Article

Characterization of Equine Infectious Anemia Virus Integration in the Horse Genome

Qiang Liu, Xue-Feng Wang, Jian Ma, Xi-Jun He, Xiao-Jun Wang * and Jian-Hua Zhou *

State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute of Chinese Academy of Agricultural Sciences, Harbin 150001, China; E-Mails: liuqiang_yyy@163.com (Q.L.); xuefengwang1982@126.com (X.-F.W.); jma@hvri.ac.cn (J.M.); hexijun@caas.cn (X.-J.H.)

* Authors to whom correspondence should be addressed; E-Mails: xjw@hvri.ac.cn (X.-J.W.); jianhua_uc@126.com (J.-H.Z.); Tel.: +86-189-4606-62855 (X.-J.W.); +86-137-9608-5512 (J.-H.Z.); Fax: +86-451-5199-7166 (X.-J.W.).

Abstract: Human immunodeficiency virus (HIV)-1 has a unique integration profile in the human genome relative to murine and avian retroviruses. Equine infectious anemia virus (EIAV) is another well-studied lentivirus that can also be used as a promising retro-transfection vector, but its integration into its native host has not been characterized. In this study, we mapped 477 integration sites of the EIAV strain $EIAV_{FDDV13}$ in fetal equine dermal (FED) cells during *in vitro* infection. Published integration sites of EIAV and HIV-1 in the human genome were also analyzed as references. Our results demonstrated that $EIAV_{FDDV13}$ tended to integrate into genes and AT-rich regions, and it avoided integrating into transcription start sites (TSS), which is consistent with EIAV and HIV-1 integration in the human genome. Notably, the integration of $EIAV_{FDDV13}$ favored long interspersed elements (LINEs) and DNA transposons in the horse genome, whereas the integration of HIV-1 favored short interspersed elements (SINEs) in the human genome. The chromosomal environment near LINEs or DNA transposons potentially influences viral transcription and may be related to the unique EIAV latency states in equids. The data on EIAV integration in its natural host will facilitate studies on lentiviral infection and lentivirus-based therapeutic vectors.

Keywords: equine infectious anemia virus; integration sites; chromosomes; RefSeq genes; repetitive elements

1. Introduction

Integration, *i.e.*, the incorporation of viral cDNA into a host cell genome, is a crucial step in the retrovirus life cycle that is mediated by preintegration complexes (PICs), which are composed of viral integrase, cellular proteins and other viral proteins [1–3]. As a result, retroviruses, in the form of proviruses, become an integral part of the host's chromosomes, persist for the duration of the host's lifetime and establish latent infections in the host cell. The irreversible integration of retroviruses makes them promising therapeutic vectors for human genetic diseases. Retrovirus-based vectors have been successfully used in human gene therapy to correct primary immunodeficiency and genetic deficiency [4–8]. However, adverse events, such as disruption of genes and activation of the proto-oncogenes, have raised awareness of the safety implications of retroviral-based vectors used in clinical gene therapy [9–12]. Not all retroviral-based vectors have the same genotoxic potential. For example, gammaretroviruses tend to cause disruption of genes more than lentiviruses [13–18]. Therefore, studies on the selection of integration sites for retroviruses in host genomes, especially for gene therapy vectors, have particular significance.

The complete human genome is available online, and data on the integration sites of some important retroviruses have been surveyed, which have shown that selection of the target site is not random and that different retroviruses have different preferences in the human genome. For example, the lentiviruses in the human genome, e.g., human immunodeficiency virus type 1 (HIV-1) [19], simian immunodeficiency virus (SIV) [20], and feline immunodeficiency virus (FIV) [21], are prone to insert into transcription units, particularly in actively transcribed host genes. Furthermore, murine leukemia virus (MLV) tends to integrate near transcription start sites (TSSs) or CpG islands [22]. In contrast, target site selection for avian sarcoma-leukosis virus (ASLV) has been shown to be random, and ASLV displays only a slight preference for transcription units during its infection of human-derived cell lines [23,24].

Equine infectious anemia virus (EIAV) is an important etiological agent to equids and has served as an animal model for HIV-1/AIDS research [25,26]. Most infected horses eventually become asymptomatic carriers of the virus after several months of the initial infection with a low plasma viral load and viral replication occurring in tissues rich in mononuclear cells [27]. However, the viral replication and clinical symptoms can be reactivated by natural or experimental suppression of the horse's immunity [28,29]. It is known that the chromosomal environment around the integration sites influences the replication and gene expression of the integrated provirus [30–32]. What, therefore, are the integration site characteristics of EIAV in the genome, and particularly in the horse genome? Hacker *et al*. [33] published a study in 2006 on EIAV integration in the human genome in HEK293T cells using a vesicular stomatitis virus G protein (VSV-G)-pseudotyped, EIAV-based transfection vector that was prepared using a three-plasmid system. In comparison with an HIV-1-based vector control, EIAV exhibited a preference similar to that of HIV-1 integration in the human genome. Both of these lentiviral vectors tended to insert within genes, especially actively transcribed genes, and they also favored AT-rich regions [33,34]. However, to the best of our knowledge, there is no published study on the integration of EIAV or other retroviruses in the chromosomes of equids.

EIAV$_{FDDV13}$ is an EIAV strain that is adapted to cultivated equine and donkey fibroblasts (fetal dermal cells), which are among the few EIAV-permissive equine cells and are widely used for EIAV infection and replication *in vitro*. This virus's genomic, virological and immunological characteristics have been

extensively studied [35–37]. With the availability of the complete horse genome sequence [38,39], it is possible to map EIAV integration sites in horse chromosomes. In this study, fetal equine dermal (FED) cells were infected with EIAV$_{FDDV13}$, and 477 EIAV integration sites were cloned, sequenced, and mapped to the horse genome. In addition, because the human genome database has been updated and additional information on EIAV integration in the human genome also needs to be analyzed, the published data on the integration sites of EIAV and HIV-1 in the human genome were also downloaded and processed using our analysis pipeline in this study for reference to our data. This study, which evaluates the features of EIAV integration in the horse genome, provides additional knowledge on lentiviral integration.

2. Materials and Methods

2.1. Cell Preparation and Infection

FED cells were cultured and infected with EIAV as described previously for fetal donkey dermal (FDD) cells [40]. Briefly, FED cells were plated in 75 cm^2 flasks (Corning, Corning, NY, USA) for 2 days until they reached approximately 90% confluency. EIAV$_{FDDV13}$ was inoculated at a multiplicity of infection (MOI) of 10 for 1 h to allow the adsorption and infection of the virus. The cells were then washed twice with phosphate buffered saline (PBS) and once with fresh α-minimal essential medium (α-MEM) (Life Technologies, Carlsbad, CA, USA) containing 10% fetal bovine serum (FBS) (Sigma, St Louis, MO, USA), and then they were re-cultured in the same medium at 37 °C under 5% CO$_2$. It is known that lentiviral cDNA enters the nucleus within 24 h post-infection, and viral particles are produced and released within 48 h post-infection [41,42]. To maximize provirus formation but minimize clonal expansion during cell growth after integration [43], the FED cells were harvested 24 h post-infection. The harvested cells were stored at -80 °C until further use.

2.2. Cloning and Sequencing of Integration Sites

Cellular genomic DNA (gDNA) was extracted using a QIAamp DNA mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. A library of integration site junctions containing 371 bp of the 5′ end of the integrated EIAV long terminal repeat (LTR) and the up-stream horse gDNA was created by ligation-mediated PCR (LM-PCR) performed as previously described by Ciuffi *et al.* [43,44]. Briefly, to remove unintegrated (circular and linear forms) viral DNA, 2 mg of gDNA was electrophoresed on a 0.6% agarose gel on ice. DNA bands longer than 8.3 kb were extracted from the gel, purified using an E.Z.N.A. TM Ploy Gel DNA Extraction Kit (Omega, Stamford, CT, USA) and digested overnight with the restriction endonucleases *Ssp* I or *Dra* I to fragment the long gDNA. The digested DNA fragments were blunt-ligated to a double-stranded adapter, which was prepared by annealing from Link-1 and Link-2.

