BMB
reports

Invited Mini Review

Retroviral integration profiles: their determinants and implications for gene therapy

*Kwang-il Lim**

Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul 140-742 , Korea

Retroviruses have often been used for gene therapy because of their capacity for the long-term expression of transgenes via stable integration into the host genome. However, retroviral integration can also result in the transformation of normal cells into cancer cells, as demonstrated by the incidence of leukemia in a recent retroviral gene therapy trial in Europe. This unfortunate outcome has led to the rapid initiation of studies examining various biological and pathological aspects of retroviral integration. This review summarizes recent findings from these studies, including the global integration patterns of various types of retroviruses, viral and cellular determinants of integration, implications of integration for gene therapy and retrovirus-mediated infectious diseases, and strategies to shift integration to safe host genomic loci. A more comprehensive and mechanistic understanding of retroviral integration processes will eventually make it possible to generate safer retroviral vector platforms in the near future. [BMB reports 2012; 45(4): 207-212]

INTRODUCTION

Retroviruses are RNA viruses that carry a diploid positive-sense RNA genome (1, 2). These viruses are classified into five retroviral genera (the alpha- to epsilon-retroviruses) and two additional genera, the spumaviruses and lentiviruses. Retroviruses cause various fatal human diseases, including acquired immunodeficiency syndrome (AIDS), inflammatory diseases and multiple types of cancer (3-6). Ironically, because of the unique ability of retroviruses to efficiently integrate genetic material within their genome into host chromosomes, these viruses have been used to reverse genetic defects in patients (2). This beneficial use of retroviruses was made possible by an early understanding of retroviral biology and the development of powerful molecular biology techniques (7, 8). After removing genes that are needed for viral replication (for safety reasons), the genomes of retroviruses were further engineered to carry therapeutic transgenes between two long terminal repeats (LTRs) (7-9). These engineered, replication-defective forms of retroviruses, which have been designated retroviral vectors, have been mainly used in making genetic modifications to stem cells for cell-based therapy and in the gene therapy–mediated treatment of various human diseases, including blood disorders, diabetes, and neurological disorders (2, 10-15). Unfortunately, there have been unexpected side effects from retroviral gene therapy that largely stem from where the retroviral genome integrates into the host genome. When retroviral integration occurs near oncogenes, normal cells can be transformed into cancerous cells, as demonstrated by the incidence of leukemia in a gene therapy trial (11). This minireview summarizes the recent findings regarding the integration patterns of various types of retroviruses, the impact of these integration patterns on the initiation and progression of retrovirus-mediated infectious diseases and sporadic virus-initiated cancer, and the viral and host factors that determine these integration patterns. In addition, the ways in which both viral and host components have been engineered to develop safer retroviral vector platforms are briefly discussed.

RETROVIRAL INTEGRATION PREFERENCES FOR GENOMIC REGIONS RICH IN GENE REGULATORY ELEMENTS

Whereas retroviruses were previously thought to randomly integrate into the host genome, many of them display a non-random integration pattern. For example, murine leukemia virus (MLV), a gamma-retrovirus that has been well studied and is often used in gene therapy, shows a strong preference for integrating into genomic regions within or near transcriptional start sites (TSSs) and CpG islands (16-19), which are enriched in gene regulatory elements. MLV demonstrates equivalent levels of integration upstream and downstream of TSSs (16, 18). This integration pattern has been suggested to be closely linked to the incidence of leukemia in young patients who were treated with retrovirally modified CD34+ cells (11). The underlying mechanism by which cells are transformed during retroviral gene therapy is related to the ability of the promoter and enhancer components in the LTR and/or in the regions upstream of the therapeutic genes within the retroviral genome to activate host oncogenes, such as LMO2, when the viruses integrate near the TSSs of these genes (11). Activated oncogenes often turn normal cells into can-

*Corresponding author. Tel: +82-2-2077-7627; Fax: +82-303-0799-0396; E-mail: klim@sm.ac.kr
http://dx.doi.org/10.5483/BMBRep.2012.45.4.207

Received 6 February 2011

Keyword: Host genome, Integration patterns, Retroviral vectors, Safety, Tumorigenicity

cerous cells by perturbing the regulation of cell growth and survival. The retroviral promoter and enhancer components have recently been shown to affect cellular transcription over a long range, independent of vector type, thereby increasing the tumorigenicity of the viral vector (20). The retroviral propensity for frequently targeting oncogenes and cell growth/survival genes and the integration preference for TSSs and CpG islands synergistically increase the malignant potential of retroviral vectors (21).

