Title	Targeted gene integration using the combination of a sequence-specific DNA-binding protein and phiC31 integrase.
Author(s)	Nakanishi, Hideyuki; Higuchi, Yuriko; Yamashita, Fumiyoshi; Hashida, Mitsuru
Citation	Journal of biotechnology (2014), 186: 139-147
Issue Date	2014-07-17
URL	http://hdl.handle.net/2433/189465
Right	© 2014 Elsevier B.V.
Type	Journal Article
Textversion	author

Targeted gene integration using the combination of a sequence-specific DNA-binding protein and phiC31 integrase

Hideyuki Nakanishi,[1,2] Yuriko Higuchi,[3,4] Fumiyoshi Yamashita,[1] and Mitsuru Hashida[1,5]

[1]Department of Drug Delivery Research, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

[2]Japan Society for the Promotion of Science (JSPS), Sumitomo-Ichibancho FS Bldg., 8 Ichibancho, Chiyoda-ku, Tokyo 102-8472, Japan

[3]Institute for Innovative NanoBio Drug Discovery and Development, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

[4]PRESTO, Japan Science and Technology Agency (JST), Kawaguchi Center Building 4-1-8, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan

[5]Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University, 69 Yoshida Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan

Correspondence should be addressed to Mitsuru Hashida (hashidam@pharm.kyoto-u.ac.jp)

This work was done in Kyoto, Japan.

Corresponding author's address: Department of Drug Delivery Research, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

Telephone: +81-75-753-4545

Fax: +81-75-753-4575

Email: hashidam@pharm.kyoto-u.ac.jp

Highlights

· DNA binding protein (DBP) was designed to raise site-selectivity of phiC31 integrase.

· DBP has two DNA binding motif to tether the donor vector to the target sequence.

· DBP did not affect integration efficiency of phiC31 integrase in human cells.

· DBP enhanced integration site-selectivity of phiC31 integrase up to 26-fold.

· Delayed expression of integrase after that of DBP leading to higher site-selectivity.

Abstract

PhiC31 integrase-based vectors can integrate therapeutic genes selectively into *attP* or pseudo-*attP* sites in genomes, but considerable numbers of pseudo-*attP* sites in human genomes exist inside endogenous gene-coding regions. To avoid endogenous gene disruptions, we aimed to enhance the integration site-specificity of the phiC31 integrase-based vector using a sequence-specific DNA-binding protein containing Gal4 and LexA DNA-binding motifs. The dual DNA-binding protein was designed to tether the UAS-containing donor vector to the target sequence, the LexA operator, and restrict integration to sites close to the LexA operator. To analyze the site-specificity in chromosomal integration, a human cell line having LexA operators on the genome was established, and the cell line was transfected with donor vectors expressing the DNA-binding protein and the phiC31 integrase expression vector (helper vector). Quantitative PCR indicated that integration around the LexA operator was 26-fold higher with the UAS-containing donor vector than with the control. Sequence analysis confirmed that the integration occurred around the LexA operator. The dual DNA-binding protein-based targeted integration strategy developed herein would allow safer and more reliable genetic manipulations for various applications, including gene and cell therapies.

Keywords

Genomic integration; Site-specific integration; Integrase; DNA binding protein; Gene therapy

Introduction

Chromosomal integration enables sustained transgene expression and cell division-dependent replication of transgenes. Therefore, integrative vectors have often been applied to gene therapies (Calos, 2006), reprogramming (Takahashi and Yamanaka, 2006), induction of differentiation (Lacoste et al., 2009), tumor-model establishments (Carlson et al., 2005) and production of recombinant proteins (Tomita et al., 2003). PhiC31 integrase, a serine recombinase of *Streptomyces* phage, is widely used for chromosomal integration (Calos, 2006). In its natural context, phiC31 integrase integrates phage genomes into bacterial genomes by recombination between phage *attP* sites and bacterial *attB* ones (Thorpe and Smith, 1998). PhiC31 integrase can also react with pseudo-*attP* sequences that partially match with bacterial *attP* sequences, and thereby integrate plasmid vectors containing *attB* sequences (called “donor vectors”) into the genomes of many organisms, including yeasts (Thomason et al., 2001), insects (Groth et al., 2004), amphibians (Allen and Weeks, 2005), birds (Leghton et al., 2008), and mammals, in both cultured cells (Groth et al., 2000 and Thyagarajan et al., 2001) and adult animals (Olivares et al., 2002). Such *attP* or pseudo-*attP*-targeting characteristics of phiC31 integrase-based vector systems enable site-specific integration, unlike other integrative vectors including lentiviral vectors or *Sleeping Beauty* (Vink et al., 2009), *piggyBac* (Nakanishi et al., 2010 and 2011), and *Tol2* (Grabundzija et al., 2010) transposon-based vectors. Gene integration into or close to endogenous genes can disrupt or dysregulate their function, and sometimes induces severe adverse effects such as cancer (Hacein-Bey-Abina et al., 2003). In addition, random integration of sites results in a variegation of integrated transgene expression, which is called “position effect variegation” (Robertson et al., 1995). While some newer techniques, including transcription activator-like effector nucleases (TALENs) (Cermak et al., 2011 and Miller et al., 2011) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 (Cong et al., 2013 and Mali et al., 2013), are available for site-specific chromosomal integration, phiC31 integrase-based vector systems still remain useful because it allows *in vivo* gene integration in adult mammals (Olivares et al., 2002). However, phiC31 integrase-based vector systems are not sufficient in terms of site-specificity of integration, considering

that the human genome has 202-764 pseudo-*attP* sequences, about 38.7% of which are associated with endogenous genes (Chalberg et al., 2006).