The ligation products were amplified by a two-step PCR. The first PCR amplification was performed with an adaptor-specific primer, APF1, and a viral specific primer, EIAV514. The PCR reaction conditions were as follows: 5 min of pre-denaturation at 98 °C, 7 cycles of denaturation at 98 °C for 25 s and an extension step at 72 °C for 3 min, followed by 32 cycles of denaturation at 98 °C for 25 s and extension at 68 °C for 3 min. A final extension was performed at 72 °C for 10 min. The PCR

products were then diluted using TE buffer (Invitrogen, Carlsbad, CA, USA) at a ratio of 1:50 (*v/v*), and 1 μL of diluted amplified DNA was used as the template for the secondary PCR reaction. The reaction was performed with an adaptor-specific primer, APF2, and a viral-specific primer, EIAV359. The conditions of the reaction were identical to those of the first-round PCR except that the number of amplification cycles was reduced (20 cycles). The PCR products were visualized using 2% agarose gel electrophoresis, and the diffuse bands longer than 371 bp were extracted and purified using a gel extraction kit (Qiagen, Germany) and then cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA). All clones were cultured and sequenced.

2.3. Sequences Analysis and Mapping of Integration Sites in the Host Genome

For reference to EIAV integration in the horse genome, 562 human genome sequences flanking the 3′ LTR for the retrovirus, which included 146 integration sites for an HIV-1 transfection vector (DQ498202 to DQ498347) and 416 sites for an EIAV vector (DQ498348 to DQ498763), were downloaded from GenBank and processed using our analysis pipeline in this study (Table 1). These vectors were produced using a three-plasmid lentiviral vector system, which included an EIAV-based transduction vector and vectors encoding the EIAV gag/pol and the VSV-G envelope. These vectors were then co-transfected into 293T cells as described in a previous report [33]. The number of sequences in GenBank was less than that published by Hacker *et al*. [33], which had been analyzed and mapped to the human genome (University of California, Santa Cruz (UCSC), assembled in May 2004). In this study, the human genome database, which has been recently updated, was used to analyze and map these sequences (UCSC, assembled in December 2013).

Table 1. Datasets for the integration sites used in this study.

Group	Virus or Vector	Cell Type	Number of Integration Sites	Accession Numbers (GenBank)	Source of Reference
HIV-1	HIV-1 vector	HEK 293T [a]	146 [b] (162 [c])	DQ498202–DQ498347	[33]
EIAV	EIAV vector	HEK 293T	416 (458)	DQ498348–DQ498763	[33]
EIAV-*Ssp* I	EIAV$_{FDDV13}$ [d]	FED [e]	287	KO454413–KO454699	This study
EIAV-*Dra* I	EIAV$_{FDDV13}$	FED	190	KO454223–KO454412	This study

[a]: Human embryonic kidney 293T cells; [b]: Actual number of sequences downloaded from the NCBI database provided by Hacker *et al.* [33]; [c]: Number of sequences reported by Hacker *et al.* [33]; [d]: EIAV fetal equine dermal cell-adapted strain; [e]: Fetal equine dermal cell.

The Blast-like Alignment Tool (BLAT) program (Available online: http://genome.ucsc.edu) was used to analyze the sequences, which were aligned and mapped to the horse genome (UCSC, assembled in September 2007) (Available online: http://genome.ucsc.edu/). A site was deemed to be an integration site for EIAV if it complied with the following criteria: (1) must be located between the adaptor sequence and the EIAV LTR; (2) must have at least 95% homology to the horse genome sequence and should be a single horse genetic locus; (3) the junction must consist of a horse genomic sequence and 371 bp of the 5′ terminal of the EIAV LTR, within which "TGTGGG" must be used as the initial sequence based on the sequence of EIAV$_{FDDV13}$; and (4) must have a minimum length of 20 bp, which was the lowest

limit recognized by the BLAT program. The sequences that were used for reference were aligned and mapped to the recently updated human genome (UCSC, assembled in December 2013) (Available online: http://genome.ucsc.edu/).

The BioMart program (Ensemble Genes 79, Available online: http://asia.ensembl.org) was used to determine whether the integration sites were located in coding genes of the September 2007 horse genome draft. Additional information regarding transcription initiation and termination sites of the coding genes was also obtained from the BioMart program. The reference mRNA sequence (RefSeq mRNA) of the September 2007 horse genome draft was downloaded from the UCSC genome annotation database (Available online: http://www.genome.ucse.edu). The base frequency around the integration sites was determined using the WebLOGO program (Available online: http://weblogo.berkeley.edu/). The repetitive elements around the integration sites were determined by the RepeatMasker analysis of the September 2007 horse genome draft (Available online: http://www.repeatmasker.org/).

2.4. Statistical Analysis

To determine integration site selection bias *in vitro*, a set of 10,000 random integration sites was generated using the Microsoft Excel "RAND()" function by choosing random numbers between 1 and 2,367,055,132, which represents the total length of the 31 autosomal chromosomes plus the X sex chromosome in the horse genome (UCSC, assembled in September 2007) as previously described for the generation of a random dataset [33]. All statistical analyses were performed using the statistics analysis system (SAS) 9.2. The one-sided Fisher exact or chi-squared test was used to determine whether the differences in chromosome distribution were significant and to determine the integration frequencies into gene coding regions and repetitive elements. $p \leq 0.05$ and $p \leq 0.01$ were considered significant or very significant, respectively, for all tests performed.

2.5. Nucleotide Sequence Accession Numbers

The integration site sequences in this study were uploaded to GenBank. The accession numbers range from KO454223 to KO454699.

3. Results

3.1. Distribution of EIAV Provirus Integration Sites in Horse Chromosomes

To determine the horse genomic sequences flanking the EIAV provirus integration sites, genomic (g)DNA from FED cells infected with EIAV$_{\text{FDDV13}}$ was isolated, digested with restriction endonucleases, and ligated to a DNA adaptor. Integration sites were amplified using the 5′ primer that bound the adaptor and the 3′ primer that bound to the provirus long terminal repeat (LTR), and then the PCR products were cloned and sequenced. To reduce the integration bias introduced by usage of restriction endonucleases in chromosomes, two different restriction endonucleases, *Ssp* I and *Dra* I, were used to single-digest the horse gDNA. Based on the criteria for authentic integration sites described in the Materials and Methods, we obtained a total of 477 horse genomic sequences flanking the EIAV integration sites, which included 287 sequences obtained from the *Ssp* I-digested (*Ssp* I group) and 190 sequences from the *Dra* I-digested (*Dra* I group) horse gDNA. These sequences were analyzed and mapped to the horse genome (UCSC,

assembled in September 2007) (Table 1). In addition, as references for the analysis of EIAV integration in the horse genome, 562 human genomic sequences flanking the lentivirus integration sites, including 146 sequences for HIV-1 and 416 sequences for EIAV, were downloaded and mapped to an updated human genomic database (UCSC, assembled in December 2013) (Table 1).

Mapping of both the *Ssp* I and *Dra* I groups showed that integration events occurred in all 31 autosomal chromosomes plus the X sex chromosome in the horse genome. No significant differences were observed between the *Ssp* I and *Dra* I groups with respect to the integration frequency in each chromosomes ($p > 0.05$) (Figure 1). Furthermore, although variations in insertion events and chromosome sizes appeared in the EIAV-integrated horse genome, no evidence of significant differences in the distribution of integration sites was found among chromosomes ($p > 0.05$) (Figure 1). Furthermore, the human genomic sequences flanking HIV-1 and EIAV integration sites in human chromosomes were also analyzed to determine their chromosomal distribution. Integration occurred in all 22 autosomal chromosomes plus the X sex chromosome in the human genome (except the eighteenth chromosome for EIAV). The integration frequencies of HIV-1 in chromosomes 3, 4 and 6 were higher than those of EIAV ($p < 0.05$). In contrast, the integration frequencies of EIAV in chromosomes 15–17 and 19 were higher than those of HIV-1 ($p < 0.05$) (Figure S1).

