

UC Merced

UC Merced Previously Published Works

Title

Efficient Nonviral Stable Transgenesis Mediated by Retroviral Integrase

Permalink

<https://escholarship.org/uc/item/1bk3v1r6>

Authors

Chiang, Chang-Ying
Ligunas, Gloria Denise
Chin, Wei-Chun
et al.

Publication Date

2020-06-01

DOI

10.1016/j.omtm.2020.04.020

Peer reviewed

Efficient Nonviral Stable Transgenesis Mediated by Retroviral Integrase

Chang-Ying Chiang,¹ Gloria Denise Ligunas,² Wei-Chun Chin,^{1,2} and Chih-Wen Ni^{1,2}

¹Department of Bioengineering, School of Engineering, University of California, Merced, Merced, CA, USA; ²Program of Quantitative and Systems Biology, University of California, Merced, Merced, CA, USA

Efficient transgene delivery is critical for genetic manipulation and therapeutic intervention of target cells. Two well-characterized integrative systems have been described that rely on viral and nonviral vectors. However, use of viral vectors for gene therapy has been associated with several safety concerns. Here, we report a virus-free method for stable transgenesis based on the reaction of retroviral integrase. We constructed a gateway cloning compatible vector containing two truncated long terminal repeat (LTR) sequences (dLTR) that flank the transgene cassette. Notably, 5'-ACTG-3' and blunt-end restriction cutting sites were also embedded at the end of dLTR to be recognized by HIV-1 integrase. When performing coinjection of transgene cassette and integrase mRNA into zebrafish embryos at one cell stage, there were 50% to 55% of injected embryos expressing a marker gene in a desired pattern. When applying our method in mammalian cells, there were 42% of cultured human epithelial cell lines showing stable integration. These results demonstrated that our method can successfully insert an exogenous gene into the host genome with highly efficient integration. Importantly, this system operates without most of the viral components while retaining effective stable transgenesis. We anticipate this method will provide a convenient, safe, and highly efficient way for applications in transgenesis and gene therapy.

INTRODUCTION

Stable transgenesis is a powerful tool in biomedical research and clinical applications. The methods for transgenesis include two well-characterized integrative systems that have been described, relying on viral vectors and nonviral vehicles.^{1,2} In general, the use of viral vectors is more efficient for stable gene expression, whereas nonviral vehicles are immune for certain virus-related, hazardous concerns.^{2,3} The safety concerns of using viral vectors include insertional mutagenesis, host's immune responses toward the transgene and/or its product, and postintegrative promoter silencing.^{3,4} Despite the mentioned concerns, efficient transgene delivery is critical for genetic manipulation and therapeutic intervention of target cells. Therefore, efforts to improve the safety of using viral vectors have been extensively imposed.

Currently, there are four major viral vector systems used for transgenesis, including nonintegrative adenoviruses, adeno-associated viruses

(AAV), and integrative retroviruses/lentiviruses. Among them, lentiviral vectors have been routinely used as a promising tool for effective delivery of transgene into cells with the major advantages of persistent gene transfer in most tissues and nondividing cells.^{1-3,5} Unlike adenovirus, lentiviral vectors do not seem to induce immune or inflammatory responses against viral proteins. With the efforts to engineer the lentiviral vector, including the pseudotyping of vesicular stomatitis virus G protein (VSV-G), retaining less than 5% of the parental genome and incorporating less than 25% of the viral genome into packaging constructs, biosafety has been further improved by the development of SIN (self-inactivating) vectors that contain the truncated LTR (long terminal repeat) sequence, which removes many regulatory elements.^{6,7} Still, the main limitation is the incident of causing oncogenesis due to insertional mutagenesis by inappropriately activating transcription of a nearby host gene.⁸ Fortunately, the understanding of the retrovirus integration mechanism could provide a new design to achieve a method for nonviral stable transgenesis. In the case of human HIV-1, the ends of LTR are recognized by a *pol*-encoding integrase.⁹ Integrase is responsible for the integration of viral DNA into a host genome during the viral replication. It possesses two major catalytic activities: 3' processing and strand transfer.¹⁰ The 3' processing is a highly specific reaction involving the removal of a dinucleotide (GT), adjacent to the highly conserved CA dinucleotide, from the 3' strand of the U3 and U5 viral DNA LTRs.¹¹ The resulting cleaved DNA is then used as a substrate for integration or strand transfer, leading to the covalent insertion of the viral DNA into the genome of the infected cell.¹⁰ Mutations in this sequence completely abolish activity, where the integrity of flanking sequences is much less important.^{11,12} Besides, the activity of integrase is strongly dependent on its oligomeric state.^{13,14} In contrast to 3' processing, which requires the dimeric form of integrase,¹⁵ it was shown that concerted integration requires at least a tetrameric form of integrase,^{16,17} and the structure analysis of intasomes revealed that a higher-order integrase multimer can self-associate into a functional intasome.^{18,19} In addition, it has been reported that for an *in vitro* integration assay, the use of LTR and integrase is sufficient for a successful DNA insertion;^{9,11,20} however, full-length LTRs are not required for this reaction, but only the

Received 23 March 2020; accepted 27 April 2020;
<https://doi.org/10.1016/j.omtm.2020.04.020>

Correspondence: Chih-Wen Ni, University of California, Merced, 5200 N. Lake Road, Merced, CA 95343, USA.

E-mail: cni3@ucmerced.edu



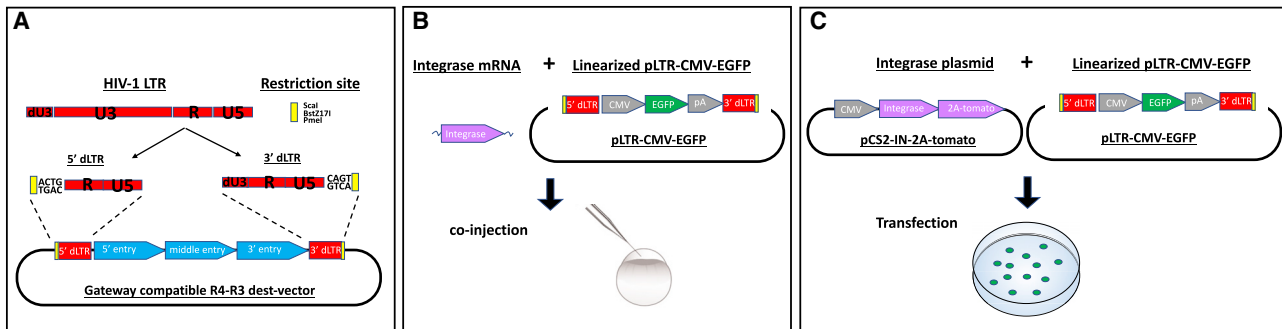


Figure 1. Schematic Diagrams of the pLTR Vectors and the Procedure of Transgenesis