In addition to the development of mutant integrases that have higher integration site-specificity (Gersbach et al., 2010 and Keravala et al., 2009), utilization of sequence-specific DNA-binding proteins might be a promising approach for enhancing the integration site-specificity of phiC31 integrase-based vectors. It could be more easily adapted to an arbitrary target sequence by customizing sequence-specific DNA-binding motifs such as zinc finger (Carroll et al., 2006, Mandell and Barbas 2006, Urnov et al., 2005, and Wright et al., 2006) or transcription activator-like effector (TALE) (Cermak et al., 2011, Miller et al., 2011, and Zhang et al., 2011). Enhanced site-specificity of integration due to fusion with DNA-binding proteins has been achieved with transposase (Ammar et al., 2012, Ivics et al., 2007, Kettlun et al., 2011, Lacoste et al., 2009, Maragathavally et al., 2006, Owens et al., 2012, Voigt et al., 2012, and Yant et al., 2007), retroviral integrase (Tan et al., 2004 and 2006), Tn3 resolvase and Gin invertase (Gordley et al., 2009). However, it has also been suggested that the coupling of phiC31 integrase with the DNA-binding protein results in loss of activity (Shinohara et al., 2007). On the other hand, Ivics *et al.* (2007) proposed a conceptually different approach whereby a dual DNA-binding protein is utilized to tether the donor vector to the binding target in chromosomes and restricts the integration to nearby sites. They demonstrated that co-transfection of cells with dual LexA/SAF-box (or LexA/TetR) DNA-binding proteins successfully enhanced the site-specificity of *Sleeping Beauty* transposase-mediated integration. We thought that this technique might be applicable to the phiC31 integrase-based vector system, since phiC31 integrase itself is in a native form.

In this context, the present study aimed to further improve the site-specificity of a phiC31 integrase-based vector system by combination with a dual DNA-binding protein. We made some modifications in the design of the donor vector as compared to the methods of Ivics *et al*. We developed a donor vector which carries both an *attP* sequence and the expression cassettes of a dual DNA-binding protein, so that we could ensure the expression of the DNA-binding protein and minimize the risk of nonspecific integration. In addition, to decrease the probability that phiC31 integrase-mediated

integration precedes the DNA-binding protein-mediated tethering, we transfected cells with the helper vector 1 or 2 days later than the donor vectors. These modifications allowed us to successfully increase the percentage of the targeted integration by 4.5-fold in the inter-plasmid integration assay and up to 26-fold in the chromosomal integration assay. The present study is the first demonstration that sequence-specific DNA-binding proteins can limit the chromosomal integration due to phiC31 integrase-based vectors to the more specific sites.

Materials and Methods

pDNA

KOD-plus ver.2 or KOD-plus Neo (Toyobo, Osaka, Japan) was used for PCRs to prepare inserts, and Rapid DNA Dephos & Ligation Kit (Roche Diagnostics, Tokyo, Japan) or Mighty Cloning Kit (blunt end) (Takara Bio, Otsu, Japan) was used for ligations. All pDNAs were amplified in the *E. coli* strains DH5α or HST08, isolated and purified using PureYield plasmid Miniprep Kit (Promega, Tokyo, Japan). For details of pDNA construction, see supplementary methods.

Cell culture

HEK293 and Hela cells were maintained in Dulbecco's modified Eagle's essential medium containing 10% fetal bovine serum.

Establishment of a Hela-attPlex4R stable cell line

Hela cells were transfected with pIR-attPlex4R and pFerH-PBTP using XtremeGene9 (Roche Diagnostics). The transfected cells were selected by antibiotic G418 (Nacalai Tesque, Kyoto, Japan) over 2 weeks from day 2 onward, and cloned. To calculate the pIR-attPlex4R-derived transposons/endogenous RNaseP gene copy number, real-time PCR was performed with genomic DNA extracted from clone cells and digested with restriction enzymes BssHII and HindIII, using a Light-Cycler instrument (Roche Diagnostics) and SYBR Premix Ex Taq (Takara Bio). The sequences

of the primer sets used to determine the copy numbers of the neomycin-resistance gene in the transposon and endogenous RNaseP gene were CGGATGGAAGCCGGTCTTGTC + AGAAGGCGATAGAAGGCGATG and AGATTTGGACCTGCGAGCG + GAGCGGCTGTCTCCACAAGT, respectively. pVITRO1-neo-RNasePfragment digested with BssHII and HindIII that contained both the neomycin-resistance gene and a fragment of the RNaseP gene was used to generate a standard curve.

Assay of *Renilla* luciferase activity

HEK293 cells were seeded onto 6- or 12-well plates. Eighteen hours later, the cells were transfected with the indicated amount of pDNA using XtremeGene9 and then lyzed using lysis buffer (0.05% Triton X-100, 2 mM EDTA, 0.1 M Tris, pH7.8) at the indicated times. The *Renilla* luciferase activity of the cell lysate was measured using a Biolux *Gaussia* luciferase assay kit (New England BioLabs Japan, Tokyo, Japan) and Lumat LB 9507 (EG & G Berthold, Bad Wildbad, Germany).

