SCID mice, which showed a weak preference for integration close to oncogenes.¹¹⁰ Given the very small size of the analyzed samples and the confounding effect of the in vivo selection, it is difficult to conclude from these studies whether FV may have evolved a tethering mechanism similar to that of MLV. The FV Gag contains a C-terminal chromatin-binding site (CBS) that allows the interaction of PICs with the core histones H2A and H2B on host chromatin.¹¹¹ FV PICs can be retargeted to satellite elements and H3K9me3-positive heterochromatin by modifying the Gag CBS and the IN proteins,¹¹² indicating that a Gag/IN-based tethering mechanism is most likely at the basis of virus-host interaction also in the case of spumaviruses.

Conclusions

A large body of studies on the biochemistry and molecular genetics of retroviral integration has provided a wealth of information on the remarkably complex and evolved patterns of interactions between the retroviral integration machinery and the mammalian, and particularly human, genome. This knowledge has helped in understanding the basis of the “genotoxic” integration events that caused the occurrence of malignant and myelodysplastic proliferation in some gene therapy trials for inherited blood diseases, and provided clues for the design of safer gene transfer tools. The current clinical applications of retroviral gene transfer technology, either ex vivo or in vivo, involve the transduction of millions of cells, in most cases with stem-progenitor characteristics, and the generation of a very high number of potentially mutagenic insertion events. The integration preferences of each vector type make some of these events more or less likely to happen, but given the numbers involved, even a completely random integration machinery would have only a few-fold lower probability of inducing a potentially oncogenic or otherwise dangerous mutation than an MLV- or an HIV-based vector. This evidence points to the design of the vector and the nature of the sequences it carries into the genome as major factors determining its potential genotoxicity: replacement of potent, long-range-acting promoter/enhancer elements of viral origin with constitutive promoters of cellular origin dramatically reduces the potential mutagenic effect of vector



integration.^{113,114} The development of high-throughput technology for mapping viral ISs in small samples has made long-term monitoring of the dynamics of genetically modified cells in patients a reality, increasing the safety of using viral vectors in the clinics while providing new knowledge on stem cell dynamics.¹¹⁵ Many approaches have been proposed in the last few years to replace transgenesis based on viral INs with more “intelligent” machineries achieving site-directed rather than semi-random integration, and gene correction rather than gene addition. Although extremely promising, these techniques will probably take years before matching the unsurpassed transduction efficiency of retroviral vectors and becoming practically applicable in a clinical context. For the time being, improving the safety profile of retroviral vectors and predicting and evaluating at best the risks and benefits of each specific therapeutic application remain the priorities in clinical gene therapy.

REFERENCES

- Coffin, J.M., Huges, S.H., and Varmus, H.E. (1997). *Retroviruses* (Cold Spring Harbor Laboratory Press).
- Schmidt, M., Hoffmann, G., Wissler, M., Lemke, N., Müssig, A., Glimm, H., Williams, D.A., Ragg, S., Hesemann, C.U., and von Kalle, C. (2001). Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. *Hum. Gene Ther.* *12*, 743–749.
- Schröder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* *110*, 521–529.
- Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* *300*, 1749–1751.
- Derse, D., Crise, B., Li, Y., Princler, G., Lum, N., Stewart, C., McGrath, C.F., Hughes, S.H., Munroe, D.J., and Wu, X. (2007). Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *J. Virol.* *81*, 6731–6741.
- Pruss, D., Bushman, F.D., and Wolffe, A.P. (1994). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* *91*, 5913–5917.
- Pryciak, P.M., and Varmus, H.E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* *69*, 769–780.
- Taganov, K.D., Cuesta, I., Daniel, R., Cirillo, L.A., Katz, R.A., Zaret, K.S., and Skalka, A.M. (2004). Integrase-specific enhancement and suppression of retroviral DNA integration by compacted chromatin structure in vitro. *J. Virol.* *78*, 5848–5855.
- Berry, C., Hannehalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* *2*, e157.
- Carteau, S., Hoffmann, C., and Bushman, F. (1998). Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* *72*, 4005–4014.
- Holman, A.G., and Coffin, J.M. (2005). Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. USA* *102*, 6103–6107.
- Stevens, S.W., and Griffith, J.D. (1996). Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* *70*, 6459–6462.