(A) The design of the pLTR vector. The pLTR vector was revised from a gateway destination vector, pmiTo2 R4-R3 by replacing the miniTo2 sequence with 5' dLTR and 3' dLTR, which are truncated sequences adapted from HIV-1 LTR. The end of dLTR also embedded with 5'-ACTG-3' and restriction cutting sites, such as *ScaI*, *BstZ171*, and *PmeI*. (B) The procedure of microinjection in zebrafish. The transgene cassette was linearized by restriction digestion to generate a blunt-end fragment. This cassette is premixed with integrase mRNA for microinjection to zebrafish embryos in one-cell stage. (C) The procedure of transfection. The transgene cassette was linearized by restriction digestion to generate a blunt-end fragment. This cassette is premixed with an overexpression plasmid of HIV integrase, pCS2-integrase-2A-tdTomato, and co-transfected to culture cells.

final U5 or U3 ends, approximately 20 nt, are imperative for the insertion.^{20,21}

By means of the *in vitro* integration mechanism of HIV-1, it is possible to develop a method to achieve efficient integration of a nonviral gene-delivery system using only two retrovirus elements: the LTR and integrase. Previous works have reported the efforts and attempts to integrate a transgene into a host genome by using integrase from bovine leukemia virus (BLV)²² or Rous sarcoma virus (RSV).²³ However, the integration efficiency in those studies was low and not compatible to the retrovirus, given their method was not based on the knowledge of the *in vitro* integration mechanism, or the design was not effective.

In this study, we report a method to perform highly efficient stable transgenesis. This method consists of nonviral plasmids and is highly efficient for stable transgenesis. We constructed a gateway cloning-compatible vector containing two truncated LTR sequences that flank the transgene of interest. We specifically embedded restriction cutting sites immediately adjacent to the LTR sequences so that the blunt-end DNA fragment cassette can be generated. The linearized cassette of transgene was codelivered with the overexpression construct of HIV-1 integrase to the target cells. This system removes most of the viral components, while retaining a highly efficient insertion rate of transgene into host genome, both in mammalian cells and vertebrate animal, such as zebrafish. We anticipate that this method will provide a convenient, safe, and highly efficient way for applications in transgenesis and gene therapy.

RESULTS

Construction of pLTR Vectors, Which Is Gateway Cloning Compatible with Engineered LTR Sequence Enabling Efficient Transgene Integration Mediated by HIV-1 Integrase

Based on the working mechanism of HIV-1 integrase, we hypothesized that directly providing a DNA fragment flanked with proper

engineered LTR sequences can serve as a substrate of HIV-1 integrase to form the preintegration complex (PIC), followed by 3' end process and stand transfer. This could provide a method for transgenesis, which skips the process of generation of viral particles and removes most of the viral proteins. Such a method is expected to have a highly efficient integration rate, since the mechanism is identical to native lentivirus. As a result, the vector design is shown in Figure 1A. First, the pLTR vector was modified from a gateway-compatible destination vector, pmiTo2 R4R3. We removed upstream and downstream miniTo2 sequences and inserted truncated LTR sequences immediately adjacent to the *attR4* and *attR3* motif. Labeled as 5' dLTR and 3' dLTR, these two dLTRs are truncated versions of HIV-1 LTRs, which remove most of the U3 region that contains numerous cis-control elements. We also embedded the 5'-ACTG-3' sequence to the end of dLTR, which is required for 3' end processing by integrase.^{9,16} To mimic the structure of viral DNA from reverse transcription, we included restriction cutting sites in the cassette in order to create a blunt-end DNA fragment that is also important for the activity of integrase. We have made three versions of this pLTR vector with different cutting sites of *ScaI*, *BstZ171*, and *PmeI* to prevent unintentional cleavage in the sequence of a given transgene. This type of vector is a critical component of our design, because it provides a convenient way to generate a cassette for transgenesis. For the process of transgenesis, the vector will be linearized with the restriction enzyme, and the transgene cassette can be directly delivered into the target cells. In addition, we also made expression constructs of HIV-1 integration in pCS backbone, which can be used as a template for *in vitro* transcription to produce mRNA. The synthetic HIV-1 integrase mRNA and the transgene cassette were co-injected into zebrafish-fertilized embryos at a one-cell stage, as shown in Figure 1B. For the use of this method in culture cells, the plasmid of pCS2-integrase-2A-tdTomato under the control of the cytomegalovirus (CMV) promoter and the transgene cassette was introduced into mammalian culture cells by cotransfection (Figure 1C).