Colony-counting assay

Hela cells (2.5×10^4 cells) were seeded onto 24-well plates, and 18 h later transfected with the indicated amount of DNA using XtremeGene9. Two days after transfection, the cells were harvested, and 10% or 90% of the cells were transferred to 6-well plates and maintained in medium containing 3 μg/ml blasticidin S (Invivogen) for two weeks. To count blasticidin-resistant colonies, cells were fixed with 4% paraformaldehyde in PBS (Wako Pure Chemical Industries, Osaka, Japan) for 10 min and stained with 0.2% methylene blue (Wako Pure Chemical Industries) in PBS. The numbers of colonies were corrected by the dilution ratio.

Analysis of integration site-specificity in inter-plasmid integration

Hela cells ($6x10^5$ cells) were transfected with 1 μg of pSV40-int, 125 ng of pAttP-lex1F, 1R, 2F, 2R, 3F, 3R, 4F, 4R and 1 μg of pTargetB-NLS-Rluc or pTargetB-LexA-Rluc using XtremeGene9. Two days later, DNA was extracted from these cells using a Genelute mammalian genomic DNA

extraction kit (Sigma-Aldrich Japan, Tokyo, Japan). The extracted DNA was used to transform the *E. coli* strain DH5α or HST08. pDNA was purified from *E. coli* resistant to both kanamycin and blasticidin using a PureYield plasmid Miniprep Kit (Promega). Purified pDNA was digested with restriction enzymes either BamHI, NdeI, SpeI, BstZ17I plus NdeI. These digested pDNAs were electrophoresed on 1% agarose S (Nippon Gene, Tokyo, Japan) gel to determine which recipient vectors were integrated with donor vectors.

Analysis of targeted chromosomal integration

Hela-attPlex4R cells (5×10^4 cells) were seeded onto 6-well plates, and 18 hr later transfected with 500 ng of the donor vectors using XtremeGene9. One or two days after transfection of the donor vectors, the cells were transfected with 500 ng of pCMV-int. From three days after transfection of the donor vectors, the cells were maintained in medium containing 3 μg/ml blasticidin S for 18 days. After blasticidin selection, DNA was extracted from these cells using a Genelute mammalian genomic DNA extraction kit. For quantitative analysis of the *attR* and blasticidin-resistance gene copy numbers, real-time PCR was performed with extracted DNA using a Light-Cycler instrument (Roche Diagnostics) and SYBR Premix Ex Taq (Takara Bio). The sequences of the primer sets used to determine the copy numbers of the *attR* and blasticidin-resistance gene were tcgagGCATCAAGCTAATTC + AGTACGCCCCCTATTGACG and gaagaccttcaacatctctcagc + atcttctcagtggcgacctc, respectively. A targeted integration product containing both the *attR* and blasticidin-resistance genes was obtained by the plasmid rescue method, and used to generate a standard curve.

Analysis of integration sites by plasmid rescue

To analyze phiC31 integrase-mediated integration sites, Hela-attPlex4R cells (5×10^4 cells) were transfected with 500 ng of pTargetB2(4x)-LexA-Rluc. Two days later, the cells were transfected with 500 ng of pCMV-int. XtremeGene9 was used for both transfections. From the next day, cells were cultured in medium containing 3 μg/ml blasticidin S for 18 days. DNA was isolated from these cells using a Genelute mammalian genomic DNA extraction kit, and digested using restriction enzyme NheI,

SpeI, and XbaI. After digestion by these restriction enzymes, the DNA was purified using a Genelute PCR Clean-up Kit (Sigma-Aldrich Japan) and ligated using a Rapid DNA Dephos & Ligation Kit (Roche Diagnostics) or Ligation convenience kit (Nippon Gene). The ligation products were used to transform *E. coli* Strain DH5α or HST08. pDNA was purified from blasticidin-resistant *E coli* using a PureYield plasmid Miniprep Kit. The nucleotide sequences of the pDNA were sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies Japan, Tokyo, Japan) and ABI3100xl (Life Technologies Japan).

Results

Vector design to analyze inter-plasmid integration

The phiC31 integrase-based vector system consists of two vector types. One is a donor vector that contains the *attB* sequence, and the other is a helper vector that expresses phiC31 integrase. PhiC31 integrase expressed by helper vectors integrates donor vectors into chromosomes.

To enhance the site-specificity of phiC31 integrase-mediated integration, we designed the expression cassette of a dual DNA-binding protein which tethers the donor vector at the target sequence. The dual DNA-binding protein contains a Gal4 DNA-binding domain and LexA DNA-binding domain, which recognize an upstream activation sequence (UAS) and a LexA operator sequence, respectively. Here, the LexA operator sequence was assumed to be a model of binding targets in chromosomes. Tethering of the donor vector by the dual DNA-binding protein limits phiC31 integrase-mediated integration to the *attP* sites close to its target sequence (**Fig. 1**). pTargetB-LexA-Rluc was constructed as a donor vector, in which both the expression cassettes of the DNA-binding protein and 4 copies of UAS were inserted in addition to the *Renilla* luciferase gene. The donor vector named pTargetB-NLS-Rluc was also constructed as a negative control. pTargetB-NLS-Rluc was the same as pTargetB-LexA-Rluc, except that it expressed a DNA-binding protein lacking the LexA DNA-binding domain. pAttB-UAS-Rluc was another negative control donor vector, which had no expression cassette of DNA-binding protein. On the other hand, pCMV-int and pSV40-int were

constructed as helper vectors that express phiC31 integrase. To investigate whether the integration of pTargetB-LexA-Rluc is dependent on the distance between the target sequence for the DNA-binding protein and *attP* sequence, eight recipient vectors (i.e., pAttP-lex1~4F and 1~4R) having different *attP*-sequence orientations and different distances between the LexA operator and *attP* sequences were constructed (**Fig. 2**).