In addition, MLV considerably prefers transcription factor binding sites (TFBSs) (22) and moderately genes, especially highly expressed genes, during integration (16, 21). Another gamma-retrovirus, porcine endogenous retrovirus (PERV), shows a similar integration pattern, favoring TSSs and CpG islands (23). Xenotropic murine leukemia virus–related virus (XMRV) has an even greater integration preference for TSSs and CpG islands than MLV (24). Relative to other genomic regions, the loci near gene regulatory elements are more deleterious sites of retroviral integration.

LENTIVIRAL INTEGRATION INTO ACTIVELY TRANSCRIBED GENES

HIV-1, a representative lentivirus, strongly prefers integrating into genes instead of genomic regions near TSSs and CpG islands (16-18, 25). In particular, HIV-1 targets actively transcribed genes, and this strong preference for genes has been consistently observed in different cell lines and multiple types of primary cells from different species (26). Furthermore, in both dividing and quiescent cells, HIV-1 demonstrates a strong preference for integration into highly transcribed genes. These findings suggest that the features of cellular physiology do not always affect the overall retroviral integration patterns (27). Whereas the extent of HIV-1's preference for genic regions positively depends upon the transcription level of the genes, interestingly, HIV-1 shows a reduced preference for the most highly expressed genes (21, 26). In contrast to HIV-1, the integration patterns of gamma-retroviruses, such as MLV and PERV, do not depend upon the expression level of the genes (23). HIV-1 normally integrates into entire transcriptional units (TUs) without a significant bias for a certain region of a gene. In addition, HIV-1 targets oncogenes and the genes involved in cell growth and survival (21). In contrast, HIV-1 disfavors CpG islands and does not show any preference for TFBSs (16, 22, 26). Interestingly, integrated lentiviral vectors can both up- and down-regulate host gene expression, in contrast to MLV (20). Another lentivirus, simian immunodeficiency virus (SIV), also shows a similar integration pattern of favoring actively transcribed genes with no preference for TSSs and CpG islands (19, 28, 29). Therefore, retroviruses and lentiviruses in the same genus often share similar integration preferences, as illustrated by MLV and PERV and HIV-1 and SIV (23).

The first molecular-level impact of viral integration on the host is the increased expression of genes that harbor retroviral integrants or are adjacent to retroviral integrants; this increased expression can be up to 3-fold compared with the control (26). Such effects are particularly apparent in HIV-1-infected cells. The site at which retroviruses and lentiviruses integrate into the host genome can affect the initiation and progression of disease. A study using *Cdkn2a*$^{-/-}$ mice, which develop tumors from cancer-triggering genetic changes because of the lack of the Rb1- and p53-dependent tumor-suppressor pathways, demonstrated that retroviral infection of the mice produced tumors following integration at oncogenes and cell cycle–controlling genes in a virus dose–dependent manner (30). In contrast, lentiviral infections did not significantly affect the extent of tumor formation even at high virus doses, indicating that lentiviral vectors may be less tumorigenic than other retroviral vectors (30).

SEMI-RANDOM INTEGRATION OF RETROVIRUSES

In contrast to MLV, PERV, HIV-1 and SIV, which have strong integration preferences for particular genomic regions, avian sarcoma leucosis virus (ASLV), an alpha-retrovirus, has an integration pattern that cannot be clearly distinguished from a hypothetical random pattern (26). ASLV has only a weak preference for actively expressed genes and no significant preference for TSSs. A spumavirus, foamy virus (FV), does not show strong preference toward either TSSs or active genes (31). Similarly, a delta-retrovirus, human T-cell leukemia virus type I (HTLV-I), does not have a strong preference for any host genomic features except alphoid sequences and demonstrates an almost random integration pattern (3). The sequence similarity between the ASLV and HTLV-1 integrases, both of which have seemingly random-like integration patterns, suggests that the viral integrase plays an important role in the selection of sites during integration (32). Retroviruses with random integration patterns can be considered safer platforms for gene delivery vehicles compared with other retroviruses that demonstrate strong integration preferences for TSSs or transcribed genes. Random integration will have a lower chance of disturbing genes that regulate cell growth and survival than integrations targeted to genomic regions enriched in genes or gene regulatory elements.