- Wu, X., Li, Y., Crise, B., Burgess, S.M., and Munroe, D.J. (2005). Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* *79*, 5211–5214.
- Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* *17*, 1186–1194.
- Hematti, P., Hong, B.K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I.E., Eckfeldt, C.E., Sharma, Y., Schmidt, M., et al. (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* *2*, e423.
- Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* *2*, E234.
- Cattoglio, C., Pellin, D., Rizzi, E., Maruggi, G., Corti, G., Miselli, F., Sartori, D., Guffanti, A., Di Serio, C., Ambrosi, A., et al. (2010). High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* *116*, 5507–5517.
- De Ravin, S.S., Su, L., Theobald, N., Choi, U., Macpherson, J.L., Poidinger, M., Symonds, G., Pond, S.M., Ferris, A.L., Hughes, S.H., et al. (2014). Enhancers are major targets for murine leukemia virus vector integration. *J. Virol.* *88*, 4504–4513.
- Wang, G.P., Levine, B.L., Binder, G.K., Berry, C.C., Malani, N., McGarrity, G., Tebas, P., June, C.H., and Bushman, F.D. (2009). Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. *Mol. Ther.* *17*, 844–850.
- Cattoglio, C., Maruggi, G., Bartholomae, C., Malani, N., Pellin, D., Cocchiarella, F., Magnani, Z., Ciceri, F., Ambrosi, A., von Kalle, C., et al. (2010). High-definition mapping of retroviral integration sites defines the fate of allogeneic T cells after donor lymphocyte infusion. *PLoS ONE* *5*, e15688.
- Marini, B., Kertesz-Farkas, A., Ali, H., Lucic, B., Lisek, K., Manganaro, L., Pongor, S., Luzzati, R., Recchia, A., Mavilio, F., et al. (2015). Nuclear architecture dictates HIV-1 integration site selection. *Nature* *521*, 227–231.
- Mulder, L.C., Chakrabarti, L.A., and Muesing, M.A. (2002). Interaction of HIV-1 integrase with DNA repair protein hRad18. *J. Biol. Chem.* *277*, 27489–27493.
- Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R., and Goff, S.P. (1994). Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* *266*, 2002–2006.
- Violot, S., Hong, S.S., Rakotobe, D., Petit, C., Gay, B., Moreau, K., Billaud, G., Priet, S., Sire, J., Schwartz, O., et al. (2003). The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1. *J. Virol.* *77*, 12507–12522.
- Li, L., Yoder, K., Hansen, M.S., Olvera, J., Miller, M.D., and Bushman, F.D. (2000). Retroviral cDNA integration: stimulation by HMG I family proteins. *J. Virol.* *74*, 10965–10974.
- Engelman, A., and Cherepanov, P. (2008). The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog.* *4*, e1000046.
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* *278*, 372–381.
- Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempé, D., Vandekerckhove, L., Moisan, F., Ben-Slama, L., Witvrouw, M., et al. (2005). Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J. Biol. Chem.* *280*, 25517–25523.
- Turlure, F., Maertens, G., Rahman, S., Cherepanov, P., and Engelman, A. (2006). A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res.* *34*, 1653–1665.
- Llano, M., Delgado, S., Vanegas, M., and Poeschla, E.M. (2004). Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *J. Biol. Chem.* *279*, 55570–55577.
- Maertens, G.N., Cherepanov, P., and Engelman, A. (2006). Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J. Cell Sci.* *119*, 2563–2571.
- Ge, H., Si, Y., and Roeder, R.G. (1998). Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J.* *17*, 6723–6729.



33. Bartholomeeusen, K., De Rijck, J., Busschots, K., Desender, L., Gijbbers, R., Emiliani, S., Benarous, R., Debyser, Z., and Christ, F. (2007). Differential interaction of HIV-1 integrase and JPO2 with the C terminus of LEDGF/p75. *J. Mol. Biol.* *372*, 407–421.
34. Hughes, S., Jenkins, V., Dar, M.J., Engelman, A., and Cherepanov, P. (2010). Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. *J. Biol. Chem.* *285*, 541–554.