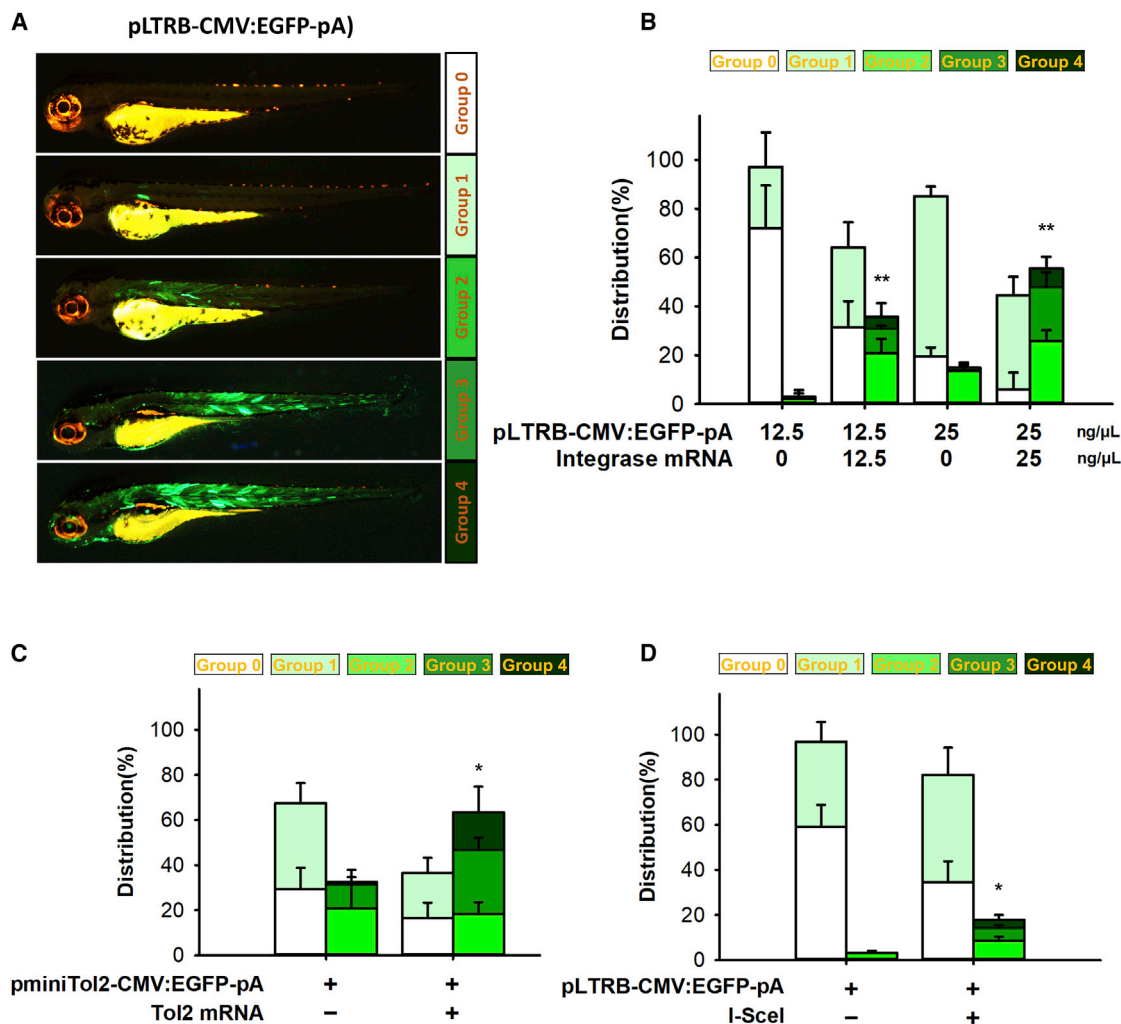


Figure 2. Highly Efficient Transgenesis Using pLTR Construct and Integrase in Zebrafish Embryos

(A) Five-category integration level of pLTR-CMV:EGFP-pA in zebrafish embryos. The images were acquired from live embryos (4 dpf) at a lateral view using a Leica M165 FC fluorescence stereomicroscope. The category is based on the coverage of the EGFP as 0%, 1%–20%, 21%–50%, 51%–80%, and 81%–100% and assigned to groups 0 to 4, respectively. (B) Percentage distribution of embryos with indicated injection dose sorted in each category. Groups 2–4 are determined as positive results of a successful genome integration with a promising percentage of transgene expression. ** $p < 0.01$ compared to the group-injected DNA only at the same concentration. (C and D) As a comparison, percentage distribution of embryos in five-category integration level of (C) pminiTol2-CMV:EGFP-pA and (D) pLTR-CMV:EGFP-pA in zebrafish embryos was shown. * $p < 0.05$ compared to the group-injected DNA only at the same concentration.

Highly Efficient, Stable Transgenesis System in Cultured Cells and Zebrafish Embryos

In order to demonstrate that our method can perform a highly efficient transgenesis, we first constructed a transgene cassette containing EGFP cDNA under the control of the CMV promoter and then tested in zebrafish embryos. As mentioned previously, the transgene cassette was coinjected with integrase mRNA into zebrafish embryos at a one-cell stage. The injected embryos were analyzed for the expression of transgene of fluorescent EGFP at 6 dpf (days postfertilization) and categorized in five different groups based on the mosaicism of the EGFP expression level (Figure 2A). We determined that groups 2–4 are positive results of successful genome integration, with a promising

percentage of transgene expression and a high potential of germline transmission in F1 fish, whereas groups 0 and 1 are negative with unsuccessful stable integration. Two injection doses have been performed (high dose, 25 ng/μL each, of mRNA and DNA or low dose of 12.5 ng/μL each), and the results were quantified in Figure 2B (Table S1). As shown, 55% and 38% of the embryos showed positive integration when injected with a high and low dose, respectively. Compared to the embryos in control experiments injected with only a transgene cassette, the HIV-1 integrase greatly facilitates the integration rate, suggesting that our method is capable of performing stable transgenesis in zebrafish with very high efficiency. It is noted that there are existing methods for making transgenic zebrafish or

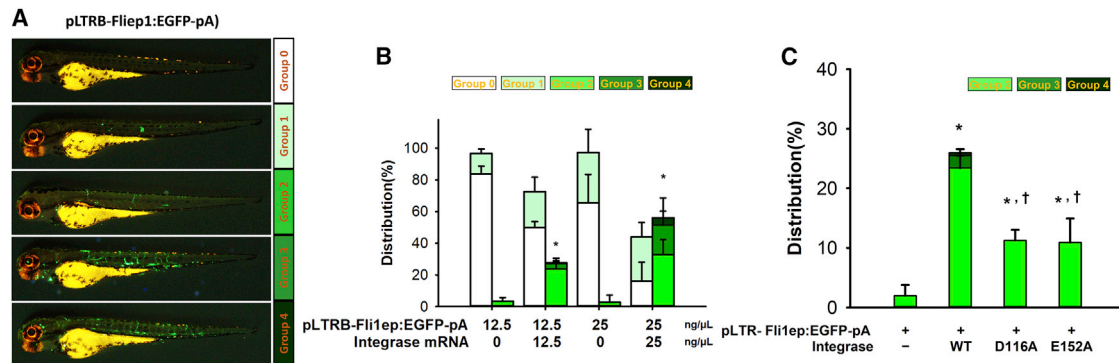


Figure 3. Tissue-Specific Licit Transgenesis Using pLTR Construct and Integrase in Zebrafish Embryos

(A) Five-category integration level of pLTR-fli1ep:EGFP-pA in zebrafish embryos. The images were acquired from live embryos (4 dpf) at a lateral view using a Leica M165 FC fluorescence stereomicroscope. The category is based on the criteria of the EGFP expression as the following: (1) group 0: “no EGFP expression,” (2) group 1: “nonspecific expression of EGFP,” (3) group 2: “EGFP expressed in 1–2 intersegmental vessels or in the heart or brain,” (4) group 3: “EGFP expressed in 3–9 intersegmental vessels and in the heart or brain vessels,” and (5) group 4: “EGFP expressed in over 10 intersegmental vessels and in the heart or brain vessels.” (B) Percentage distribution of embryos with indicated injection dose sorted in each category. Groups 2–4 are determined as positive results of successful genome integration with promising percentage of transgene expression. * $p < 0.05$ compared to the group-injected DNA only at the same concentration. (C) Percentage distribution of embryos with indicated integrase variants sorted in each category. Groups 2–4 are determined as positive results of successful genome integration with promising percentage of transgene expression. * $p < 0.05$ compared to the group-injected DNA only at the same concentration; † $p < 0.05$ compared to the group-injected wild-type (WT) integrase.

other organisms, other than using viral vectors. Those are the following: (1) Tol2-mediated and (2) meganuclease (I-SceI)-mediated transgenesis. Accordingly, we compared our system to these two methods of performing transgenesis in zebrafish. As shown in Figures 2C and 2D (Tables S2 and S3), in similar experimental conditions, Tol2 and I-SceI showed the successful rate of stable transgenesis of 62% and 20%, respectively. These results were consistent with previous reports,^{24,25} and the efficiency of our method is compatible to the Tol2 transgenic but much higher than using I-SceI. These results suggest that our method provides a new tool for stable transgenesis with similar or better efficiency than current methods. However, although the Tol2 transgenic is well established and very effective, our method has the advantages over the Tol2-mediated transgenic, since the transgene will not be mobilized, which is the major concern of the Tol2 transgenic of the mobilization of the transgene in the presence of the Tol2 transposon.