3.2. Specificity of the 40 bp Region around the Integration Sites

After retroviral integration into host chromosomes, the primary sequence is found to play a minor role in the selection of the target sites, but a weak palindromic sequence is often observed at the insertion sites of retroviruses, including HIV-1, MLV and ASLV [45,46]. In this study, the 40 bp sequence around the integration sites was analyzed. We found that as for other lentiviruses, AT-rich regions were characteristic of the target site selection (Figure 2A,B). In addition, in both the *Ssp* I and *Dra* I groups, a palindromic sequence centered around the insertion site (centered between the site −1 of the palindromic sequence for the *Ssp* I group and +1 for the *Dra* I group) was observed in the horse genome (Figure 2A,B). However, these characteristics were not observed in the corresponding regions of EIAV and HIV in the human genome. In contrast, a palindromic structure centered at +3 between the −1 and +6 position from the insertion sites was observed at the integration sites of HIV-1 and EIAV in human chromosomes (Figure S2A,B). Our study found that EIAV integration sites, particularly in the *Dra* I group, also had a weak palindromic sequence centered around +3 in the horse chromosomes (Figure 2A,B).
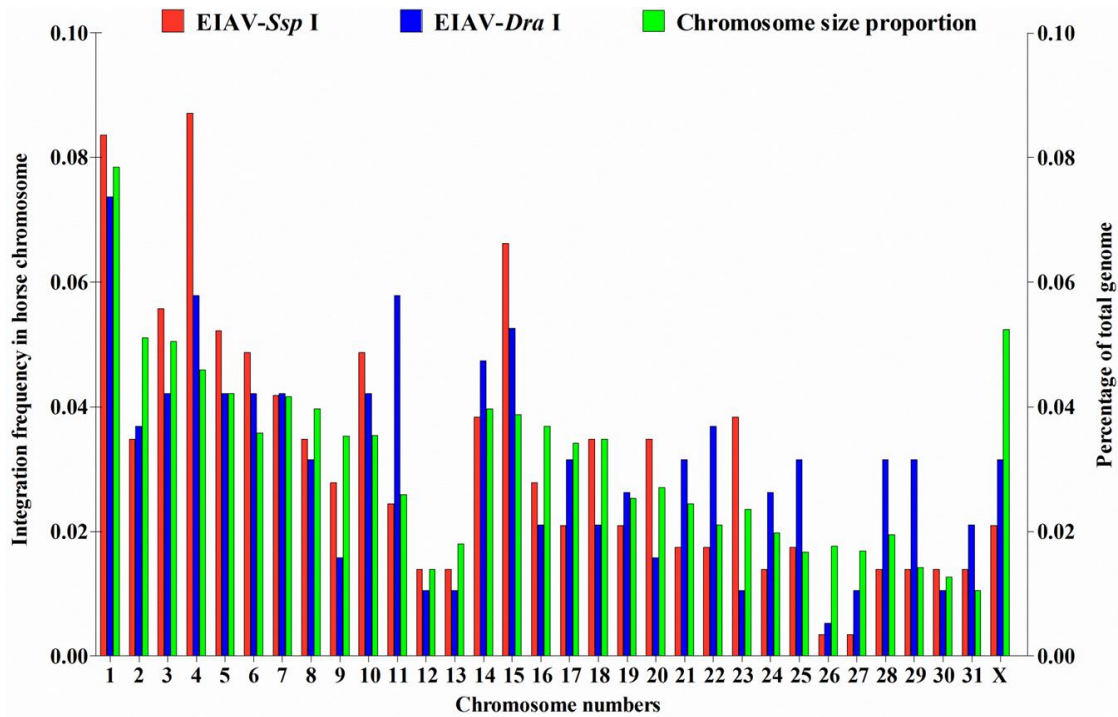
Figure 1. Chromosome distribution of EIAV$_{FDDV13}$ integration sites in horse chromosomes. Chromosomal DNA of FED cells infected with EIAV$_{FDDV13}$ was digested with the restriction endonuclease *Ssp* I or *Dra* I. Genomic fragments with EIAV integration sites were subcloned after being amplified by LM-PCR. The proportion of integration sites in horse chromosomes digested with *Ssp* I (287 sites) or *Dra* I (190 sites) is indicated as a percentage, and the percentage of each corresponding chromosome size was calculated based on the length of the entire horse genome. The chi-squared test was used.
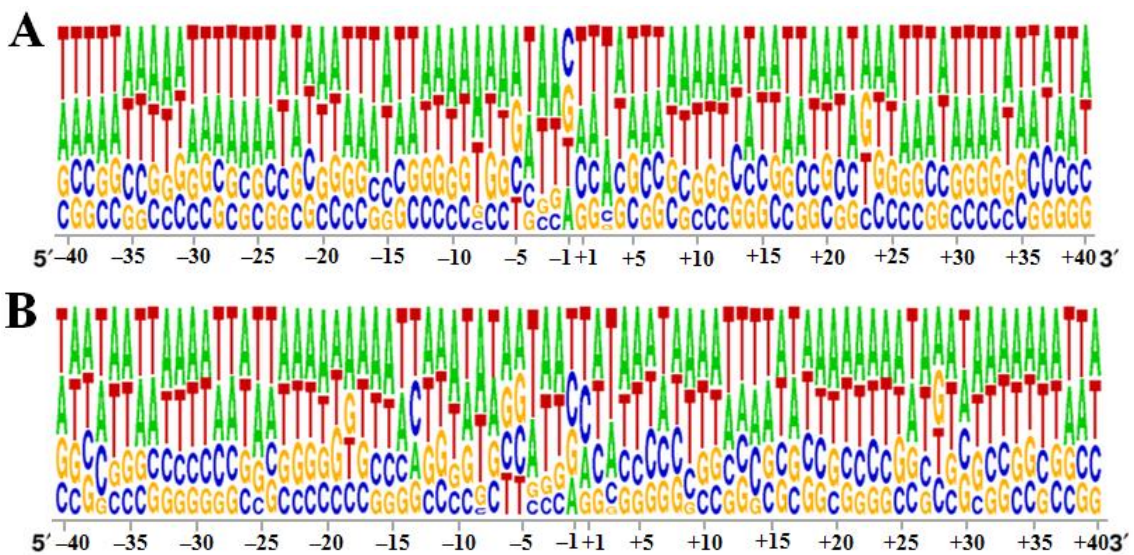

Figure 2. Base frequency within 40 bp around the integration site of EIAV$_{FDDV13}$ in the horse genome. (**A**) *Ssp* I-digested horse genome; (**B**) *Dra* I-digested horse genome. The diagrams were generated using the WebLOGO program (Available online: http://weblogo.berkeley.edu/). The relative lengths of the nucleotides on the Y-axis represent the frequency of each base.

3.3. EIAV_{FDDV13} Provirus Tended to Integrate into Host RefSeq Genes

A previous study on lentiviral integration examining HIV-1, FIV and SIV integration into the human genome found that integrations more frequently occurred in regions between the transcriptional start and stop sites of genes from the Reference Sequence (RefSeq) database [19–21]. In the present study, we asked whether this pattern occurred in the integration of the EIAV provirus into the horse genome. Our results showed that 53.0% (152/287) of integrations in the *Ssp* I group and 55.3% (105/190) in the *Dra* I group occurred in RefSeq genes. The integration rate was not notably different between these two groups ($p > 0.05$) (Table 2), but it was significantly higher than those from a set of 10,000 computer-generated random insertion sites, of which only 33.2% occurred in RefSeq genes ($p < 0.01$) (Table 2). We also analyzed the frequency of integration of the EIAV and HIV-1 in RefSeq genes from human chromosomes using the updated genome database (UCSC, assembled in December 2013). In total, 77.4% (113/146) of HIV-1 and 75.2% (313/416) of EIAV integrations occurred in RefSeq genes, which is generally consistent with the 72% (116/162) for HIV-1 and 68% (311/458) for EIAV ($P > 0.05$) previously reported by Hacker *et al*. [33] in 2006 (Table 2). The integration frequencies for both EIAV and HIV-1 in the human chromosomes were significantly higher than that of EIAV in horse chromosomes observed in this study ($p < 0.01$).

Table 2. Comparison between groups for the integration sites within genes.