Whereas integrated retroviral genomes affect the expression of host genes that harbor or are adjacent to the integrants, the host genomic environment surrounding the viral genomes also affects the expression of viral proteins and genomic RNA, ultimately regulating the production of viral progeny. Such mutual genetic effects between proviruses and host cells can lead to complications in the initiation and progression of infectious diseases and sporadic cancer, both of which are caused by retroviral infections. This scenario has been experimentally observed *in vivo* even in infections with retroviruses that display seemingly random integration patterns. For example, HTLV-I infection often results in an asymptomatic carrier disease state that can progress to adult T cell leukemia (ATL) (3, 33). In the carrier state, many infected cells harbor viral genomes at random loci without a strong bias for TSSs and TUs, but only cells in which the viral genome is integrated into the regions not actively transcribed, such as alphoid sequences, can escape from the im-

mune response; cells that produce viral proteins and whole viruses have a greater chance of being identified by the immune system and eliminated (3). Retroviral integration into alphoid repeats appears to result in the formation of an inactive or latent retroviral infection state, as demonstrated by HIV-1 infection. During the carrier stage of the disease, a fraction of infected cells carrying provirus integrated into TSSs but not into alphoid repeats begin to express viral proteins and proliferate, thereby dominating the infected cell population. The proliferative advantage conferred on these cells results in their positive selection, and these cells ultimately can become transformed into leukemia cells (ATL stage) (3). During the transition from the carrier state to the ATL stage of the disease, retroviral integration patterns in the infected cell population appear to shift from a seemingly random pattern with no preference for transcribed regions to a pattern favoring TSSs. Because of this shift, two conflicting integration patterns can be experimentally observed for a given retrovirus. Therefore, characterization of the integration patterns with averaged frequencies for particular genomic regions is not sufficient to confirm the safety of therapeutically used retroviruses.

VIRAL AND HOST FACTORS INVOLVED IN DETERMINING RETROVIRAL INTEGRATION PATTERNS

Few studies have mechanistically and comprehensively deciphered how retroviruses select their preferred genomic regions for integration. Researchers have identified only a fraction of the viral and cellular factors that play a role in determining particular retroviral integration patterns. A previous study showed that when the HIV-1 integrase or the HIV-1 Gag sequence is replaced with the corresponding MLV sequence, the resultant chimeric HIV-1 has an integration pattern similar to that of MLV or a new pattern that is different from that of HIV-1 or MLV, respectively; the introduction of sequences of the two MLV proteins results in an HIV-1 integration pattern closer to that of MLV (34). These experimental observations suggest that the integrase is one of the main components that determine the integration pattern for each retrovirus and that Gag protein, a structural polyprotein, is also involved as a cofactor in shaping the integration pattern. Several other studies have indicated that the central core of the retroviral integrase, especially the α2 helix in the core, plays a key role in generating distinct integration patterns (35). Recently, a structurally resolved molecular mechanism of the retroviral integration reaction has been proposed (36), but it is not sufficient to explain the host genomic site preference of retroviruses in detail.

Although there is no concrete or complete model to explain retrovirus-specific variations in integration preference, several plausible mechanisms have been suggested. Retroviruses can associate with their preferred genomic loci by interacting with cellular proteins that are localized at the genomic sites. For example, MLV and HIV-1 integrate into the host genome by interacting with host proteins that bind to genomic regions rich in either TSSs and CpG islands or genes. In contrast, ASLV, which demonstrates an almost random integration pattern, would interact with proteins that are distributed across the host genome but do not have a significant affinity for TSSs, CpG islands or genes (26).