In addition, our pLTR vector is gateway compatible, which can easily include any enhancer/promoter in the p5E insert to generate a tissue-specific expression transgenic. To demonstrate this, we constructed a transgene cassette containing EGFP cDNA under the control of the fli1ep enhancer, which guides the expression of EGFP enriched in endothelial cells. Similar to the previous experiments, because of the expression mosaicism, five categories were used to determine the positive (groups 2–4) and negative (groups 0 and 1) results for successful stable transgenesis, as shown in Figure 3A. The injected embryos showed a nicely vascular expression pattern and were quantified as 57% and 27% successful rate when injected with a high and low dose, respectively (Figure 3B; Table S1). These results further demonstrated that our method can provide a spatial control for stable transgene expression if a tissue-specific enhancer/promoter is available. Moreover, to demonstrate further that the integration is licit,

we performed similar experiments with well-studied, catalytically deficient integrase mutants (D116A and E152A).^{26–28} As shown in Figure 3C (Table S4), the efficiency of integrase activity-independent integration is 10%, and notably, all were from group 2. However, these results were significantly higher than the negative control experiment (no integrase), suggesting that a catalytically deficient integrase may still facilitate the import process of DNA fragments and increase the chance for random integration during the development. Nevertheless, this result repeatedly showed that our method significantly improves the integration efficiency in a licit manner dependent on the integrase activity.

Finally, our method can also be used in culture cells, which may be useful for future cellular engineering for immunotherapy. As shown in Figure 1C, pCS2-integrase-2A-tdTomato, a plasmid containing HIV-1 integrase cDNA under the control of the CMV promoter, and the transgene cassette can be introduced into mammalian culture cells by cotransfection and execute genome integration for stable transgenesis. To demonstrate this, we cotransfected linearized pLTR-CMV:EGFP and pCS2-integrase-2A-tdTomato plasmid into two human cell lines, A549 and PANC-1, common human epithelial cancer cell lines used for transfection. After normalization to the cotransfection efficiency, the results showed that 42% of the cells expressed EGFP as a positive indication of stable transgenesis in both cell lines compared to the control group of 12% of random integration (Figure 4). This further supports our claim that our method can provide a highly efficient transgene, both in vertebrate animals and cultured mammalian cells, and that is valuable for cellular engineering and immunotherapy.

The Effect of the End Sequence of LTRs on Integration Efficiency

One important design aspect of our method is its ability to generate a blunt-end DNA fragment flanked with a specific sequence

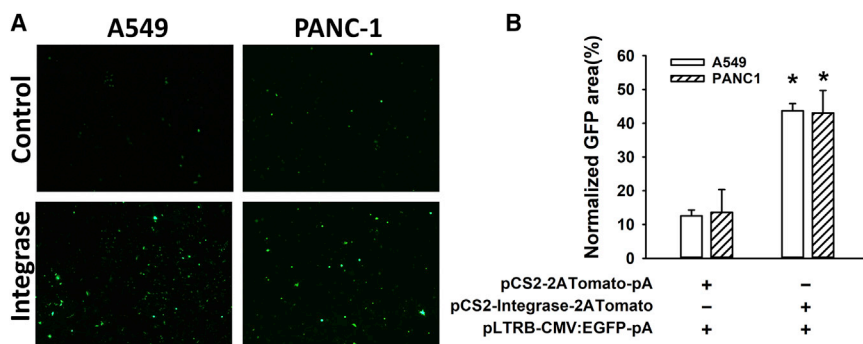


Figure 4. Stable Transgenesis Using pLTR Vector in Culture Cells

(A) A549 and PANC-1 cells were transfected with transgene cassette of pLTR-CMV:EGFP-pA and either pCS2-integrase-2A-tdTomato (indicated as integrase) or pCS2-tdTomato (indicated as control) expression plasmid. The images were acquired at three passages after transfection, and the expression of green fluorescence showed successful stable integration. (B) Normalized percentage of positive cells was analyzed by imaging process using ImageJ. * $p < 0.01$ compared to the group-injected DNA only at the same concentration.

(5'-ACTG-3') that can be identified and processed by HIV-1 integrase for strand transfer to host genome. However, in order to generate a blunt-end DNA fragment, a certain restriction enzyme would have to create extra nucleotides, which may impair the function of integrase. To verify our design, we generated three versions of the pLTR vector with different restriction cutting sites of ScaI, BstZ17I, and PmeI to test the integration efficiency. Among them, ScaI cutting will generate a perfect 5'-ACTG-3' blunt end of DNA cassette, whereas BstZ17I and PmeI digestion will create one and two extra nucleotides, respectively, as 5'-TACTG-3' and 5'-AAACTG-3'. Interestingly, there were no significant differences in the integration efficiency using these three pLTR vectors, and the one or two extra nucleotides appeared to have no effect on the process of 3' processing and strand transfer by integrase (Figures 5A and 5B; Table S1). In contrast, if we used the pLTR vector and digested it with different restricted enzymes that generate sticky ends and possess several extra nucleotides next to 5'-ACTG-3', then the integration efficiency was totally abolished to the level of the control experiment, suggesting that the integrase cannot perform 3' processing or strand transfer for the transgene. Besides the effect of the end structure of LTR, it was reported previous that only the final U5 or U3 ends, approximately 20 nt, are sufficient for integrase-mediated DNA insertion.^{20,21} Therefore, to test whether the dLTRs in our design are required and sufficient for the integration, we generated multiple transgene cassettes, flanked with or without the dLTR sequence in different lengths, as shown in Figure 5C. As shown in Figure 5D (Table S5), it is sufficient for a transgene cassette, flanked by only a 21-nt end sequence of U3 plus 5'-ACTG-3' blunt-end structure, to perform DNA insertion, seen in the C5 group, with comparable efficiency of our design in the C2 group. This observation is consistent with previous reports;^{20,21} however, the cassette flanked by such 21 nt sequences also showed a higher percentage of a nonspecific expression pattern (Figure S1), which may hamper the applications of this approach concerning the nonspecific transgene expression. It is also noted that the transgene cassette flanked by 3' dLTR only showed a significant increase for DNA insertion (C4 group), but the same increase was not observed in the construct with transgene flanked by 5' dLTR (C3 group). This may be due to the semi-specific reaction of integrase.⁹ It had higher efficiency of integration, but the nonspecific expression also increased because of the end processing of the DNA causing

random deletion of the enhancer sequence in the 5' end. As a comparison, the C3 group should have higher insertion efficiency. However, if the similar DNA end processing happens in the C3 construct from the 3' sequence, then the coding sequence of the transgene will be damaged, and the transgene expression will not be detected (Figure S1).