	Random	HIV-1 vector	EIAV vector		EIAV-*Ssp* I group				EIAV-*Dra* I group				
	No. (%)	No. (%)	No. (%)	P2 [d]	No. (%)	P1 [c]	P2 [d]	P3 [e]	No. (%)	P1 [c]	P2 [d]	P3 [e]	P4 [f]
Total	10000 (100.0)	146 (100.0)	416 (100.0)	NA	287 (100.0)	NA	NA	NA	190 (100.0)	NA	NA	NA	NA
In RefSeq genes [a]	3315 (33.2)	113 (77.4)	313 (75.2)	NS	152 (53.0)	**	**	**	105 (55.3)	**	**	**	NS
Sense [b]	ND (ND)	65 (57.5)	149 (47.6)	NS	79 (52.0)	NA	NS	NS	63 (60.0)	NA	NS	*	NS

RefSeq genes represent reference sequence genes; [a]: The percentage of the provirus integrated in RefSeq genes was calculated and indicated in brackets; [b]: The percentage of sense-direction insertion in the genome was calculated among RefSeq genes in which the viral genome was integrated; [c]: *P1* shows a comparison with the random group using the chi-squared test; [d]: *P2* shows a comparison with the HIV-1 vector group using the chi-squared test; [e]: *P3* shows a comparison with the EIAV vector group using the chi-squared test; [f]: *P4* shows a comparison with the *Ssp* I group using the chi-squared test; NA indicates not applicable; ND indicates not done. NS indicates not significant; * indicates significant, *i.e.*, $p < 0.05$; ** indicates very significant, *i.e.*, $p < 0.01$.

3.4. EIAV_{FDDV13} Integration Targeted Low Gene Density Regions and Avoided the TSS

Our results presented above suggest that EIAV_{FDDV13} has a preference for integration into RefSeq genes, *i.e.*, the transcription units. Therefore, we further investigated some other characteristics of the genome that may relate to retrovirus integration. These characteristics include (a) the insertion orientation relative to RefSeq genes; (b) the gene density within a 2 Mb region around the integration site; (c) the integration frequency near the TSS.

Our results showed that no differences existed in the insertion orientation either for the total integration sites (Figure 3) or for the sites integrated into RefSeq genes in both horse and human genome (Table 2). In addition, the frequencies of insertion orientation were similar between the *Ssp* I and *Dra* I groups of the EIAV integrated into the horse genome and the EIAV and HIV-1 integrated into the human genome ($p > 0.05$). Furthermore, analytic data revealed that EIAV$_{FDDV13}$ appeared to have a preference for integration into regions with low (11–20 RefSeq/2 Mb for *Ssp* I groups) or moderate gene density (21–40 RefSeq/2 Mb for *Ssp* I and *Dra* I groups) in the horse chromosomes ($p < 0.05$), which was similar to the mean value of gene density in the horse chromosomes (the mean value was 29.0 RefSeq/2 Mb). In contrast, our analytic data showed that integration events of EIAV tended to occur in regions of moderate gene density (21–40 RefSeq/2 Mb) in the human chromosomes (39.3 RefSeq/2 Mb) ($p < 0.05$). However, HIV-1 tended to occur in regions of high gene density (>61 RefSeq/2 Mb) in the human chromosomes ($p < 0.01$) (Figure 4), which was consistent with previously published results [33]. Our data also revealed that EIAV$_{FDDV13}$ integration generally avoided the TSS in the horse chromosomes, similarly to the results obtained on the transfection of EIAV and HIV-1 vectors into human cells (Figure 5).
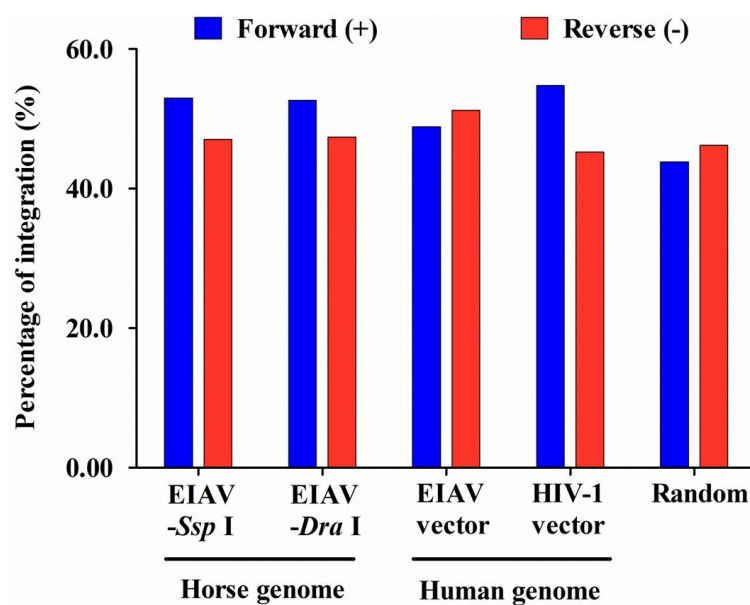

Figure 3. Orientation of the integration sites relative to the TSS. The percentages were calculated according to the total number of integration sites in each type of sample. The chi-squared test was used.
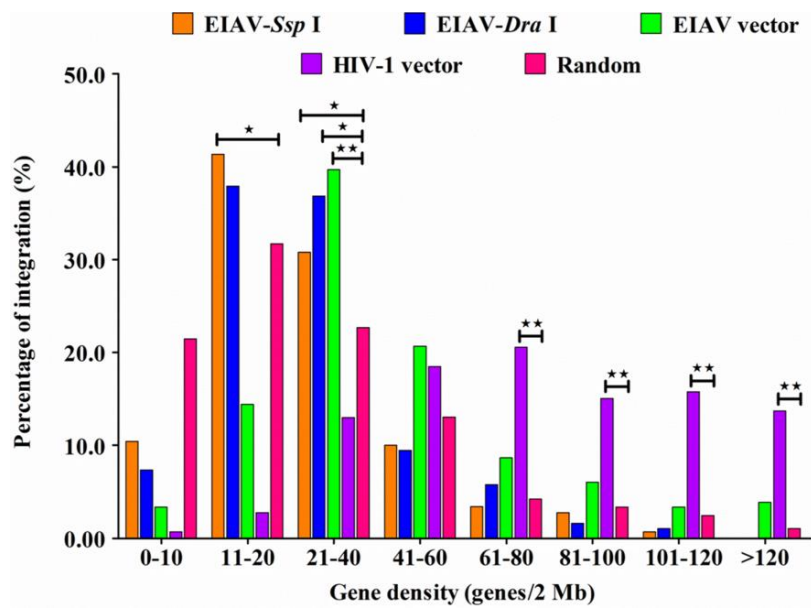
Figure 4. Gene density within a 2 Mb window centered on the insertion site was determined and classified into eight groups (0–10, 11–20, 21–40, 41–60, 61–80, 81–100, 101–120, >120). The percentage of integration within each group was calculated. The number of integration sites analyzed in the *Ssp* I and *Dra* I group was based on the horse genome; and the HIV-1 and EIAV vector group was based on the human genome. The chi-squared test was used. * indicates significant, *i.e.*, $p < 0.05$; ** indicates very significant, *i.e.*, $p < 0.01$.
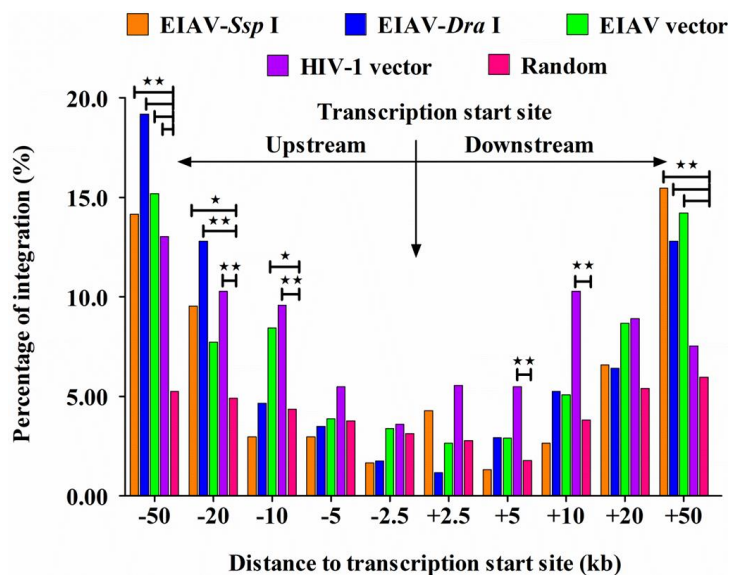

Figure 5. Frequencies of integration at different nucleotide windows up- and downstream of TSS. Five symmetric variable-size windows near the TSS in the host genome were indicated, and their corresponding frequencies were calculated (0–2.5, 2.5–5.0, 5.0–10, 10–20, and 20–50 kb). The number of integration sites analyzed in the *Ssp* I and *Dra* I group was based on the horse genome; and the HIV-1 and EIAV vector group was based on the human genome. The chi-squared test was used. * indicates significant, *i.e.*, $p < 0.05$; ** indicates very significant, *i.e.*, $p < 0.01$.