Observations based on yeast two-hybrid systems have identified potential host proteins that may interact with viral proteins in the retroviral preintegration complex (PIC) (37). These host proteins in MLV-infected cells include chromatin-binding factors and transcription factors (TFs). More host proteins have been identified in HIV-1-infected cells, including Ini1 and human lens epithelium–derived growth factor (LEDGF/p75) (37). HIV-1 has been shown to integrate into genes that are responsive to LEDGF/p75 in different cell types. Interestingly, depletion of LEDGF/p75 in host cells leads to significantly reduced HIV-1 integration into TUs, but the remaining integration bias for TUs suggests that LEDGF/p75 is not the only factor that tethers HIV-1 PICs to TUs (38). In the absence of LEDGF/p75, HIV-1 integration occurs more frequently in promoter regions and CpG islands, demonstrating that this host protein also causes the virus to avoid such genomic regions (39). The integration defect of HIV-1 arising from the absence of LEDGF/p75 suggests that the protein is a necessary factor for HIV-1 integration under some conditions. In addition, the knockdown of transportin-3 and the nuclear pore protein RanBP2 affects HIV-1 integration site selection, shifting the integration site away from gene-dense or gene-associated regions. However, this newly shifted integration pattern is different from the pattern observed in the absence of LEDGF/p75 (40). Research groups have used genome-wide siRNA analyses to more systematically identify host factors involved in HIV-1 genome integration (41). These researchers have proposed that multiple host proteins, including KPNB1, NUP98, ANAPC2, PRPF38A, SNW1 and AQR, which have functions in nuclear import, nuclear pore formation/structure, nucleic acid binding, transcriptional regulation, and DNA repair, may function as integration coupling or tethering factors during HIV-1 integration. The observed participation of nuclear pore proteins and Gag protein-mediated trafficking of the HIV-1 PIC through nuclear pores to the host genome suggest a new potential model in which retroviruses use the cellular routes connecting the nuclear pore to the favored genomic loci during integration (40). A comprehensive and complete catalog of host proteins interacting with the viral integration machinery is still not available, and further experiments need to be performed to confirm the effects of host proteins on retroviral integration patterns.

The sequence of retroviral genomes may also affect the integration pattern via a mechanism in which promoter and enhancer elements in the viral genome interact with certain TFs or other transcriptional components that have an affinity for particular genomic loci. For example, the interactions of MLV enhancer elements with TFs lead to integration into cell-type specific TFBSs in the host genome (22). However, there are also experimental data that conflict with this proposed mechanism for certain types of retroviruses; for example, the deletion of promoter and enhancer elements in the SIV genome does not change the viral integration preference (29). Therefore, the pro-

posed integration mechanism involving gene regulatory elements or sequences in the retroviral genome is not universally applied. In contrast, there have been no noticeable effects of the host genomic DNA sequence on retroviral integration site selection (27). Instead, retroviral integration patterns are dependent upon the transcription level of genes when viruses target TUs (28). Because many retroviruses prefer actively transcribed regions and each cell type has a different transcriptional profile, there are cell-type dependent variations in the integration frequency in each gene (26). However, the main qualitative trend in which each type of retrovirus favors particular genomic features is mostly maintained irrespective of cell type. Different types of retroviruses have also shown common dependence on other cellular factors. For example, the loci in which MLV, HIV-1 and ASLV integrate exhibit a positive correlation with gene density, although ASLV has the weakest level of dependence (26). Physical aspects of host genomic DNA are also important in the retroviral selection of integration loci, as demonstrated by the observations that retroviruses prefer the outward-facing side of chromatin and favor nucleosomal DNA over naked DNA (42).