Effect of the DNA-binding protein on the integration efficiency

It was preliminarily investigated whether binding of the DNA-binding protein to the donor vector prevents phiC31 integrase from interacting with the donor vector. Following transfection of HEK293 cells with pTargetB-NLS-Rluc or pAttB-UAS-Rluc and with or without pCMV-int, sustainability of the gene expression of Rluc was evaluated as an indicator of integrations. When the effect of co-transfection with pCMV-int on the Rluc expression was evaluated on day 12, the sustained gene expression was enhanced by 2.32-fold and 1.45-fold in pTargetB-NLS-Rluc-transfected and pAttB-UAS-Rluc-transfected cells, respectively (**Fig. 3**). It should be remembered that the DNA-binding protein could be expressed and associated with the donor vector in pTargetB-NLS-Rluc, but this was not the case in pAttB-UAS-Rluc. Nevertheless, the enhancement of sustained gene expression by phiC31 integrase was not lower in pTargetB-NLS-Rluc. This suggests that phiC31 integrase-mediated integration was not inhibited even though the DNA-binding protein bound to the donor vectors.

Integration site-specificity in inter-plasmid integration

The effects of the DNA-binding protein on the integration site-specificity were investigated in an inter-plasmid integration assay. To avoid phiC31 integrase-mediated inter-plasmid integration during the amplification process in *E. coli* (Thorpe and Smith, 1998), pSV40-int was used as a helper vector instead of pCMV-int (**Fig. 2**). pSV40-int was chosen because, unlike the CMV promoter, the SV40 promoter does not express proteins in *E. coli* (Goussard et al., 2003). A preliminary experiment indicated that pSV40-int cannot integrate donor vectors in *E. coli* but can do so in Hela cells (data not

shown).

First, Hela cells were transfected simultaneously with pSV40-int as a helper vector, pTargeB-NLS-Rluc or pTargetB-LexA-Rluc as donor vectors, and pAttP-lex1~4F and 1~4R as recipient vectors. Two days later, DNA was extracted from these cells and transformed to *E. coli*. Taking into account that the inter-plasmid integration products contain blasticidin- and kanamycin-resistance genes originated from donor and recipient vectors, respectively, transformed *E. coli* was selected by both blasticidin and kanamycin. The inter-plasmid integration products extracted from *E. coli* were digested with restriction enzymes and subjected to gel electrophoresis to determine to what extent each recipient vector was integrated. **Figure 4a and b** shows the percentages of integration into each recipient vector per total integration, and the ratio of the pTargetB-LexA-Rluc transfected group per the pTargetB-NLS-Rluc group, respectively. pTargetB-LexA-Rluc was designed to express a DNA-binding protein (i.e., the Gal4-LexA DNA-binding protein) that could bind to both the donor and recipient vectors, whereas pTargetB-NLS-Rluc expressed a protein that bound only to the donor vector. As shown in **Fig. 4b**, the percentage of integration into pAttP-lex4R, which has the smallest LexA-*attP* distance of 219 bp, was increased 4.5-fold in the pTargetB-LexA-Rluc group.

Cell line establishment and vector construction for the analysis of chromosomal integration

Prior to the evaluation of Gal4-LexA DNA-binding protein-enhanced integration site-specificity in chromosomes, we established a stable cell line designated Hela-attPlex4R, in which both the *attP* and LexA operator sequences were chromosomally integrated. pAttP-lex4R was selected for the integration to establish Hela-attPlex4R, based on the results of the inter-plasmid integration assay (**Fig. 4**). At first, pAttP-lex4R was linearized and inserted into *piggyBac* transposon (hereafter pIR-attPlex4R). Then, pIR-attPlex4R together with the *piggyBac* transposase expression vector pFerH-PBTP were transfected into Hela cells to integrate the transposon containing the pAttP-lex4R-derived sequence into chromosomes (**Fig. 5**). Following selection, a stably integrated HeLa-attPlex4R cell clone was obtained and subjected to quantitative evaluation of chromosomal integration. Integration of attPlex4R sequences into the genome was evaluated with primers for neomycin-resistance gene in

the transposon, in reference with endogenous RNaseP gene. Real-time PCR analysis following extraction and digestion of genomic DNA revealed that the number of integrated copies was 15 copies per haploid genome in the Hela-attPlex4R clone.

The Hela-attPlex4R thus established was transfected with pTargetB-NLS-Rluc or pTargetB-LexA-Rluc with pCMV-int. Unfortunately, transfection with these vectors produced very few blasticidin-resistant cells (data not shown), despite the fact that the donor vectors contained the blasticidin-resistance gene. We redesigned the donor vector constructs so that they contained the *attB* sequence at a different position (**Fig. 6a**), taking into account a previous report that the location of the *attB* sequences in the donor vectors affects sustained gene expression by the phiC31 integrase-based vector system (Watanabe et al., 2011). To minimize the effect on blasticidin-resistance gene expression, the *attB* sequence was placed apart from the blasticidin-resistance gene in the newly developed donor vector construct pTargetB2-NLS-Rluc (**Fig. 6a**). As shown in **Fig. 6b and c**, transfection with pTargetB2-NLS-Rluc and pCMV-int provided significantly more blasticidin-resistant colonies than that with pTargetB-NLS-Rluc and pCMV-int.