3.5. Integration Frequency of EIAV_{FDDV13} in Repetitive Sequences

A previous study on HIV-1 integration sites in repetitive elements indicated that this virus favors *Alu* repeats of the short interspersed nuclear elements (SINEs) family, and no bias was seen in insertion in the long interspersed nuclear elements (LINEs) [19,46]. In this study, we investigated whether EIAV integration in horse repetitive elements exhibited preferences. Repetitive sequences consist of approximately half of the horse genome (49.5%), and contain four major classes of transposable elements (TEs): LINE retrotransposons, SINE retrotransposons, DNA transposons and LTR transposons [39]. The relative proportions of the various repetitive sequences in the horse or human genome were obtained from a previous report [39,47] and are listed in Table 3 and Table S2, respectively.

Regarding the EIAV_{FDDV13} integration sites that occurred in repetitive elements, 69 of 120 sites (57.5%) from the *Ssp* I group and 38 of 69 sites (55.1%) from the *Dra* I group were found to occur in LINEs. Compared to the 37.9% of integration within LINEs among the matched random insertion sites occurred in repetitive elements, EIAV had a significant preference for integration within LINEs in the horse genome ($p < 0.01$) (Table 3). Similar analytic results were obtained when these integration data were compared with a set of sites of matched random control (MRC) generated based on the methods as described by Mitchell *et al.* [24] and Berry *et al.* [48]. Conversely, the frequency of EIAV_{FDDV13} integration in SINEs of horse chromosomes was 12.5% (15 of 120 sites) for the *Ssp* I group and 13.0% (9 of 69 sites) for the *Dra* I group. Considering the portions of integration within LINEs and SINEs among the matched random insertion sites occurred in repetitive elements and the insertion sites in each of the elements (37.9% and 13.8%, respectively), EIAV appeared to prefer to integrate in LINEs compared to SINEs in the horse genome ($p < 0.01$) (Table 3). In contrast, the EIAV and HIV-1 had similar frequenciy of integration into both LINEs (37.5% for EIAV and 25.9% for HIV-1) and SINEs (36.3% for EIAV and 49.4% for HIV-1) in the human genome (Table S2). Compared with the proportions of integration within LINEs or SINEs among the matched random insertion sites occurred in repetitive elements (40.2% or 29.7%, respectively), the HIV-1 favored SINEs to LINEs in the human genome ($p < 0.01$) (Table S2).

Table 3. Correlation between repetitive elements and EIAV integration sites in the horse genome.

	Random control	EIAV-*Ssp* I		EIAV-*Dra* I		
	No. (%) [a]	No. (%)	*P1* [c]	No. (%)	*P1* [c]	*P2* [d]
Total [b]	4876	120		69		
LINEs	1848 (37.9)	69 (57.5)	**	38 (55.1)	**	NS
SINEs	675 (13.8)	15 (12.5)	NS	9 (13.0)	NS	NS
DNA transposons	322 (6.6)	21 (17.5)	**	15 (21.7)	**	NS
LTR transposons	658 (13.5)	11 (9.2)	NS	6 (8.7)	NS	NS

[a]: Percentages of repetitive elements in the host genome were based on the total numbers of integration sites inserted into repetitive elements in the random control; [b]: The number of integration sites inserted into the repetitive elements was calculated based on the horse genome; [c]: *P1* displays a comparison of the frequency of corresponding repetitive elements in the random control using the chi-squared test; [d]: *P2* presents a comparison with *Ssp* I group integration using the chi-squared test. NS indicates not significant; ND indicates not determined; * indicates significant, *i.e.*, $p < 0.05$; ** indicates very significant, *i.e.*, $p < 0.01$.

In addition, the *Ssp* I group had 17.5% (21/120) and the *Dra* I group had 21.7% (15/69) of integration sites in repetitive sequences located in DNA transposons of the horse genome, percentages that were significantly higher than the proportion of integration within DNA transposons among the matched random insertion sites occurred in repetitive elements in the horse genome ($p < 0.01$) (Table 3). Similarly, 20 of 160 insertions (12.5%) of EIAV were identified in DNA transposons of human repetitive sequences, a level that was significantly higher than the proportion of integration within this type of element among the matched random insertion sites occurred in repetitive elements in the human genome ($p < 0.01$) (Table S2). However, HIV-1 did not display any preference for DNA transposons in the human genome ($p > 0.05$) (Table S2). The differences in the integration preferences within other repetitive elements were not significant.

4. Discussion

In this study, we examined the integration characteristics of an FED cell-adapted EIAV strain, EIAV$_{FDDV13}$, in the horse genome. Our results demonstrated that during infection of FED cells, one of the most frequently used types of host cells for *in vitro* EIAV studies, the integration sites of EIAV$_{FDDV13}$ provirus were not randomly distributed but instead were prone to fall within RefSeq genes. Overall, these data implied that EIAV$_{FDDV13}$ shares several similarities in integration site selection with other lentiviruses, such as HIV-1 [19], SIV [20], and FIV [21] but obviously differs from MLV and ASLV in the human genome [22,23]. Notably, our results showed that EIAV$_{FDDV13}$ preferred to integrate into LINEs compared to SINEs and other repetitive elements in the horse genome, while HIV-1 tended to insert within SINEs in the human genome. More interestingly, EIAV favored integration within DNA transposable elements in both the horse and human genomes, but HIV-1 did not display this preference in the human genome. Therefore, this study showed some unique integration features of EIAV$_{FDDV13}$ integration in horse genomes, and it also revealed similarities of EIAV integration site selection in the horse and human genomes, which improves our understanding of lentiviral integration and latency.

The usage of restriction enzymes to digest host gDNA during viral integration site cloning could have introduced biases, as this method tends to isolate integration sites near the restriction sites. However, the observation of integration in *in vitro* experimental controls suggested that the methods used for isolation and analysis of integration sites did not detectably bias the conclusions [19]. Additionally, various cloning and analytic strategies for examining HIV-1 integration sites in the human genome, including the fragmentation of gDNA with restriction enzymes or physical approaches [49], amplification with inverse PCR (I-PCR) [50] or LM-PCR [19,30] and high-throughput sequencing [49,51], obtained similar conclusions on the integration characteristics in different types of human cells [48,52–54]. Similarly, consistent integration features were observed for MLV in the human genome using different restriction enzymes and PCR methods [22,54–56]. Therefore, these techniques all could effectively clone retroviral integration sites without causing significant bias. In the present study, two different restriction endonucleases, *Ssp* I and *Dra* I, were applied for LM-PCR of EIAV insertion sites in FED cells, which further increased the unbiased nature of the cloning of gDNA fragments that integrated with EIAV provirus.