35. Bartholomeeusen, K., Christ, F., Hendrix, J., Rain, J.C., Emiliani, S., Benarous, R., Debyser, Z., Gijbbers, R., and De Rijck, J. (2009). Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PoggZ. *J. Biol. Chem.* *284*, 11467–11477.
36. Yokoyama, A., and Cleary, M.L. (2008). Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* *14*, 36–46.
37. Maertens, G., Cherepanov, P., Plumers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* *278*, 33528–33539.
38. Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S., and Poeschla, E.M. (2006). Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J. Mol. Biol.* *360*, 760–773.
39. Cherepanov, P., Devroe, E., Silver, P.A., and Engelman, A. (2004). Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.* *279*, 48883–48892.
40. Gijbbers, R., Vets, S., De Rijck, J., Ocwieja, K.E., Ronen, K., Malani, N., Bushman, F.D., and Debyser, Z. (2011). Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF)/p75 cofactor in lentiviral integration targeting. *J. Biol. Chem.* *286*, 41812–41825.
41. Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S., et al. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* *41*, 3924–3936.
42. van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M., and van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* *6*, 12.
43. Shun, M.C., Botbol, Y., Li, X., Di Nunzio, F., Daigle, J.E., Yan, N., Lieberman, J., Lavigne, M., and Engelman, A. (2008). Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J. Virol.* *82*, 11555–11567.
44. Pruss, D., Reeves, R., Bushman, F.D., and Wolfé, A.P. (1994). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* *269*, 25031–25041.
45. Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., and Debyser, Z. (2005). The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* *280*, 17841–17847.
46. Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M., and Poeschla, E.M. (2004). LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* *78*, 9524–9537.
47. Cherepanov, P. (2007). LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* *35*, 113–124.
48. Marshall, H.M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E., and Bushman, F.D. (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* *2*, e1340.
49. Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.* *21*, 1767–1778.
50. Schrijvers, R., De Rijck, J., Demeulemeester, J., Adachi, N., Vets, S., Ronen, K., Christ, F., Bushman, F.D., Debyser, Z., and Gijbbers, R. (2012). LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs. *PLoS Pathog.* *8*, e1002558.
51. Fadel, H.J., Morrison, J.H., Saenz, D.T., Fuchs, J.R., Kvaratskhelia, M., Ekker, S.C., and Poeschla, E.M. (2014). TALEN knockout of the PSIP1 gene in human cells: analyses of HIV-1 replication and allosteric integrase inhibitor mechanism. *J. Virol.* *88*, 9704–9717.
52. Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., and Poeschla, E.M. (2006). An essential role for LEDGF/p75 in HIV integration. *Science* *314*, 461–464.
53. De Rijck, J., Vandekerckhove, L., Gijbbers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F., and Debyser, Z. (2006). Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J. Virol.* *80*, 11498–11509.
54. Meehan, A.M., Saenz, D.T., Morrison, J., Hu, C., Peretz, M., and Poeschla, E.M. (2011). LEDGF dominant interference proteins demonstrate prenuclear exposure of HIV-1 integrase and synergize with LEDGF depletion to destroy viral infectivity. *J. Virol.* *85*, 3570–3583.
55. Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* *11*, 1287–1289.
56. Ferris, A.L., Wu, X., Hughes, C.M., Stewart, C., Smith, S.J., Milne, T.A., Wang, G.G., Shun, M.C., Allis, C.D., Engelman, A., and Hughes, S.H. (2010). Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc. Natl. Acad. Sci. USA* *107*, 3135–3140.
57. Gijbbers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., McNeely, M., Bushman, F.D., and Debyser, Z. (2010). LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. *Mol. Ther.* *18*, 552–560.
58. De Rijck, J., Bartholomeeusen, K., Ceulemans, H., Debyser, Z., and Gijbbers, R. (2010). High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region. *Nucleic Acids Res.* *38*, 6135–6147.
59. Pandey, K.K., Sinha, S., and Grandgenett, D.P. (2007). Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration. *J. Virol.* *81*, 3969–3979.
60. McKee, C.J., Kessl, J.J., Shkriabai, N., Dar, M.J., Engelman, A., and Kvaratskhelia, M. (2008). Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J. Biol. Chem.* *283*, 31802–31812.
61. Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M., and Poeschla, E. (2005). Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J. Cell Sci.* *118*, 1733–1743.