Considering that displacement of the *attB* sequence was effective for increased expression of the blasticidin-resistance gene, new donor vectors that expressed the Gal4-LexA DNA-binding protein were constructed and designated pTargetB2(0x)-LexA-Rluc, pTargetB2(4x)-LexA-Rluc, pTargetB2(18x)-LexA-Rluc, and pTargetB2(66x)-LexA-Rluc according to the number of UAS repeats (**Fig. 6a**). We expected that binding between the donor vectors and the Gal4-LexA DNA-binding protein would become more likely as the number of UAS increases.

Integration site-specificity in chromosomal integration

To evaluate the integration site-specificity in chromosomal integration, Hela-attPlex4R cells were transfected with pCMV-int and either pTargetB2-NLS-Rluc, pTargetB2(0x)-LexA-Rluc, pTargetB2(4x)-LexA-Rluc, pTargetB2(18x)-LexA-Rluc, or pTargetB2(66x)-LexA-Rluc. To allow the DNA-binding protein to be expressed and tether the donor vectors in advance, the cells were transfected with the donor vectors 1 or 2 days prior to transfection with pCMV-int. After transfections

of pCMV-int, the cells were subjected to blasticidin selection and extraction of genomic DNA. Quantitative PCR of genomic DNA was performed using forward and reverse primers designed to anneal to a sequence neighboring the LexA operator and a sequence inside a donor vector, respectively. These primers allowed us to count the copies of the donor vector integrated into the *attP* close to the LexA operator sequence (**Fig. 7a**). The copy number of total donor vectors was also evaluated as an internal control. As for the 1-day delay of pCMV-int transfection, the targeted integration efficiency of pTargetB2(4 ~ 18x)-LexA-Rluc was minimally higher than that of the negative controls (pTargetB2-NLS-Rluc and pTargetB2(0x)-LexA-Rluc) (**Fig. 7b**). As for the 2 day delay of pCMV-int transfection, the targeted integration efficiency of pTargetB2(4~18x)-LexA-Rluc was remarkably (up to 26-fold) higher than that of the negative controls (**Fig. 7c**). Unexpectedly, the targeted integration efficiencies of pTargetB2(66x)-LexA-Rluc were lower than that of pTargetB2(4~18x)-LexA-Rluc under both transfection conditions.

Confirmation of targeted integration by sequencing analysis

To confirm integration into the *attP* close to the LexA operator sequence, a plasmid rescue method was adopted. First, Hela-attPlex4R cells were transfected with pTargetB2(4x)-LexA-Rluc and, 2 days later, with pCMV-int, and then selected by treatment with blasticidin. Secondly, genomic DNA was extracted from the cells, digested with restriction enzymes, ligated, and used for *E. coli* transformation. As pTargetB2(4x)-LexA-Rluc contains a blasticidin-resistance gene and *E. coli* replication origin, fragments of genomic DNA containing an insertion of pTargetB2(4x)-LexA-Rluc can be replicated in the presence of blasticidin. Of the obtained 2 clones of phiC31 integrase-mediated integration products, one is the targeted integration product (**Table 1**). Surprisingly, the other is integrated into pseudo-*attP* in pCMV-int.

Discussion

The present study was designed to enhance the integration site-specificity of phiC31

integrase-based vectors by tethering the donor vector to chromosomes with a target sequence-specific dual DNA-binding protein. Although the strategic concept has already been applied to *Sleeping Beauty*-based vectors (Ivics et al., 2007), we made some further improvements to adapt the method for phiC31 integrase-mediated integration. First, we incorporated an expression cassette for a dual DNA-binding protein in the donor vectors, instead of constructing donor vectors and the DNA-binding protein expression vectors separately. This guarantees the expression of the DNA-binding protein in cells transfected with the donor vectors. Secondly, we inserted multiple (4~66x) binding sequences into the donor vectors to increase the probability of binding the DNA-binding protein to the donor vectors. Thirdly, we transfected cells with the donor vectors in advance of transfection with the helper vector in order to ensure expression of the DNA-binding protein and tethering of the donor vectors at the target sequence of chromosomes prior to phiC31 integrase expression.

Under the modified conditions, the donor vectors containing dual DNA-binding protein provided a higher integration site-specificity than their negative controls in both the inter-plasmid integration assay (**Fig. 4b**) and chromosomal integration assay (**Fig. 7c**). In previous studies using sequence-specific DNA-binding motifs, the fold increases of targeted chromosomal integration ranged from 2.0 to 9.9 (Ivics et al., 2007, Kettlun et al., 2011, Owens et al., 2012, Tan et al., 2006, Voigt et al., 2012). Even though our data cannot simply be compared with the previous data because of the differences in evaluation and analysis methods, a 26-fold increase in targeted chromosomal integration (**Fig. 7c**) would be considerably large. However, the percentage of the donor vector integrated into native *attP* (not pseudo-*attP*) neighboring the target sequence was not necessarily high (0.38% of tonal donor vectors; **Fig. 7b**). One reason for the low rate of targeted integration might be associated with the context of *attP*-surrounding sequence (Calos, 2006). The inter-plasmid integration assay showed that the integration efficiency of pTargetB-NLS-Rluc into pAttP-lex4R was lowest among all recipient vectors (**Fig. 4a**). This suggests that the context of pAttP-lex4R-derived surrounding sequences might not be suitable for phiC31 integrase-mediated integration. Therefore, if we select *attP* or pseudo-*attP* sequences with more suitable surrounding contexts, the targeted integration percentage may increase.