MOLECULAR ENGINEERING OF RETROVIRAL INTEGRATION PATTERNS

To alter retroviral integration patterns with tumorigenic potential, retroviral integrases have been targeted for engineering. Several studies have used polydactyl zinc finger proteins that have specificity for particular genomic loci in the host cell (43). The DNA-binding motifs, as tethering molecules for retroviral PICs, are inserted into the N- or the C-terminus of integrase, but these fusion proteins only allow a low level of retroviral integration specificity for the intended target site in the host genome (43). This limited outcome may be due to the remaining functional motifs in the intact integrase. In addition, other DNA-binding domains in the LexA repressor of *Escherichia coli* and the λ repressor have also been used, but the intended outcome was not obtained at a satisfactory level (44-46). We recently constructed a retroviral library that was screened in a high-throughput format to identify novel fusion proteins where zinc finger complexes (ZFCs), composed of six finger subunits, were inserted within MLV Gag-Pol proteins, with the goal of enhanced safety and higher integration specificity of retroviruses *in vivo* (17). Several fusion proteins successfully shifted the characteristic retroviral integration pattern into new genomic regions that were distant from TSSs and CpG islands. Furthermore, when such ZFCs were inserted into different regions of Gag-Pol, integrations occurred at common genomic sites with a high frequency, indicating that the inserted ZFCs provided viral integration specificity. However, the found common integration sites were not the predicted sites given the ZFCs, suggesting that the protein context of the inserted ZFCs and the chromatin structure in the targeted genomic regions may affect the conformation and specificity of the zinc finger units (17). Alternatively, integration-defective retroviruses have been generated via site-specific mutagenesis of integrase to develop safer vector systems (47, 48). Depending upon the host cell type, the mutant viruses allow transient or long-term expression of transgenes without a significant level of integration. However, the integration ability of retroviruses is not completely abolished, which remains a safety concern. The limited successes of the above-mentioned approaches are mainly the result of an incomplete understanding of how retroviral integration preferences are determined and controlled *in vivo*.

CONCLUSION

As summarized above and in Table 1, most retroviruses prefer integrating into particular host genomic regions, thereby revealing different levels of oncogenic potential when these retroviruses are used in gene therapy. Such differences in integration patterns might arise from different evolutionary strategies of retroviruses; MLV maximizes progeny production during infection, whereas HIV-1 maximizes long-term coexistence with the host by minimizing perturbation of host gene regulation (49). Viruses do not generally evolve to become beneficial to humans. Therefore, to maximally use the retroviral ability of prolonged gene expression in gene therapy without detrimental side effects, these viruses need to be carefully engineered. Recent and near-future advances in the understanding of retroviral integration, in terms of virology, genetics, and molecular and cell biology, will make it possible to develop efficient and safe retroviral gene delivery platforms. Such advances will also lead to better elucidation of the initiation and

Table 1. Summary of the integration patterns of retroviruses

Retrovirus	Genus	Favored genomic regions	Disfavored or not favored genomic region
MLV	Gamma-retrovirus	TSSs[16-19, 26, 49] CpG islands[16-18, 49]	
HIV-1	Lentivirus	Genes[16, 18, 25-27] TSSs[16 (weak)]	CpG islands[18, 26 (dis)]
ASLV	Alpha-retrovirus	Genes[26 (weak)]	TSSs[26 (not)]
SIV	Lentivirus	Genes[19, 28, 29]	TSSs[28, 29 (not)] CpG islands[28 (not), 29 (dis)]
PERV	Gamma-retrovirus	TSSs[23] CpG islands[23]	
Foamy virus	Spumavirus	CpG islands[31 (weak)]	TSSs[31 (not)] Genes[31 (not)]
XMRV	Gamma-retrovirus	TSSs[24] CpG islands[24]	

Only the genomic regions within (or near) TSSs, CpG islands and Genes are denoted in this table. Superscript numbers indicate the relevant references. No superscript: strong integration preference. Weak: weak integration preference, Not: not favored for integration. dis: disfavored for integration

progression of various retrovirus-mediated infectious diseases.