62. Vandegraaff, N., Devroe, E., Turlure, F., Silver, P.A., and Engelman, A. (2006). Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology* *346*, 415–426.
63. Wang, H., Jurado, K.A., Wu, X., Shun, M.C., Li, X., Ferris, A.L., Smith, S.J., Patel, P.A., Fuchs, J.R., Cherepanov, P., et al. (2012). HRP2 determines the efficiency and specificity of HIV-1 integration in LEDGF/p75 knockout cells but does not contribute to the antiviral activity of a potent LEDGF/p75-binding site integrase inhibitor. *Nucleic Acids Res.* *40*, 11518–11530.
64. Schrijvers, R., Vets, S., De Rijck, J., Malani, N., Bushman, F.D., Debyser, Z., and Gijbbers, R. (2012). HRP-2 determines HIV-1 integration site selection in LEDGF/p75 depleted cells. *Retrovirology* *9*, 84.
65. Cattoglio, C., Facchini, G., Sartori, D., Antonelli, A., Miccio, A., Cassani, B., Schmidt, M., von Kalle, C., Howe, S., Thrasher, A.J., et al. (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* *110*, 1770–1778.
66. Ambrosi, A., Glad, I.K., Pellin, D., Cattoglio, C., Mavilio, F., Di Serio, C., and Frigessi, A. (2011). Estimated comparative integration hotspots identify different behaviors of retroviral gene transfer vectors. *PLoS Comput. Biol.* *7*, e1002292.
67. Lelek, M., Casartelli, N., Pellin, D., Rizzi, E., Souque, P., Severgnini, M., Di Serio, C., Fricke, T., Diaz-Griffero, F., Zimmer, C., et al. (2015). Chromatin organization at the nuclear pore favours HIV replication. *Nat. Commun.* *6*, 6483.
68. Di Nunzio, F., Fricke, T., Miccio, A., Valle-Casuso, J.C., Perez, P., Souque, P., Rizzi, E., Severgnini, M., Mavilio, F., Charneau, P., and Diaz-Griffero, F. (2013). Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology* *440*, 8–18.



69. Koh, Y., Wu, X., Ferris, A.L., Matreyek, K.A., Smith, S.J., Lee, K., KewalRamani, V.N., Hughes, S.H., and Engelman, A. (2013). 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.* *87*, 648–658.
70. Ocwieja, K.E., Brady, T.L., Ronen, K., Huegel, A., Roth, S.L., Schaller, T., James, L.C., Towers, G.J., Young, J.A., Chanda, S.K., et al. (2011). HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog.* *7*, e1001313.
71. Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* *2*, e60.
72. Moloney, J.B. (1960). Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. I. Origin and introductory investigations. *J. Natl. Cancer Inst.* *24*, 933–951.
73. Cavazza, A., Moiani, A., and Mavilio, F. (2013). Mechanisms of retroviral integration and mutagenesis. *Hum. Gene Ther.* *24*, 119–131.
74. Demeulemeester, J., De Rijck, J., Gijssbers, R., and Debyser, Z. (2015). Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. *BioEssays* *37*, 1202–1214.
75. Felice, B., Cattoglio, C., Cittaro, D., Testa, A., Miccio, A., Ferrari, G., Luzi, L., Recchia, A., and Mavilio, F. (2009). Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. *PLoS ONE* *4*, e4571.
76. Biasco, L., Ambrosi, A., Pellin, D., Bartholomae, C., Brigida, I., Roncarolo, M.G., Di Serio, C., von Kalle, C., Schmidt, M., and Aiuti, A. (2011). Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. *EMBO Mol. Med.* *3*, 89–101.
77. Cavazza, A., Miccio, A., Romano, O., Petiti, L., Malagoli Tagliazucchi, G., Peano, C., Severgnini, M., Rizzi, E., De Bellis, G., Bicciato, S., and Mavilio, F. (2016). Dynamic transcriptional and epigenetic regulation of human epidermal keratinocyte differentiation. *Stem Cell Reports* *6*, 618–632.
78. Romano, O., Peano, C., Tagliazucchi, G.M., Petiti, L., Poletti, V., Cocchiarella, F., Rizzi, E., Severgnini, M., Cavazza, A., Rossi, C., et al. (2016). Transcriptional, epigenetic and retroviral signatures identify regulatory regions involved in hematopoietic lineage commitment. *Sci. Rep.* *6*, 24724.