These results not only showed the equivalence of our three vectors but also suggested the importance of our design on the pLTR vectors that mimic the structure of the viral DNA after reverse transcription and work nicely with integrase. This also implies that without a design like ours, all other existing retroviral vectors cannot be used in this method by just cutting out the LTR-flanked gene and expecting that it will be integrated into the genome with the help of integrase. Lack of the design of the blunt-end structure of the transgene cassette and the end sequence of 5'-ACTG-3' will be the major reason of the failure of the process.

Integration Analysis of Zebrafish Embryos and Human Cell Line

There are two major questions associated with our method for stable transgenesis, which raise the concerns in gene therapy regarding the gene expression and insertional mutagenesis. (1) In the insertion site, what is the structure of the dLTR end sequence flanking the transgene cassette, and (2) in the host genome, is there any integration preference of this method? To answer these questions, integration site analysis was performed on zebrafish embryos and A549 cells that have successfully expressed the EGFP transgene using the pLTRB vector. As summarized in Table 1, Illumina sequencing found that only 8% and 1% mapped reads had an intact end pLTR structure containing 5'-TACTG-3' in zebrafish and A549 cells, respectively, and the rest of the reads showed deletions in such sequence, whereas apparently, the integrations were still successful (Figure S2). The sequencing data further showed that about 70% of the integration sites are unique, both in zebrafish and A549 cells. Interestingly, our data showed that only less than 40% of the integration sites were associated with known transcripts (genic region of the genome). Compared to previous reports, it showed around 80% of such intergration in the genic region of the genome in HIV-1-positive patients.²⁹⁻³¹ Among those reads, around 40% were mapped inside the exon that was different from the previous reports showing the range from 2% to 42%, with

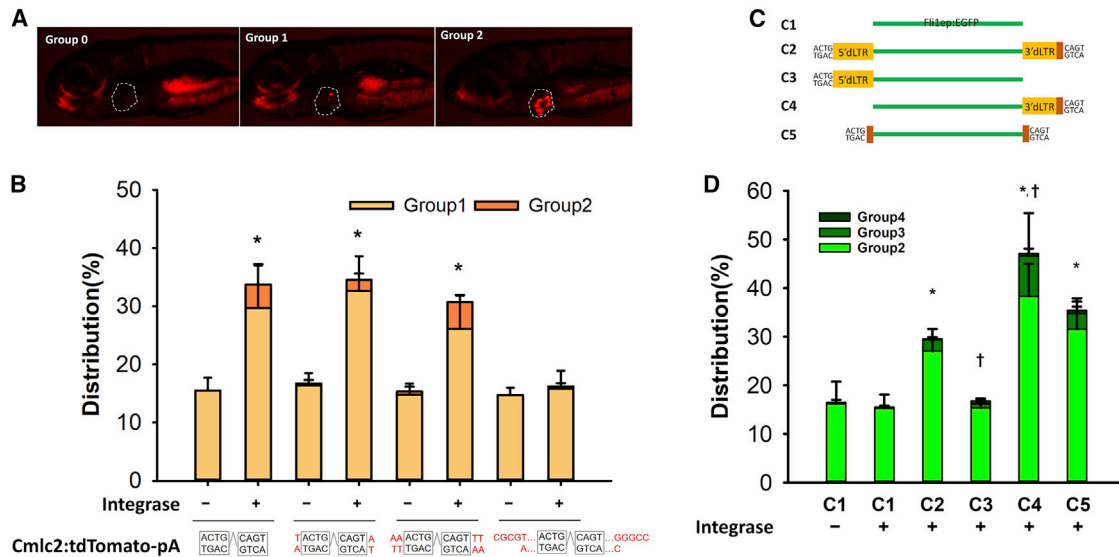


Figure 5. The Transgene Cassette Flanked by LTR Sequences with Proper End Structure Is Essential for Integrase-Mediated DNA Integration

(A) Three-category integration level of pLTR-cmlc2:tdTomato-pA in zebrafish embryos. The images were acquired from live embryos (6 dpf) at a lateral view using a Leica M165 FC fluorescence stereomicroscope. The category is based on the coverage of the red fluorescence as the following: (1) group 0: “no fluorescence,” (2) group 1: “partial fluorescence expression in the heart,” and (3) group 2: “full fluorescence expression in the heart.” The injected dose is 12.5 ng/μL each of DNA and mRNA. (B) Percentage distribution of embryos in groups 1 and 2 with the injected dose of 12.5 ng/μL DNA, with or without of mRNA (12.5 ng/μL). Groups 1 and 2 are determined as positive results of successful genome integration with a promising percentage of transgene expression. * $p < 0.05$ compared to the group-injected DNA only at the same concentration. (C) Schematic diagrams of the transgene cassettes showing different flanked LTR sequences. C1: flil1ep:EGFP; C2: flil1ep:EGFP flanked by both 5' dLTR and 3' dLTR; C3: flil1ep:EGFP flanked by only 5' dLTR; C4: flil1ep:EGFP flanked by only 3' dLTR; and C5: flil1ep:EGFP flanked by both 21 nt and 5'-ACTG-3' sequences. (D) Five-category integration level of pLTR-flil1ep:EGFP-pA in zebrafish embryos was used for quantifying the insertion rate as described in Figure 3. Embryos were injected with 12.5 ng/μL purified DNA amplicons and mRNA each. Percentage distribution of embryos with indicated injection constructs sorted in each category. Groups 2–4 are determined as positive results of successful genome integration with a promising percentage of transgene expression. * $p < 0.05$ compared to the C1 group with integrase; † $p < 0.05$ compared to the group-injected C2 group with integrase.

a median value of 7% in patients. Notably, the insertion profile in A549 cells using our method was also distinct from the results in patients. The top insertion sites from A549 cells had no overlapping with HIV-1-specific sites in patients³⁰ (Table S6). However, some of these sites were associated with epigenetic markers, such as H3K4me3 and H3K27ac, suggesting the insertion is preferred in the enhancer region (data not shown). Nevertheless, these results showed the difference between our method and HIV-1 infection, suggesting a distinct mechanism in the process of integration, while mediated by integrase. This can be due to the lack of most of the HIV-1 proteins, but the details need to be investigated further.