EIAV$_{FDDV13}$ integration clearly tended to occur in RefSeq genes. Similar patterns of integration targeting have been reported for other lentiviruses [19–21]. However, the frequency of integration of

EIAV$_{FDDV13}$ in these gene-containing regions of the horse genome was significantly lower than that of the integration of HIV-1 and SIV (72% for HIV-1 and 74% for SIV) [20] or EIAV and HIV-1 in the human genome as analyzed in this study (77.4% and 75.2%, respectively). It is logical to presume that this difference may be partially attributed to the incomplete annotation of RefSeq genes in horse chromosomes. Although the horse genome shares extensive conserved synteny with the human genome, the total size is smaller (2.5–2.7 Gb for horses compared with 2.9 Gb for humans), and only 26,740 RefSeq genes have been identified in horse chromosomes (ENSEMBL Genes 79), which is significantly fewer than the 59,492 genes identified in human chromosomes (ENSEMBL Genes 79). In addition, different types of cells from the same host or cells from different species of hosts have variant replication rates and activation states, which might alter interactions between viral pre-integration complex (PIC) and host cell proteins and thereafter influence EIAV integration in RefSeq genes. These factors may partially explain our observation of a relatively low integration frequency of EIAV in RefSeq genes in the horse genome. Certainly, the specific structures and functions of viral proteins in each type of virus also affect insertion rates in RefSeq genes.

The results from this study and others revealed some specific integration features among different species of retroviruses. It is known that integration site selection in the human genome is largely determined by viral integrases and cellular factors that bind to and interact with integrases [57,58]. Indeed, integrases from all lentivirus members share highly conserved structural features. Lens epithelium-derived growth factor (LEDGF), which is conserved in natural hosts of lentiviruses, interacts with lentiviral integrases by forming a complex and mediates viral integration in the host chromosomes [59,60]. In addition, viral proteins other than integrase have also been shown to affect integration site selection in HIV-1. Studies using HIV-1/MLV chimeric viruses indicated that HIV-1 capsid proteins have implicated a role in influencing target site preference for integration [57,61]. In addition, recent studies have demonstrated that the bromo- and extra-terminal domain (BET) protein, which mediates gammaretrovirus integration near the transcription initiation site, is a major cellular factor that covalently binds MLV integrase [62–64]. Our data on some unique integration characteristics of EIAV$_{FDDV13}$ from this study provide additional evidence for the specific integration preference of the *Retroviridae* genus of lentivirus.

Our analysis revealed that the EIAV provirus is more prone to integrate in LINEs than in SINEs in the horse genome. Concurrently, consistent with some publications, HIV-1 prefers inserting in SINEs (especially *Alus*) in the human genome compared to LINEs ([19] and this study). There are two possible explanations for these differences: The first explanation is that host proteins in cells of different origins involved in retroviral integration are in different cell types and have different activities. Because the integration of the EIAV strain in horse cells favored LINEs but the HIV in human cell line 293 preferred SINEs, this possibility is at least partially responsible for this difference. The second explanation involves the different copy numbers and ratios of LINEs to SINEs between the horse and human genome. These two types of elements consist of 19.7% and 7.0% of the horse genome and 20% and 13% of the human genome, respectively. The lower percentage of SINEs identified in the horse genome could occur because (1) the horse genome contains fewer SINE elements; or (2) the horse genome has been incompletely annotated, which is less likely because we compared the proportion of LINEs and SINEs from clones of the same genome and it is unlikely that only SINEs were largely unidentified in the

genomic database. For other lentiviruses, the chromosomal environment around the integration sites influences the replication and gene expression of the integrated provirus [30–32,50]. LINE elements are significantly enriched in GC- and gene-poor regions [65]. Genes with high expression usually do not occur near LINEs, and the presence of LINEs will reduce the transcriptional activity of genes, possibly due to the presence of a strong polyA signal within LINEs, which can destroy transcription elongation [66]. Most horses infected with EIAV will eventually become asymptomatic carriers [25,26]. The integration site preference, including in LINEs and DNA transposons, is presumed to be one possible driving force that contributes to this infection behavior.

DNA transposon sequences, which are able to move from one genomic location to another by a cut-and-paste mechanism (transposition), make up approximately 3% of the host genome [47]. Currently, DNA transposons, in particular *Sleeping Beauty* (SB), a member of the Tc1/mariner family of DNA transposons, are most widely used as a gene therapy tool for human genetic diseases [67]. Studies on integration site selection found that the integration of SB was nonrandom in the host chromosome. This transposon strongly favored integration into non-coding regions, including heterochromatin and microsatellite repeats [68]. Using an HIV-1-based vector transducing GFP in Jurkat cells, Lewinski *et al*. [30] reported that inducible expression (similar to the latent state) of GFP was more frequently identified in (1) very highly expressed host genes; (2) gene deserts; and (3) heterochromatin. These results suggest that the chromosomal environment surrounding the DNA transposons' insertion sites influences the activity of the transposons. In this study, EIAV tended to insert into DNA transposons in the horse and human genome. It is logical to infer that the activation of EIAV proviruses integrated within DNA transposons could be suppressed by the chromosomal environment surrounding the DNA transposon insertion sites, which may be one of the reason that EIAV tend to be silenced in long-term infected horse.

The EIAV vector, similarly to other lentiviral vectors such as HIV-1, has certain advantages as a therapeutic gene transfer vector. It is important to determine what integration is like for the EIAV vector in horse genome, *i.e*., if the same result that was observed in human cells would be seen in equid cells. We recently completed the construction of a three-plasmid, VSV-G pseudotyped EIAV transfection system, which will be used together with a three-plasmid HIV-1 transfection system, to compare the integration of EIAV vector in the horse genome to that of the human genome.

Acknowledgments

Author Contributions

Jian-Hua Zhou, Xiao-Jun Wang and Qiang Liu designed the experiments and wrote the paper. Qiang Liu performed the experiments and analyzed the data, Qiang Liu, Xue-Feng Wang, Xi-Jun He, and Jian Ma uploaded the sequences. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Coffin, J.M.; Hughes, S.H.; Varmus, H.E. The Interactions of Retroviruses and their Hosts. In *Retroviruses*; Coffin, J.M., Hughes, S.H., Varmus, H.E., Eds.; Cold Spring Harbor: New York, NY, USA, 1997.

2. Van Maele, B.; Debyser, Z. HIV-1 integration: An interplay between HIV-1 integrase, cellular and viral proteins. *AIDS Rev.* **2005**, *7*, 26–43. [PubMed]

3. Debyser, Z.; Christ, F.; De Rijck, J.; Gijsbers, R. Host factors for retroviral integration site selection. *Trends Biochem. Sci.* **2015**, *40*, 108–116. [CrossRef] [PubMed]

4. Biasco, L.; Baricordi, C.; Aiuti, A. Retroviral integrations in gene therapy trials. *Mol. Ther. J. Am. Soc. Gene Ther.* **2012**, *20*, 709–716. [CrossRef] [PubMed]

5. Boztug, K.; Dewey, R.A.; Klein, C. Development of hematopoietic stem cell gene therapy for Wiskott-Aldrich syndrome. *Curr. Opin. Mol. Therapeutics* **2006**, *8*, 390–395.

6. Fischer, A.; Hacein-Bey-Abina, S.; Cavazzana-Calvo, M. Gene therapy for primary adaptive immune deficiencies. *J. Allergy Clin. Immunol.* **2011**, *127*, 1356–1359. [CrossRef] [PubMed]

7. Rivat, C.; Santilli, G.; Gaspar, H.B.; Thrasher, A.J. Gene therapy for primary immunodeficiencies. *Hum. Gene Ther.* **2012**, *23*, 668–675. [CrossRef] [PubMed]

8. Verma, I.M.; Weitzman, M.D. Gene therapy: Twenty-first century medicine. *Annu. Rev. Biochem.* **2005**, *74*, 711–738. [CrossRef] [PubMed]

9. Pike-Overzet, K.; van der Burg, M.; Wagemaker, G.; van Dongen, J.J.; Staal, F.J. New insights and unresolved issues regarding insertional mutagenesis in X-linked SCID gene therapy. *Mol. Ther. J. Am. Soc. Gene Ther.* **2007**, *15*, 1910–1916. [CrossRef] [PubMed]

10. Cavazzana-Calvo, M.; Fischer, A. Gene therapy for severe combined immunodeficiency: Are we there yet? *J. Clin. Investig.* **2007**, *117*, 1456–1465. [CrossRef] [PubMed]