In Hela-attPlex4R that was used for chromosomal integration assay, multiple target *attP*

sequences were inserted into the genome (15 copies per haploid genome). However, all the target *attP* sites might not necessarily be available for transgene expression due to their surrounding chromosomal contexts and epigenetic modifications. Selection of blasticidin-resistant cells could lead to underestimation of the degree of site-specific chromosomal integration of pTargetB2-LexA-Rluc, since it rules out integration to the target *attP* sites that do not allow the expression of the resistance marker. Taking together with the existence of pseudo-*attP* sites in some regions of the genome such as heterochromatin, it would be difficult to determine an exact efficiency count of targeted chromosomal integration. Thus, it should be noted that the present assay method of targeted integration simply allows relative comparison among the vector systems.

In the chromosomal integration assay, delayed transfection of the helper vector after that of the donor vector transfections increased the difference between the pTargetB2(4~18x)-LexA-Rluc) and the negative controls (pTargetB2-NLS-Rluc and pTargetB2(0x)-LexA-Rluc) (F**ig. 7b, c**). These results suggest that the time-lag transfection provides an opportunity for the DNA-binding proteins to tether the donor vectors to the target sequences before phiC31 integrase-mediated integration. In addition to time-lag transfection, the use of chemical-regulatable gene expression systems is another option for producing an expression time-lag (Sharma et al., 2008 and Yen et al., 2004).

Because of the simplicity of evaluation, the LexA operator sequence was selected as a target sequence of the DNA-binding protein by inserting it into chromosomes exogenously. However, when this targeted integration strategy is intended for practical applications such as gene therapies, the target sequences should be selected from native genomic sequences. To achieve safer chromosomal integration, pseudo-*attP* sequences in genomic safe harbors (Sadelain et al., 2012) should be selected as integration sites, and a specific DNA-binding motif should be customized so that it binds near the pseudo-*attP* sequences. In addition, it has been established that pseudo-*attP* sequences that can be recognized as substrates for phiC31 integrase exhibit certain variations in chromosomal context among cell types (Calos, 2006). Selection of different target sequences might thus be required depending on the cell types.

The present targeted integration approach is theoretically applicable to other integrative

vector systems. Ivics *et al.* (2007) have shown that a LexA-SAFbox DNA-binding protein did not inhibit *Sleeping Beauty*-mediated integration, and during the preparation of this manuscript, Owens *et al.* (2013) reported targeted *piggyBac* integration by tethering of the donor vectors. As experienced with problem associated with the *attB* site (**Fig. 6**), the design of donor vectors appears to be important in targeted integration. As long as the vectors are carefully designed, the present targeted integration approach would be useful for other integrative vector systems, including *Sleeping Beauty* and *piggyBac*.

In conclusion, we demonstrated by using phiC31 integrase-based integration systems that a multi-functional donor vector which expresses a sequence-specific dual DNA-binding protein that tethers itself to the target sequence and which has multiple sequences for binding of the protein works cooperatively with a helper vector when cells are subjected to both vector transfections separated by an appropriate time lag. We believe that the present study provides important information toward the achievement of site-specific transgene integration, and paves the way for more reliable gene therapies and genetic studies.

Acknowledgements

We thank Hajime Mori (Kyoto Institute of Technology, Kyoto, Japan) and Takahiro Kusakabe (Kyushu University, Fukuoka, Japan) for p3E1.2 (a plasmid DNA containing the *piggyBac* original fragment) and pZerofd31attB3xP3EGFP (a plasmid DNA containing *attB* sequence), respectively. This work was supported by the Japan Society for the Promotion of Science (KAKENHI Grant Number 11J01309), Grants-in-Aid for Scientific Research (Number 23659284 and 25670259) from the Ministry of Education, Culture, Sports, and Science, and the PRESTO program (Number 3424) of the Japan Science and Technology Agency. None of the authors have any conflicts of interest associated with this study.

References

Allen, B.G., Weeks, D.L., 2005. Transgenic Xenopus laevis embryos can be generated using ΦC31 integrase. Nat. Methods 2, 975–979.

Ammar, I., Gogol-Döring, A., Miskey, C., Chen, W., Cathomen, T., Izsvák, Z., Ivics, Z., 2012. Retargeting transposon insertions by the adeno-associated virus Rep protein. Nucleic Acids Res. 40, 6693–712.

Calos, M.P., 2006. The ΦC31 Integrase System for Gene Therapy. Curr. Gene Ther. 6, 633–645.

Carlson, C.M., Frandsen, J.L., Kirchhof, N., Mcivor, R.S., Largaespada, D.A., 2005. Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. Proc. Natl. Acad. Sci. Unites States Am. 102, 17059–17064.