DISCUSSION

Here, we describe a method for highly efficient stable transgenesis. Although numerous methods have been described for gene transfer, the method we developed has unique features that will make it a useful tool in the applications of making transgenic, cellular engineering, and gene therapy. The major advantage of this method is the highly efficient integration rate of the transgene, while most of the viral components have been removed. In addition, the gateway cloning compatible vector enables an easy cloning to construct a transgene that also allows a spatial control of the transgene using a tissue-specific enhancer/promoter.

It is noteworthy that previous works attempted to improve the integration efficiency using retroviral integrase.^{22,23} Based on a knowledge of the *in vitro* integration mechanism of retrovirus, Mizuarai et al.²³ first reported that integrase alone can improve the stable transfection efficiency by adding partial LTR sequences that flank the transgene. They also reported that the length and the end structure of the LTR sequences have a significant effect on the integration rate. This is consistent with our finding that the end structure is required and critical for the integrase to perform 3' processing or strand transfer. However, as a proof of concept, this report only used PCR to add partial LTR sequences to the transgene and relied on PFU DNA polymerase to generate the blunt-end end structure of the transgene. This design greatly limited the efficiency of their method of improving the stable integration rate. Although the integration rate was significantly improved in their study compared to the transgene without the flanked partial LTR, it is still relatively low (0.73%) and not applicable to certain applications. As mentioned, the relative low efficiency could be due to the length of the LTR sequences (40 bp) or the selection of virus machinery (RSV). In contrast, our method utilizes predesigned vectors that include the restriction cutting sites to generate blunt-end structure of the transgene cassette. Such vectors are also pre-embedded with the verified truncated LTR sequences. Along with the gateway cloning, our method

Table 1. Integration Site Analysis

Sample	Structure of the Junction End Sequence (%) ^a		Unique Integration Sites (%)		Genomic Region (%)		Genomic Annotation (%)	
	Intact	Deletion	Single Site	Expanded Site	Intergenic	Genic	Intron	Exon
Zebrafish embryos	8	92	69	31	69	31	64	36
A549 cells	1	99	70	30	62	38	60	40

^aThe definition of “intact” or “deletion” for the structure of the junction end sequence is shown in Figure S2.

provides an easy way to clone desired transgene without the limitation of using PCR. Importantly, with the powerful HIV-1 machinery, our method is highly efficient for stable transgenes compared to most of the nonviral methods for transgenesis.³²

Previously, there was a similar approach using HIV-1 integrase to insert a DNA carrying two LTRs into a genome.²¹ This study nicely demonstrated that this approach can successfully insert DNA into the yeast genome, and such assays were used to show that this integration process is inhibited by RAD51. However, it was not easy to evaluate the integration efficiency on a yeast system. Our method further shows high efficiency of such approach in a *in vivo* setting and can be used to develop an assay to study the integration mechanism.

As mentioned, other than viral vectors, there are several nonviral vectors that are popular for performing transgenesis. One major category is using DNA transposon as a tool for such purpose. Several well-established systems, such as Tol2, Sleeping Beauty, and PiggyBac, have been widely used in all areas with countless applications.^{24,33–36} The features of using transposon include the following: (1) decent delivery capacity (~10 kb), (2) no dramatic immune response to repeated administration, (3) fairly random integration, and (4) inexpensive and easy to prepare. Therefore, transposon seems an ideal tool for stable transgenesis. Unfortunately, in general, DNA transposons are less efficient for gene transfer relative to viral systems. Furthermore, once integrated, the transposon insertions will be moved in the presence of transposase, which hampers the stability of the transgene for some gene therapeutic purposes and research applications. In this report, our method as a nonviral system possesses all of the features similar to DNA transposons, while showing a highly efficient integration rate, and the integrated transgene will not be moved in any circumstance. Nevertheless, the efficiency of integration can be further improved in our method. As mentioned, it was shown that concerted integration requires a tetrameric organization of integrase.¹⁶ It is similar to the working mechanism of transposase functioning as a tetramer. Within a certain range of concentration, more transposase in the cell results in an increase in transposition.³⁷ However, when transposase concentration exceeds an optimal concentration in a given cell, an inhibitory effect, known as “overproduction inhibition,” may occur, which has been described for several transposable elements.³⁸ It means that an increase in the amount of transposase results in a decrease in transpositional activity. In our method, this is also possible to occur so that the optimizations of the ratio of transgene cassette and integrase may further improve our integration efficiency.

Besides, there are still certain limitations in our current method. First, it will be more efficient to develop a one-plasmid vector system compared to our two-component systems. However, since the transgene cassette needs to be linearized before delivering into target cells, it seems unlikely to combine the transgene and integrase in one single plasmid. Second, the transgene cassette and integrase need to be actively delivered into cells, since it is unlikely for the DNA to cross a cellular membrane without any treatment. Our system can potentially be delivered into cells by any of the established nonviral delivery techniques, such as complex formation with liposomes³⁹ or polyethylenimine,⁴⁰ electroporation, or direct microinjection, but the effective integration efficiency will greatly depend on the delivery efficiency. Third, the insert size of the transgene is limited strictly in viral vectors due to the packaging capacity. Our method skips the steps of packaging, viral production, and reverse transcription inside the cells. Thus, the delivery capacity in our system should be superior than the retrovirus. However, depending on the structure of the PIC, it still has a limitation of the cargo size, which may be similar to the transposons system, considering the similar working mechanism. Finally, the transgene insertion sites using our method are still random and very different from the HIV-1-specific sites. Although we removed most of the regulatory elements in the LTR, and the integrase does not preferably insert the transgene close to certain oncogenes, it is likely that our method will induce insertional mutagenesis, given that we observed certain insertions in the exon regions. This seems inevitable, like all other methods for stable transgenesis, when performing random integration. Nevertheless, the future direction is to develop a method for site-specific integration based on our current method so that the mentioned concerns will be addressed.

In sum, we developed a method for stable transgene integration that removes most of the viral components while retaining a highly efficient insertion rate of transgene. This approach presents a potential opportunity for highly efficient transgenesis anticipation in future applications of biomedical research and gene therapy.