11. Voigt, K.; Izsvak, Z.; Ivics, Z. Targeted gene insertion for molecular medicine. *J. Mol. Med.* **2008**, *86*, 1205–1219. [CrossRef] [PubMed]

12. 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.; *et al.* *LMO2*-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **2003**, *302*, 415–419. [CrossRef] [PubMed]

13. Hacein-Bey-Abina, S.; Garrigue, A.; Wang, G.P.; Soulier, J.; Lim, A.; Morillon, E.; Clappier, E.; Caccavelli, L.; Delabesse, E.; Beldjord, K.; *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Investig.* **2008**, *118*, 3132–3142. [CrossRef] [PubMed]

14. Hacein-Bey-Abina, S.; von Kalle, C.; Schmidt, M.; Le Deist, F.; Wulffraat, N.; McIntyre, E.; Radford, I.; Villeval, J.L.; Fraser, C.C.; Cavazzana-Calvo, M.; Fischer, A. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **2003**, *348*, 255–256. [CrossRef] [PubMed]

15. Wang, G.P.; Garrigue, A.; Ciuffi, A.; Ronen, K.; Leipzig, J.; Berry, C.; Lagresle-Peyrou, C.; Benjelloun, F.; Hacein-Bey-Abina, S.; Fischer, A.; Cavazzana-Calvo, M.; Bushman, F.D. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucl. Acids Res.* **2008**, *36*, e49. [CrossRef] [PubMed]

16. Wang, G.P.; Levine, B.L.; Binder, G.K.; Berry, C.C.; Malani, N.; McGarrity, G.; Tebas, P.; June, C.H.; Bushman, F.D. Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. *Mol. Ther. J. Am. Soc. Gene Ther.* **2009**, *17*, 844–850. [CrossRef] [PubMed]

17. Montini, E.; Cesana, D.; Schmidt, M.; Sanvito, F.; Bartholomae, C.C.; Ranzani, M.; Benedicenti, F.; Sergi, L.S.; Ambrosi, A.; Ponzoni, M.; Doglioni, C.; di Serio, C.; von Kalle, C.; Naldini, L. 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. Investig.* **2009**, *119*, 964–975. [CrossRef] [PubMed]

18. Papayannakos, C.; Daniel, R. Understanding lentiviral vector chromatin targeting: Working to reduce insertional mutagenic potential for gene therapy. *Gene Ther.* **2013**, *20*, 581–588. [CrossRef] [PubMed]

19. Schroder, A.R.; Shinn, P.; Chen, H.; Berry, C.; Ecker, J.R.; Bushman, F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **2002**, *110*, 521–529. [CrossRef]

20. Crise, B.; Li, Y.; Yuan, C.; Morcock, D.R.; Whitby, D.; Munroe, D.J.; Arthur, L.O.; Wu, X. Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* **2005**, *79*, 12199–12204. [CrossRef] [PubMed]

21. Kang, Y.; Moressi, C.J.; Scheetz, T.E.; Xie, L.; Tran, D.T.; Casavant, T.L.; Ak, P.; Benham, C.J.; Davidson, B.L.; McCray, P.B., Jr. Integration site choice of a feline immunodeficiency virus vector. *J. Virol.* **2006**, *80*, 8820–8823. [CrossRef] [PubMed]

22. Wu, X.; Li, Y.; Crise, B.; Burgess, S.M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **2003**, *300*, 1749–1751. [CrossRef] [PubMed]

23. Narezkina, A.; Taganov, K.D.; Litwin, S.; Stoyanova, R.; Hayashi, J.; Seeger, C.; Skalka, A.M.; Katz, R.A. Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol.* **2004**, *78*, 11656–11663. [CrossRef] [PubMed]

24. Mitchell, R.S.; Beitzel, B.F.; Schroder, A.R.; Shinn, P.; Chen, H.; Berry, C.C.; Ecker, J.R.; Bushman, F.D. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2004**, *2*, e234. [CrossRef] [PubMed]

25. Craigo, J.K.; Montelaro, R.C. Lessons in AIDS vaccine development learned from studies of equine infectious, anemia virus infection and immunity. *Viruses* **2013**, *5*, 2963–2976. [CrossRef] [PubMed]

26. Leroux, C.; Cadore, J.L.; Montelaro, R.C. Equine Infectious Anemia Virus (EIAV): What has HIV's country cousin got to tell us? *Vet. Res.* **2004**, *35*, 485–512. [CrossRef] [PubMed]

27. Harrold, S.M.; Cook, S.J.; Cook, R.F.; Rushlow, K.E.; Issel, C.J.; Montelaro, R.C. Tissue sites of persistent infection and active replication of equine infectious anemia virus during acute disease and asymptomatic infection in experimentally infected equids. *J. Virol.* **2000**, *74*, 3112–3121. [CrossRef] [PubMed]

28. Kono, Y.; Hirasawa, K.; Fukunaga, Y.; Taniguchi, T. Recrudescence of equine infectious anemia by treatment with immunosuppressive drugs. *Natl. Inst. Anim. Health Q.* **1976**, *16*, 8–15.

29. Craigo, J.K.; Leroux, C.; Howe, L.; Steckbeck, J.D.; Cook, S.J.; Issel, C.J.; Montelaro, R.C. Transient immune suppression of inapparent carriers infected with a principal neutralizing domain-deficient equine infectious anaemia virus induces neutralizing antibodies and lowers steady-state virus replication. *J. Gen. Virol.* **2002**, *83*, 1353–1359. [PubMed]

30. Lewinski, M.K.; Bisgrove, D.; Shinn, P.; Chen, H.; Hoffmann, C.; Hannenhalli, S.; Verdin, E.; Berry, C.C.; Ecker, J.R.; Bushman, F.D. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **2005**, *79*, 6610–6619. [CrossRef] [PubMed]

31. Pace, M.J.; Graf, E.H.; Agosto, L.M.; Mexas, A.M.; Male, F.; Brady, T.; Bushman, F.D.; O'Doherty, U. Directly infected resting CD4+ T cells can produce HIV Gag without spreading infection in a model of HIV latency. *PLoS Pathog.* **2012**, *8*, e1002818. [CrossRef] [PubMed]

32. Han, Y.; Lassen, K.; Monie, D.; Sedaghat, A.R.; Shimoji, S.; Liu, X.; Pierson, T.C.; Margolick, J.B.; Siliciano, R.F.; Siliciano, J.D. Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J. Virol.* **2004**, *78*, 6122–6133. [CrossRef] [PubMed]

33. Hacker, C.V.; Vink, C.A.; Wardell, T.W.; Lee, S.; Treasure, P.; Kingsman, S.M.; Mitrophanous, K.A.; Miskin, J.E. The integration profile of EIAV-based vectors. *Mol. Ther. J. Am. Soc. Gene Ther.* **2006**, *14*, 536–545. [CrossRef] [PubMed]

34. Marshall, H.M.; Ronen, K.; Berry, C.; Llano, M.; Sutherland, H.; Saenz, D.; Bickmore, W.; Poeschla, E.; Bushman, F.D. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* **2007**, *2*, e1340. [CrossRef] [PubMed]

35. Ma, J.; Shi, N.; Jiang, C.G.; Lin, Y.Z.; Wang, X.F.; Wang, S.; Lv, X.L.; Zhao, L.P.; Shao, Y.M.; Kong, X.G.; Zhou, J.H.; Shen, R.X. A proviral derivative from a reference attenuated EIAV vaccine strain failed to elicit protective immunity. *Virology* **2011**, *410*, 96–106. [CrossRef] [PubMed]

36. Ma, J.; Wang, S.S.; Lin, Y.Z.; Liu, H.F.; Wei, H.M.; Du, C.; Wang, X.F.; Zhou, J.H. An attenuated EIAV strain and its molecular clone strain differentially induce the expression of Toll-like receptors and type-I interferons in equine monocyte-derived macrophages. *Vet. Microbiol.* **2013**, *166*, 263–269. [CrossRef] [PubMed]