79. Roth, S.L., Malani, N., and Bushman, F.D. (2011). Gammaretroviral integration into nucleosomal target DNA in vivo. *J. Virol.* *85*, 7393–7401.
80. Aiuti, A., Cassani, B., Andolfi, G., Miolo, M., Biasco, L., Recchia, A., Urbinati, F., Valacca, C., Scaramuzza, S., Aker, M., et al. (2007). Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *J. Clin. Invest.* *117*, 2233–2240.
81. Deichmann, A., Brugman, M.H., Bartholomae, C.C., Schwarzwaelder, K., Versteegen, M.M., Howe, S.J., Arens, A., Ott, M.G., Hoelzer, D., Seger, R., et al. (2011). Insertion sites in engrafted cells cluster within a limited repertoire of genomic areas after gammaretroviral vector gene therapy. *Mol. Ther.* *19*, 2031–2039.
82. Babaei, S., Akhtar, W., de Jong, J., Reinders, M., and de Ridder, J. (2015). 3D hotspots of recurrent retroviral insertions reveal long-range interactions with cancer genes. *Nat. Commun.* *6*, 6381.
83. Lin, C.W., and Engelman, A. (2003). The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J. Virol.* *77*, 5030–5036.
84. Lee, M.S., and Craigie, R. (1994). Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc. Natl. Acad. Sci. USA* *91*, 9823–9827.
85. Studamire, B., and Goff, S.P. (2008). Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology* *5*, 48.
86. De Rijck, J., de Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N., Bushman, F.D., Landuyt, B., Husson, S.J., Busschots, K., et al. (2013). The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* *5*, 886–894.
87. Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastrius, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* *146*, 904–917.
88. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.L., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* *153*, 320–334.
89. Belkina, A.C., and Denis, G.V. (2012). BET domain co-regulators in obesity, inflammation and cancer. *Nat. Rev. Cancer* *12*, 465–477.
90. Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* *468*, 1067–1073.
91. Umehara, T., Nakamura, Y., Jang, M.K., Nakano, K., Tanaka, A., Ozato, K., Padmanabhan, B., and Yokoyama, S. (2010). Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J. Biol. Chem.* *285*, 7610–7618.
92. Larue, R.C., Plumb, M.R., Crowe, B.L., Shkriabai, N., Sharma, A., DiFiore, J., Malani, N., Aiyyer, S.S., Roth, M.J., Bushman, F.D., et al. (2014). Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. *Nucleic Acids Res.* *42*, 4868–4881.
93. LeRoy, G., Chepelev, I., DiMaggio, P.A., Blanco, M.A., Zee, B.M., Zhao, K., and Garcia, B.A. (2012). Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biol.* *13*, R68.
94. Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessler, J.J., Shkriabai, N., Coward, E., Aiyyer, S.S., et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. USA* *110*, 12036–12041.
95. Gupta, S.S., Maetzig, T., Maertens, G.N., Sharif, A., Rothe, M., Weidner-Glunde, M., Galla, M., Schambach, A., Cherepanov, P., and Schulz, T.F. (2013). Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* *87*, 12721–12736.
96. Aiyyer, S., Swapna, G.V., Malani, N., Aramini, J.M., Schneider, W.M., Plumb, M.R., Ghanem, M., Larue, R.C., Sharma, A., Studamire, B., et al. (2014). Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res.* *42*, 5917–5928.
97. Elis, E., Ehrlich, M., Prizan-Ravid, A., Laham-Karam, N., and Bacharach, E. (2012). p12 tethers the murine leukemia virus pre-integration complex to mitotic chromosomes. *PLoS Pathog.* *8*, e1003103.
98. Schneider, W.M., Brzezinski, J.D., Aiyyer, S., Malani, N., Gyuricz, M., Bushman, F.D., and Roth, M.J. (2013). Viral DNA tethering domains complement replication-defective mutations in the p12 protein of MuLV Gag. *Proc. Natl. Acad. Sci. USA* *110*, 9487–9492.