MATERIALS AND METHODS

Plasmid Construction

The pLTR vectors were modified from pminiTol2 R4-R3 (Addgene; #40970), which is a gateway cloning destination vector containing attR4 and attR3 gateway cassette. Briefly, the upstream and downstream miniTol2 sequences were replaced by two truncated HIV-1 LTRs, 5' dLTR (template from Addgene; #14883) and 3' dLTR (template from Addgene; #19319), respectively (Figure 1). These replacing

DNA fragments were amplified by PCR using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) with the primer-embedded 5'-ACTG-3' sequence at the 3' end and specific restriction cutting sites (ScaI, BstZ17I, or PmeI), which can generate the blunt-end fragment later (Figure 1). The pCS2-integrase and pCS2-integrase-2A-tdTomato overexpression vectors were cloned by the standard gateway cloning protocol using pCSDest2 (Addgene; #22424), p3E-2A-tdTomato (Addgene; #67707), and wild-type (WT) HIV-1 integrase in pET15b (Addgene; #61668) as a template to clone pME-integrase. The mutant constructs of integrase (D116A and E152A) were generated by PCR using pME-integrase as a template and followed by the mentioned protocol to create pCS-integrase-D116A and pCS-integrase-E152A. All generated constructs were verified by DNA sequencing at the University of California (UC) Berkeley Sequencing Facility. The p5E-CMV/SP6 and p5E-flil1ep are kindly provided by Dr. Nathan Lawson. p5E-cmlc2 is acquired from the zebrafish Tol2 kit generated by Dr. Chien Chi-Bin.⁴¹ Gateway cloning was used to construct the transgene cassette, such as pLTR-CMV:EGFP-pA, pLTRB-flil1ep:EGFP-pA, and pLTR-cmlc2:EGFP-pA. The variants of transgene cassettes with different LTR-flanked sequences were generated by PCR using pLTRB-flil1ep:EGFP-pA as a template and Phusion DNA polymerase (NEB). The amplicons were purified by gel extraction and directly used for microinjection. All plasmids were amplified in One Shot TOP10 *E. coli* cells (Invitrogen) and purified by the PureLink Quick Plasmid Miniprep Kit (Invitrogen) for subsequent microinjection, transfection, or *in vitro* transcription for making mRNA.

Zebrafish Housing and Microinjection

All zebrafish work was conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee (IACUC) at UC Merced with approved protocol. Adult zebrafish were housed in the Aquaneering zebrafish housing system at 28 °C in a 14-h light and 10 -h dark cycle. Single-pair crossing of the zebrafish generated fertilized embryos for microinjection to test genomic integration of stable transgenesis. After analysis, selected embryos were incubated in the egg water at 28 °C for up to 6 days post-fertilization (dpf) before raised in the main system. mRNA used for microinjection was generated by the mMACHINE SP6 Transcription Kit (Invitrogen), followed by one-time extraction with phenol/chloroform, and ethanol precipitated. We performed zebrafish embryos microinjection as the same methods for Tol2 and I-SceI transgenics, as described previously.^{24,25} We injected one-cell stage embryos with 25 ng/μL each of DNA and RNA or at a lower dose with 12.5 ng/μL each of DNA and RNA for 0.5 nL in each embryo.

Cell Culture and Transfection

Human lung cancer cell line A549 was acquired from ATCC (#CCL-185) and maintained in F12 medium. Human pancreatic cancer cell line PANC-1 was acquired from Sigma (#87092802) and maintained in DMEM. Both cell lines were supplied with 10% fetal bovine serum at 37°C in a humidified atmosphere of 5% CO₂/95% air in the presence of antibiotics. Transfection was followed by the Invitrogen Lipofectamine 3000 manual. Briefly, 1 day

before transfection, we seeded cells at a density of 2×10^5 cells/well in a 12-well plate. After 24 h, the cells were first rinsed with phosphate-buffered saline (PBS). Each group was transfected with 2 μg plasmids and Lipofectamine-p3000 mixture in Opti-MEM for 4 h, and then an equal amount of complete medium was added. The next day, the cells were subcultured and analyzed by flow cytometry for the initial transfection rate. The integration rate was analyzed after three passages by imaging.

Flow Cytometry Analysis

The analysis of flow cytometry was used to determine the efficiency of successful double-positive cotransfection. A portion of transfected A549 cells was harvested 24 h after transfection, whereas untransfected cells served as a negative control. Single-cell suspension of the samples was prepared by trypsinization. Enhanced Green Fluorescent Protein (EGFP) and red (tdTomato) fluorescent signals were used as an indicator for successful cotransfection of transgene and integrase, respectively. The fluorescence intensity of each sample was evaluated on a BD LSR II flow cytometer (BD Biosciences). For each analysis, at least 10,000 events were recorded, and the percentages of positive events were calculated using BD FACSDiva software (BD Biosciences).

Imaging and Processing

Live embryos were placed in 1× tricaine containing egg water. Fluorescence images were acquired using a Leica M165 FC stereo microscope. Individual embryos images were categorized in different groups based on the mosaicism of transgene expression. For cultured cells, fluorescence images were also acquired using Leica M165 FC. At least four images were taken in random locations of the dish for each experimental group. To quantify the percentage of the cells with positive EGFP expression, all images were analyzed and processed consistently using ImageJ by adjusting the threshold and counting the positive pixels. Quantified results were averaged and normalized to the transfection efficiency.

Next-Generation Sequencing and Data Analysis

Genomic DNA from EGFP-positive zebrafish embryos or A549 cells, which have been successfully inserted by pLTRB transgene cassette, were extracted using DNeasy Blood & Tissue Kits (QIAGEN; #69504). Deep-sequencing library preparation was carried out as previous described.⁴² Briefly, 10 μg of genomic DNA was digested with XhoI and MseI, and the digested DNA fragments were ligated to a linker with a 5' overhang compatible to a MseI cutting site. The viral-LTR-host genomic DNA junctions were amplified by semi-nested PCR using two sets of forward primers containing the sequences from the U5 region of LTR, paired with the reverse primer including the sequence from the linker. The amplified libraries were size selected by gel extraction and sequenced on the HiSeq platform (Illumina) to generate 150 bp paired-end reads. Reads were first cropped to remove LTR and linker sequence from genomic sequences and then aligned using Rbowtie2⁴³ to GRCz11 and hg38 for zebrafish and A549 cells, respectively. The analysis of the integration site was performed as described previously.²⁹

Statistical Analysis

All assays were carried out in triplicate or more. Data were expressed as a mean or stacked mean with standard deviation (SD). Student's t test was used to compare the mean between groups to determine statistical significance. A p value <0.05 was considered statistically significant.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtm.2020.04.020>.

AUTHOR CONTRIBUTIONS

Conceptualization, C.-W.N.; Investigation, C.-Y.C., G.D.L., and C.-W.N.; Writing - Original Draft, C.-W.N.; Writing - Review & Editing, C.-W.N. and W.-C.C.; Resources, W.-C.C.; Funding Acquisition, C.-W.N.; Supervision, C.-W.N.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work is supported by the Hellman Fellows Fund to C.-W.N.