37. Qi, X.; Wang, X.; Wang, S.; Lin, Y.; Jiang, C.; Ma, J.; Zhao, L.; Lv, X.; Shen, R.; Wang, F.; Kong, X.; Su, Z.; Zhou, J. Genomic analysis of an effective lentiviral vaccine-attenuated equine infectious anemia virus vaccine EIAV FDDV13. *Virus Genes* **2010**, *41*, 86–98. [CrossRef] [PubMed]

38. Huang, J.; Zhao, Y.; Shiraigol, W.; Li, B.; Bai, D.; Ye, W.; Daidiikhuu, D.; Yang, L.; Jin, B.; Zhao, Q.; *et al*. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci. Rep.* **2014**, *4*. [CrossRef] [PubMed]

39. Wade, C.M.; Giulotto, E.; Sigurdsson, S.; Zoli, M.; Gnerre, S.; Imsland, F.; Lear, T.L.; Adelson, D.L.; Bailey, E.; Bellone, R.R.; *et al*. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **2009**, *326*, 865–867. [CrossRef] [PubMed]

40. Jiang, C.G.; Gao, X.; Ma, J.; Lin, Y.Z.; Wang, X.F.; Zhao, L.P.; Hua, Y.P.; Liu, D.; Zhou, J.H. *C*-terminal truncation of the transmembrane protein of an attenuated lentiviral vaccine alters its *in vitro* but not *in vivo* replication and weakens its potential pathogenicity. *Virus Res.* **2011**, *158*, 235–245. [CrossRef] [PubMed]

41. Morrow, C.D.; Park, J.; Wakefield, J.K. Viral gene products and replication of the human immunodeficiency type 1 virus. *Am. J. Physiol.* **1994**, *266*, C1135–C1156. [PubMed]

42. Ranki, A.; Lagerstedt, A.; Ovod, V.; Aavik, E.; Krohn, K.J. Expression kinetics and subcellular localization of HIV-1 regulatory proteins Nef, Tat and Rev in acutely and chronically infected lymphoid cell lines. *Arch. Virol.* **1994**, *139*, 365–378. [CrossRef] [PubMed]

43. Ciuffi, A.; Ronen, K.; Brady, T.; Malani, N.; Wang, G.; Berry, C.C.; Bushman, F.D. Methods for integration site distribution analyses in animal cell genomes. *Methods* **2009**, *47*, 261–268. [CrossRef] [PubMed]

44. Ciuffi, A.; Barr, S.D. Identification of HIV integration sites in infected host genomic DNA. *Methods* **2011**, *53*, 39–46. [CrossRef] [PubMed]

45. Holman, A.G.; Coffin, J.M. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6103–6107. [CrossRef] [PubMed]

46. Wang, G.P.; Ciuffi, A.; Leipzig, J.; Berry, C.C.; Bushman, F.D. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **2007**, *17*, 1186–1194. [CrossRef] [PubMed]

47. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; *et al*. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [CrossRef] [PubMed]

48. Berry, C.; Hannenhalli, S.; Leipzig, J.; Bushman, F.D. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* **2006**, *2*, e157. [CrossRef] [PubMed]

49. Maldarelli, F.; Wu, X.; Su, L.; Simonetti, F.R.; Shao, W.; Hill, S.; Spindler, J.; Ferris, A.L.; Mellors, J.W.; Kearney, M.F.; Coffin, J.M.; Hughes, S.H. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **2014**, *345*, 179–183. [CrossRef] [PubMed]

50. Shan, L.; Yang, H.C.; Rabi, S.A.; Bravo, H.C.; Shroff, N.S.; Irizarry, R.A.; Zhang, H.; Margolick, J.B.; Siliciano, J.D.; Siliciano, R.F. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J. Virol.* **2011**, *85*, 5384–5393. [CrossRef] [PubMed]

51. Cohn, L.B.; Silva, I.T.; Oliveira, T.Y.; Rosales, R.A.; Parrish, E.H.; Learn, G.H.; Hahn, B.H.; Czartoski, J.L.; McElrath, M.J.; Lehmann, C.; *et al*. HIV-1 integration landscape during latent and active infection. *Cell* **2015**, *160*, 420–432. [CrossRef] [PubMed]

52. Sherrill-Mix, S.; Lewinski, M.K.; Famiglietti, M.; Bosque, A.; Malani, N.; Ocwieja, K.E.; Berry, C.C.; Looney, D.; Shan, L.; Agosto, L.M.; *et al*. HIV latency and integration site placement in five cell-based models. *Retrovirology* **2013**, *10*. [CrossRef] [PubMed]

53. Ciuffi, A.; Mohammadi, P.; Golumbeanu, M.; di Iulio, J.; Telenti, A. Bioinformatics and HIV Latency. *Curr. HIV/AIDS Rep.* **2015**, *12*, 97–106. [CrossRef] [PubMed]

54. Desfarges, S.; Ciuffi, A. Retroviral integration site selection. *Viruses* **2010**, *2*, 111–130. [CrossRef] [PubMed]

55. 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.; Malech, H.L.; Wu, X. Enhancers are major targets for murine leukemia virus vector integration. *J. Virol.* **2014**, *88*, 4504–4513. [CrossRef] [PubMed]

56. Tsukahara, T.; Agawa, H.; Matsumoto, S.; Matsuda, M.; Ueno, S.; Yamashita, Y.; Yamada, K.; Tanaka, N.; Kojima, K.; Takeshita, T. Murine leukemia virus vector integration favors promoter regions and regional hot spots in a human T-cell line. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 1099–1107. [CrossRef] [PubMed]

57. 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.; Bushman, F.D. Retroviral DNA integration: Viral and cellular determinants of target-site selection. *PLoS Pathog.* **2006**, *2*, e60. [CrossRef] [PubMed]

58. Kvaratskhelia, M.; Sharma, A.; Larue, R.C.; Serrao, E.; Engelman, A. Molecular mechanisms of retroviral integration site selection. *Nucl. Acids Res.* **2014**, *42*, 10209–10225. [CrossRef] [PubMed]

59. Busschots, K.; Vercammen, J.; Emiliani, S.; Benarous, R.; Engelborghs, Y.; Christ, F.; Debyser, Z. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* **2005**, *280*, 17841–17847. [CrossRef] [PubMed]

60. Cherepanov, P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity *in vitro*. *Nucl. Acids Res.* **2007**, *35*, 113–124. [CrossRef] [PubMed]

61. Schaller, T.; Ocwieja, K.E.; Rasaiyaah, J.; Price, A.J.; Brady, T.L.; Roth, S.L.; Hue, S.; Fletcher, A.J.; Lee, K.; KewalRamani, V.N.; *et al*. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* **2011**, *7*, e1002439. [CrossRef] [PubMed]

62. 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.; Gijsbers, R.; Debyser, Z. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* **2013**, *5*, 886–894. [CrossRef] [PubMed]

63. Gupta, S.S.; Maetzig, T.; Maertens, G.N.; Sharif, A.; Rothe, M.; Weidner-Glunde, M.; Galla, M.; Schambach, A.; Cherepanov, P.; Schulz, T.F. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* **2013**, *87*, 12721–12736. [CrossRef] [PubMed]

64. Sharma, A.; Larue, R.C.; Plumb, M.R.; Malani, N.; Male, F.; Slaughter, A.; Kessl, J.J.; Shkriabai, N.; Coward, E.; Aiyer, S.S.; *et al*. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12036–12041. [CrossRef] [PubMed]

65. Smit, A.F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **1999**, *9*, 657–663. [CrossRef]

66. Han, J.S.; Szak, S.T.; Boeke, J.D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **2004**, *429*, 268–274. [CrossRef] [PubMed]

67. Munoz-Lopez, M.; Garcia-Perez, J.L. DNA transposons: Nature and applications in genomics. *Curr. Genomics* **2010**, *11*, 115–128. [CrossRef] [PubMed]

68. Yant, S.R.; Wu, X.; Huang, Y.; Garrison, B.; Burgess, S.M.; Kay, M.A. High-resolution genome-wide mapping of transposon integration in mammals. *Mol. Cell. Biol.* **2005**, *25*, 2085–2094. [CrossRef] [PubMed]