Carroll, D., Morton, J.J., Beumer, K.J., Segal, D.J., 2006. Design, construction and in vitro testing of zinc finger nucleases. Nat. Protoc. 1, 1329–1341.

Cermak, T., Doyle, E.L., Christian, M., et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39, e82.

Chalberg, T.W., Portlock, J.L., Olivares, E.C., et al., 2006. Integration specificity of phage phiC31 integrase in the human genome. J. Mol. Biol. 357, 28–48.

Cong, L., Ran, F.A., Cox, D., et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–23.

Gersbach, C.A., Gaj, T., Gordley, R.M., Barbas, C.F., 2010. Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Res. 38, 4198–206.

Gordley, R.M., Gersbach, C.A., Barbas, C.F., 2009. Synthesis of programmable integrases. Proc. Natl. Acad. Sci. U. S. A. 106, 5053–5058.

Goussard, S., Grillot-Courvalin, C., Courvalin, P., 2003. Eukaryotic promoters can direct protein synthesis in Gram-negative bacteria. J. Mol. Microbiol. Biotechnol. 6, 211–218.

Grabundzija, I., Irgang, M., Mátés, L., et al., 2010. Comparative analysis of transposable element vector systems in human cells. Mol. Ther. 18, 1200–9.

Groth, A.C., Fish, M., Nusse, R., Calos, M.P., 2004. Construction of Transgenic Drosophila by Using

the Site-Specific Integrase From Phage ΦC31. Genetics 166, 1775–1782.

Groth, A.C., Olivares, E.C., Thyagarajan, B., Calos, M.P., 2000. A phage integrase directs efficient site-specific integration in human cells. Proc. Natl. Acad. Sci. U. S. A. 97, 5995–6000.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., et al., 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302, 415–419.

Ivics, Z., Katzer, A., Stüwe, E.E., Fiedler, D., Knespel, S., Izsvák, Z., 2007. Targeted Sleeping Beauty transposition in human cells. Mol. Ther. 15, 1137–1144.

Keravala, A., Lee, S., Thyagarajan, B., Olivares, E.C., Gabrovsky, V.E., Woodard, L.E., Calos, M.P., 2009. Mutational Derivatives of PhiC31 Integrase With Increased Efficiency and Specificity. Mol. Ther. 17, 112–120.

Kettlun, C., Galvan, D.L., George, A.L., Kaja, A., Wilson, M.H., 2011. Manipulating piggyBac transposon chromosomal integration site selection in human cells. Mol. Ther. 19, 1636–44.

Lacoste, A., Berenshteyn, F., Brivanlou, A.H., 2009. An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic stem cells. Cell Stem Cell 5, 332–42.

Leighton, P. a, van de Lavoir, M.-C., Diamond, J.H., Xia, C., Etches, R.J., 2008. Genetic modification of primordial germ cells by gene trapping, gene targeting, and phiC31 integrase. Mol. Reprod. Dev. 75, 1163–75.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-guided human genome engineering via Cas9. Science 339, 823–6.

Mandell, J.G., Barbas, C.F., 2006. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res. 34, W516–W523.

Maragathavally, K.J., Kaminski, J.M., Coates, C.J., 2006. Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J. 20, 1880–2.

Miller, J.C., Tan, S., Qiao, G., et al., 2011. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143–8.

Nakanishi, H., Higuchi, Y., Kawakami, S., Yamashita, F., Hashida M., 2010. piggyBac Transposon-

mediated Long-term Gene Expression in Mice. Mol. Ther. 18, 707–714.

Nakanishi, H., Higuchi, Y., Kawakami, S., Yamashita, F., Hashida, M., 2011. Comparison of piggyBac transposition efficiency between linear and circular donor vectors in mammalian cells. J. Biotechnol. 154, 205–208.

Olivares, E.C., Hollis, R.P., Chalberg, T.W., Meuse, L., Kay, M.A., Calos, M.P., 2002. Site-specific genomic integration produces therapeutic Factor IX levels in mice. Nat. Biotechnol. 20, 1124–1128.

Owens, J.B., Mauro, D., Stoytchev, I., Bhakta, M.S., Kim, M.-S., Segal, D.J., Moisyadi, S., 2013. Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. Nucleic Acids Res. 41, 9197-9207.

Owens, J.B., Urschitz, J., Stoytchev, I., et al., 2012. Chimeric piggyBac transposases for genomic targeting in human cells. Nucleic Acids Res. 40, 6978–6991.

Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D., Whitelaw, E., 1995. Position-dependent variegation of globin transgene expression in mice. Proc. Natl. Acad. Sci. U. S. A. 92, 5371–5.

Sadelain, M., Papapetrou, E.P., Bushman, F.D., 2012. Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer 12, 51–58.

Sharma, N., Moldt, B., Dalsgaard, T., Jensen, T.G., Mikkelsen, J.G., 2008. Regulated gene insertion by steroid-induced PhiC31 integrase. Nucleic Acids Res. 36, e67.

Shinohara, E.T., Kaminski, J.M., Segal, D.J., et al., 2007. Active integration: new strategies for transgenesis. Transgenic Res. 16, 333–9.

Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Tan, W., Dong, Z., Wilkinson, T.A., Barbas, C.F.I., Chow, S.A., 2006. Human Immunodeficiency Virus Type 1 Incorporated with Fusion Proteins Consisting of Integrase and the Designed Polydactyl Zinc Finger Protein E2C Can Bias Integration of Viral DNA into a Predetermined Chromosomal Region in Human Cells. J. Virol. 80, 1939–1948.