REFERENCES

- Dunbar, C.E., High, K.A., Joung, J.K., Kohn, D.B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. *Science* 359, eaan4672.
- Nayerossadat, N., Maedeh, T., and Ali, P.A. (2012). Viral and nonviral delivery systems for gene delivery. *Adv. Biomed. Res.* 1, 27.
- Thomas, C.E., Ehrhardt, A., and Kay, M.A. (2003). Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* 4, 346–358.
- Sadelain, M. (2004). Insertional oncogenesis in gene therapy: how much of a risk? *Gene Ther.* 11, 569–573.
- Lee, E.J., Guenther, C.M., and Suh, J. (2018). Adeno-Associated Virus (AAV) Vectors: Rational Design Strategies for Capsid Engineering. *Curr. Opin. Biomed. Eng.* 7, 58–63.
- Joglekar, A.V., and Sandoval, S. (2017). Pseudotyped Lentiviral Vectors: One Vector, Many Guises. *Hum. Gene Ther. Methods* 28, 291–301.
- Elsner, C., and Bohne, J. (2017). The retroviral vector family: something for everyone. *Virus Genes* 53, 714–722.
- Poletti, V., and Mavilio, F. (2017). Interactions between Retroviruses and the Host Cell Genome. *Mol. Ther. Methods Clin. Dev.* 8, 31–41.
- Delelis, O., Carayon, K., Saïb, A., Deprez, E., and Mouscadet, J.F. (2008). Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* 5, 114.
- Li, M., and Craigie, R. (2005). Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J. Biol. Chem.* 280, 29334–29339.
- Agapkina, J., Smolov, M., Barbe, S., Zubin, E., Zatselin, T., Deprez, E., Le Bret, M., Mouscadet, J.F., and Gottikh, M. (2006). Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA. *J. Biol. Chem.* 281, 11530–11540.
- Esposito, D., and Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* 17, 5832–5843.
- Lee, S.P., Xiao, J., Knutson, J.R., Lewis, M.S., and Han, M.K. (1997). Zn²⁺ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry* 36, 173–180.
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J.F., Auclair, C., and Brochon, J.C. (2000). Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* 39, 9275–9284.
- Guiot, E., Carayon, K., Delelis, O., Simon, F., Tauc, P., Zubin, E., Gottikh, M., Mouscadet, J.F., Brochon, J.C., and Deprez, E. (2006). Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J. Biol. Chem.* 281, 22707–22719.
- Delelis, O., Carayon, K., Guiot, E., Leh, H., Tauc, P., Brochon, J.C., Mouscadet, J.F., and Deprez, E. (2008). Insight into the integrase-DNA recognition mechanism. A specific DNA-binding mode revealed by an enzymatically labeled integrase. *J. Biol. Chem.* 283, 27838–27849.
- Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S., and Parissi, V. (2005). HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.* 33, 977–986.
- Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jónsson, S.R., Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., et al. (2017). A supramolecular assembly mediates lentiviral DNA integration. *Science* 355, 93–95.
- Hare, S., Gupta, S.S., Valkov, E., Engelman, A., and Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232–236.
- Chow, S.A. (1997). In vitro assays for activities of retroviral integrase. *Methods* 12, 306–317.
- Desfarges, S., San Filippo, J., Fournier, M., Calmels, C., Caumont-Sarcos, A., Litvak, S., Sung, P., and Parissi, V. (2006). Chromosomal integration of LTR-flanked DNA in yeast expressing HIV-1 integrase: down regulation by RAD51. *Nucleic Acids Res.* 34, 6215–6224.
- Shoji-Tanaka, A., Mizuuchi, T., and Komuro, K. (1994). Gene transfer using purified retroviral integrase. *Biochem. Biophys. Res. Commun.* 203, 1756–1764.
- Mizuurai, S., Kamihira, M., Nishijima, K., and Iijima, S. (1999). Integrase-mediated nonviral gene transfection with enhanced integration efficiency. *J. Biosci. Bioeng.* 88, 461–467.
- Kawakami, K. (2007). Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* 8 (Suppl 1), S7.
- Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., and Joly, J.S. (2002). I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev.* 118, 91–98.
- Drelich, M., Wilhelm, R., and Mous, J. (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology* 188, 459–468.
- Shin, C.G., Taddeo, B., Haseltine, W.A., and Farnet, C.M. (1994). Genetic analysis of the human immunodeficiency virus type 1 integrase protein. *J. Virol.* 68, 1633–1642.
- Wiskerchen, M., and Muesing, M.A. (1995). Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J. Virol.* 69, 376–386.
- Sunshine, S., Kirchner, R., Amr, S.S., Mansur, L., Shakhbatyan, R., Kim, M., Bosque, A., Siliciano, R.F., Planelles, V., Hofmann, O., et al. (2016). HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay. *J. Virol.* 90, 4511–4519.
- Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 345, 179–183.
- Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.L., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 345, 570–573.
- Helal, N.A., Osami, A., Helmy, A., McDonald, T., Shaaban, L.A., and Nounou, M.I. (2017). Non-viral gene delivery systems: hurdles for bench-to-bedside transformation. *Pharmazie* 72, 627–693.
- Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvák, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91, 501–510.
- Zayed, H., Izsvák, Z., Khare, D., Heinemann, U., and Ivics, Z. (2003). The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. *Nucleic Acids Res.* 31, 2313–2322.

35. Wilson, M.H., Coates, C.J., and George, A.L., Jr. (2007). PiggyBac transposon-mediated gene transfer in human cells. *Mol. Ther.* *15*, 139–145.
36. Zhao, S., Jiang, E., Chen, S., Gu, Y., Shangguan, A.J., Lv, T., Luo, L., and Yu, Z. (2016). PiggyBac transposon vectors: the tools of the human gene encoding. *Transl. Lung Cancer Res.* *5*, 120–125.
37. Izsvák, Z., Ivics, Z., and Plasterk, R.H. (2000). Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* *302*, 93–102.
38. Muñoz-López, M., and García-Pérez, J.L. (2010). DNA transposons: nature and applications in genomics. *Curr. Genomics* *11*, 115–128.
39. Kumar, V.V., Singh, R.S., and Chaudhuri, A. (2003). Cationic transfection lipids in gene therapy: successes, set-backs, challenges and promises. *Curr. Med. Chem.* *10*, 1297–1306.
40. Lemkine, G.F., and Demeneix, B.A. (2001). Polyethylenimines for in vivo gene delivery. *Curr. Opin. Mol. Ther.* *3*, 178–182.
41. Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* *236*, 3088–3099.
42. Serrao, E., Cherepanov, P., and Engelman, A.N. (2016). Amplification, Next-generation Sequencing, and Genomic DNA Mapping of Retroviral Integration Sites. *J. Vis. Exp.* (109), 53840.
43. Wei, Z., Zhang, W., Fang, H., Li, Y., and Wang, X. (2018). esATAC: an easy-to-use systematic pipeline for ATAC-seq data analysis. *Bioinformatics* *34*, 2664–2665.