REFERENCES

1. Coffin, J. M., Hughes, S. H. and Varmus, H. E. (1997) Retroviruses, Cold Spring Harbor Lab. Press, New York, U.S.A.
2. Knipe, D. M. and Howley, P. M. (2006) Fields Virology, 5th ed., Lippincott Williams & Wilkins, Philadelphia, USA.
3. Doi, K., Wu, X., Taniguchi, Y., Yasunaga, J., Satou, Y., Okayama, A., Nosaka, K. and Matsuoka, M. (2005) Preferential selection of human T-cell leukemia virus type I provirus integration sites in leukemic versus carrier states. *Blood* **106**, 1048-1053.
4. Guss, D. A. (1994) The acquired immune deficiency syndrome: an overview for the emergency physician, Part 2. *J. Emerg. Med.* **12**, 491-497.
5. Guss, D. A. (1994) The acquired immune deficiency syndrome: an overview for the emergency physician, Part 1. *J. Emerg. Med.* **12**, 375-384.
6. Fan, H. (2007) A new human retrovirus associated with prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1449-1450.
7. Shimotohno, K. and Temin, H. M. (1982) Loss of intervening sequences in genomic mouse alpha-globin DNA inserted in an infectious retrovirus vector. *Nature* **299**, 265-268.
8. Watanabe, S. and Temin, H. M. (1983) Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. *Mol. Cell. Biol.* **3**, 2241-2249.
9. Schaffer, D. V., Koerber, J. T. and Lim, K. I. (2008) Molecular engineering of viral gene delivery vehicles. *Annu. Rev. Biomed. Eng.* **10**, 169-194.
10. Schambach, A. and Baum, C. (2008) Clinical application of lentiviral vectors - concepts and practice. *Curr. Gene Ther.* **8**, 474-482.
11. Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C. S., Pawliuk, R., Morillon, E., Sorensen, R., Forster, A., Fraser, P., Cohen, J. I., de Saint Basile, G., Alexander, I., Wintergerst, U., Frebourg, T., Aurias, A., Stoppa-Lyonnet, D., Romana, S., Radford-Weiss, I., Gross, F., Valensi, F., Delabesse, E., Macintyre, E., Sigaux, F., Soulier, J., Leiva, L. E., Wissler, M., Prinz, C., Rabbitts, T. H., Le Deist, F., Fischer, A. and Cavazzana-Calvo, M. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419.
12. Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., Cavallesco, R., Gillet-Legrand, B., Caccavelli, L., Sgarra, R., Maouche-Chretien, L., Bernaudin, F., Girot, R., Dorazio, R., Mulder, G. J., Polack, A., Bank, A., Soulier, J., Larghero, J., Kabbara, N., Dalle, B., Gourmel, B., Socie, G., Chretien, S., Cartier, N., Aubourg, P., Fischer, A., Cornetta, K., Galacteros, F., Beuzard, Y., Gluckman, E., Bushman, F., Hacein-Bey-Abina, S. and Leboulch, P. (2010) Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**, 318-322.
13. Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., Marinello, E., Cattaneo, F., Vai, S., Servida, P., Miniero, R., Roncarolo, M. G. and Bordignon, C. (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* **296**, 2410-2413.
14. Ott, M. G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H., Naundorf, S., Brinkmann, A., Deichmann, A., Fischer, M., Ball, C., Pilz, I., Dunbar, C., Du, Y., Jenkins, N. A., Copeland, N. G., Luthi, U., Hassan, M., Thrasher, A. J., Hoelzer, D., von Kalle, C., Seger, R. and Grez, M. (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* **12**, 401-409.
15. Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.
16. 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.
17. Lim, K. I., Klimczak, R., Yu, J. H. and Schaffer, D. V. (2010) Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12475-12480.
18. Tsukahara, T., Agawa, H., Matsumoto, S., Matsuda, M., Ueno, S., Yamashita, Y., Yamada, K., Tanaka, N., Kojima, K. and Takeshita, T. (2006) Murine leukemia virus vector integration favors promoter regions and regional hot spots in a human T-cell line. *Biochem. Biophys. Res. Commun.* **345**, 1099-1107.
19. Hematti, P., Hong, B. K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I. E., Eckfeldt, C. E., Sharma, Y., Schmidt, M., von Kalle, C., Persons, D. A., Billings, E. M., Verfaillie, C. M., Nienhuis, A. W., Wolfsberg, T. G., Dunbar, C. E. and Calmels, B. (2004) Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* **2**, e423.
20. Maruggi, G., Porcellini, S., Facchini, G., Perna, S. K., Cattoglio, C., Sartori, D., Ambrosi, A., Schambach, A., Baum, C., Bonini, C., Bovolenta, C., Mavilio, F. and Recchia, A. (2009) Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. *Mol. Ther.* **17**, 851-856.
21. Beard, B. C., Dickerson, D., Beebe, K., Gooch, C., Fletcher, J., Okbinoglu, T., Miller, D. G., Jacobs, M. A., Kaul, R., Kiem, H. P. and Trobridge, G. D. (2007) Comparison of HIV-derived lentiviral and MLV-based gammaretroviral vector integration sites in primate repopulating cells. *Mol. Ther.* **15**, 1356-1365.