99. Hu, J., Renaud, G., Gomes, T.J., Ferris, A., Hendrie, P.C., Donahue, R.E., Hughes, S.H., Wolfsberg, T.G., Russell, D.W., and Dunbar, C.E. (2008). Reduced genotoxicity of avian sarcoma leukosis virus vectors in rhesus long-term repopulating cells compared to standard murine retrovirus vectors. *Mol. Ther.* *16*, 1617–1623.
100. Moiani, A., Suerth, J.D., Gandolfi, F., Rizzi, E., Severgnini, M., De Bellis, G., Schambach, A., and Mavilio, F. (2014). Genome-wide analysis of alpharetroviral integration in human hematopoietic stem/progenitor cells. *Genes (Basel)* *5*, 415–429.
101. Suerth, J.D., Labenski, V., and Schambach, A. (2014). Alpharetroviral vectors: from a cancer-causing agent to a useful tool for human gene therapy. *Viruses* *6*, 4811–4838.
102. Faschinger, A., Rouault, F., Sollner, J., Lukas, A., Salmons, B., Günzburg, W.H., and Indik, S. (2008). Mouse mammary tumor virus integration site selection in human and mouse genomes. *J. Virol.* *82*, 1360–1367.
103. Konstantoulas, C.J., and Indik, S. (2014). Mouse mammary tumor virus-based vector transduces non-dividing cells, enters the nucleus via a TNPO3-independent pathway and integrates in a less biased fashion than other retroviruses. *Retrovirology* *11*, 34.
104. Gillet, N.A., Malani, N., Melamed, A., Gormley, N., Carter, R., Bentley, D., Berry, C., Bushman, F.D., Taylor, G.P., and Bangham, C.R. (2011). The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* *117*, 3113–3122.
105. Heneine, W., Switzer, W.M., Sandstrom, P., Brown, J., Vedapuri, S., Schable, C.A., Khan, A.S., Lerche, N.W., Schweizer, M., Neumann-Haefelin, D., et al. (1998).



- Identification of a human population infected with simian foamy viruses. *Nat. Med.* 4, 403–407.
106. Olszko, M.E., and Trobridge, G.D. (2013). Foamy virus vectors for HIV gene therapy. *Viruses* 5, 2585–2600.
 107. Trobridge, G.D., Miller, D.G., Jacobs, M.A., Allen, J.M., Kiem, H.P., Kaul, R., and Russell, D.W. (2006). Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci. USA* 103, 1498–1503.
 108. Nowrouzi, A., Dittrich, M., Klanke, C., Heinkelein, M., Rammling, M., Dandekar, T., von Kalle, C., and Rethwilm, A. (2006). Genome-wide mapping of foamy virus vector integrations into a human cell line. *J. Gen. Virol.* 87, 1339–1347.
 109. Nasimuzzaman, M., Kim, Y.S., Wang, Y.D., and Persons, D.A. (2014). High-titer foamy virus vector transduction and integration sites of human CD34(+) cell-derived SCID-repopulating cells. *Mol. Ther. Methods Clin. Dev.* 1, 14020.
 110. Olszko, M.E., Adair, J.E., Linde, I., Rae, D.T., Trobridge, P., Hocum, J.D., Rawlings, D.J., Kiem, H.P., and Trobridge, G.D. (2015). Foamy viral vector integration sites in SCID-repopulating cells after MGMP140K-mediated in vivo selection. *Gene Ther.* 22, 591–595.
 111. Tobaly-Tapiero, J., Bittoun, P., Lehmann-Che, J., Delelis, O., Giron, M.L., de Thé, H., and Saïb, A. (2008). Chromatin tethering of incoming foamy virus by the structural Gag protein. *Traffic* 9, 1717–1727.
 112. Hocum, J.D., Linde, I., Rae, D.T., Collins, C.P., Matern, L.K., and Trobridge, G.D. (2016). Retargeted foamy virus vectors integrate less frequently near proto-oncogenes. *Sci. Rep.* 6, 36610.
 113. Maruggi, G., Porcellini, S., Facchini, G., Perna, S.K., Cattoglio, C., Sartori, D., Ambrosi, A., Schambach, A., Baum, C., Bonini, C., et al. (2009). Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. *Mol. Ther.* 17, 851–856.
 114. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J. Clin. Invest.* 119, 964–975.
 115. Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M.P., et al. (2016). In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. *Cell Stem Cell* 19, 107–119.