22. 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.
23. Moalic, Y., Felix, H., Takeuchi, Y., Jestin, A. and Blanchard, Y. (2009) Genome areas with high gene density and CpG island neighborhood strongly attract porcine endogenous retrovirus for integration and favor the formation of hot spots. *J. Virol.* **83**, 1920-1929.
24. Kim, S., Kim, N., Dong, B., Boren, D., Lee, S. A., Das Gupta, J., Gaughan, C., Klein, E. A., Lee, C., Silverman, R. H. and Chow, S. A. (2008) Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. *J. Virol.* **82**, 9964-9977.
25. Schroder, 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.
26. 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.
27. Ciuffi, A., Mitchell, R. S., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J. R. and Bushman, F. D. (2006) Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. *Mol. Ther.* **13**, 366-373.
28. Crise, B., Li, Y., Yuan, C., Morcock, D. R., Whitby, D., Munroe, D. J., Arthur, L. O. and Wu, X. (2005) Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* **79**, 12199-12204.
29. Monse, H., Laufs, S., Kuate, S., Zeller, W. J., Fruehauf, S. and Uberla, K. (2006) Viral determinants of integration site preferences of simian immunodeficiency virus-based vectors. *J. Virol.* **80**, 8145-8150.
30. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C., Doglioni, C., von Kalle, C. and Naldini, L. (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.* **24**, 687-696.
31. 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.
32. Meekings, K. N., Leipzig, J., Bushman, F. D., Taylor, G. P. and Bangham, C. R. (2008) HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog.* **4**, e1000027.
33. 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.
34. Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C. C., Ecker, J. R. and Bushman, F. D. (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**, e60.
35. Appa, R. S., Shin, C. G., Lee, P. and Chow, S. A. (2001) Role of the nonspecific DNA-binding region and alpha helices within the core domain of retroviral integrase in selecting target DNA sites for integration. *J. Biol. Chem.* **276**, 45848-45855.
36. Maertens, G. N., Hare, S. and Cherepanov, P. (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326-329.
37. 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.
38. 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.
39. 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.
40. 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., Konig, R., Malani, N., Berry, C. C. and Bushman, F. D. (2011) HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog.* **7**, e1001313.
41. Konig, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M., Irelan, J. T., Chiang, C. Y., Tu, B. P., De Jesus, P. D., Lilley, C. E., Seidel, S., Opaluch, A. M., Caldwell, J. S., Weitzman, M. D., Kuhen, K. L., Bandyopadhyay, S., Ideker, T., Orth, A. P., Miraglia, L. J., Bushman, F. D., Young, J. A. and Chanda, S. K. (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49-60.
42. 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.
43. Tan, W., Dong, Z., Wilkinson, T. A., Barbas, C. F.,3rd and Chow, S. A. (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J. Virol.* **80,** 1939-1948.
44. Katz, R. A., Merkel, G. and Skalka, A. M. (1996) Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. *Virology.* **217,** 178-190.
45. Goulaouic, H. and Chow, S. A. (1996) Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. *J. Virol.* **70,** 37-46.
46. Bushman, F. D. (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9233-9237.
47. Philippe, S., Sarkis, C., Barkats, M., Mammeri, H., Ladroue, C., Petit, C., Mallet, J. and Serguera, C. (2006) Lentiviral vectors with a defective integrase allow efficient and sustained transgene expression in vitro and *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 17684-17689.
48. Yanez-Munoz, R. J., Balaggan, K. S., MacNeil, A., Howe, S. J., Schmidt, M., Smith, A. J., Buch, P., MacLaren, R. E., Anderson, P. N., Barker, S. E., Duran, Y., Bartholomae, C., von Kalle, C., Heckenlively, J. R., Kinnon, C., Ali, R. R. and Thrasher, A. J. (2006) Effective gene therapy with nonintegrating lentiviral vectors. *Nat. Med.* **12,** 348-353.
49. Cattoglio, C., Pellin, D., Rizzi, E., Maruggi, G., Corti, G., Miselli, F., Sartori, D., Guffanti, A., Di Serio, C., Ambrosi, A., De Bellis, G. and Mavilio, F. (2010) High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* **116,** 5507-5517.