Tan, W., Zhu, K., Segal, D.J., Barbas, C.F.I., Chow, S.A., 2004. Fusion Proteins Consisting of Human Immunodeficiency Virus Type 1 Integrase and the Designed Polydactyl Zinc Finger Protein E2C Direct Integration of Viral DNA into Specific Sites. J. Virol. 78, 1301–1313.

Thomason, L.C., Calendar, R., Ow, D.W., 2001. Gene insertion and replacement in Schizosaccharomyces pombe mediated by the Streptomyces bacteriophage φC31 site-specific recombination system. Mol. Genet. Genomics 265, 1031–1038.

Thorpe, H.M., Smith, M.C., 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc. Natl. Acad. Sci. U. S. A. 95, 5505–10.

Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., Calos, M.P., 2001. Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage ΦC31 Integrase. Mol. Cell. Biol. 21, 3926–3934.

Tomita, M., Munetsuna, H., Sato, T., et al., 2003. Transgenic silkworms produce recombinant human type III procollagen in cocoons. Nat. Biotechnol. 21, 52–6.

Urnov, F.D., Miller, J.C., Lee, Y.-L., et al., 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435, 646–651.

Vink, C.A., Gaspar, H.B., Gabriel, R., Schmidt, M., Mcivor, R.S., Thrasher, A.J., Qasim, W., 2009. Sleeping Beauty Transposition From Nonintegrating Lentivirus. Mol. Ther. 17, 1197–1204.

Voigt, K., Gogol-Döring, A., Miskey, C., Chen, W., Cathomen, T., Izsvák, Z., Ivics, Z., 2012. Retargeting Sleeping Beauty Transposon Insertions by Engineered Zinc Finger DNA-binding Domains. Mol. Ther. 20, 1852–1862.

Watanabe, S., Nakamura, S., Sakurai, T., Akasaka, K., Sato, M., 2011. Improvement of a phiC31 integrase-based gene delivery system that confers high and continuous transgene expression. N. Biotechnol. 28, 312–9.

Wright, D.A., Thibodeau-Beganny, S., Sander, J.D., et al., 2006. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat. Protoc. 1, 1637–1652.

Yant, S.R., Huang, Y., Akache, B., Kay, M.A., 2007. Site-directed transposon integration in human

cells. Nucleic Acids Res. 35, e50.

Yen, L., Svendsen, J., Lee, J., Gray, J.T., 2004. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. Nature 431, 471–476.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., Arlotta, P., 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. 29, 149–154.

Fig. 1 Schematic diagram of targeted integration using a sequence-specific DNA-binding protein and phiC31 integrase

Gal4 and LexA DNA-binding domains of the Gal4-NLS-LexA DNA-binding protein bind to UAS and the LexA operator (the target sequence), respectively. Then, phiC31 integrase integrates the donor vector by recombining the *attB* and *attP* sequences. Because the donor vector is tethered to the target sequence, its integration site is restricted to the vicinity of the target sequence.

Fig. 2 Vectors for targeted integration

Vector construction. Amp^R, ampicillin-resistance gene; *attB*, *attB* sequence for phiC31 integrase-mediated recombination; *attP*, *attP* sequence for phiC31integrase-mediated recombination; $Blast^R$, blasticidin-resistance gene; CMV, cytomegalovirus promoter; EM7, bacterial EM7 promoter; Gal4, Gal4 DNA-binding domain; hEF1, human elongation factor 1α promoter; LexA, LexA DNA-binding domain; NLS, nuclear localization signal; Rluc, *Renilla* luciferase gene; Neo^R, neomycin-resistance gene; SV40, simian virus 40 promoter; ori, *E. coli* origin of replication; 4×UAS, four copies of upstream activating sequences; Kan/Neo^R, kanamycin/neomycin-resistance gene; LexA operator, LexA operator sequence; SV40/BP, simian virus 40 promoter/Bacterial promoter. Each recipient vector (pAttP-lex1~4F and 1~4R) has one copy of *attP* at 6.5-6.6 (1F, 1R), 1.7-1.8 (2F, 2R), 0.6-0.7 (3F, 3R), or 0.2-0.3 (4F, 4R) kb upstream, respectively, from LexA operator start point as indicated orientations.

Fig. 3 Effect of DNA-binding proteins on phiC31 integrase-mediated sustained gene expression

HEK293-lexluc cells ($2x10^5$ cells/well) were transfected with 500 ng of the indicated donor vector and 500 ng of pCMV-int or pcDNA3.1(+). The bars show the results for pAttB-UAS-Rluc and pcDNA3.1(+) (white), pAttB-UAS-Rluc and pCMV-int (hatched), pTargetB-NLS-Rluc and pcDNA3.1(+) (dotted), pTargetB-NLS-Rluc and pCMV-int (black), respectively. *Renilla* luciferase activities were measured at the indicated time points after transfection. Each value represents the mean + SD (n = 3).

Fig. 4 Effect of the DNA-binding protein on the integration site selection of phiC31 integrase in inter-plasmid integration

Hela cells were transfected with pAttP-lex1F~4R (125 ng each), 1 μg pSV40-int (1 μg), and pTargetB-LexA-Rluc or pTargetB-NLS-Rluc (1 μg). DNA extracted from Hela cells was subjected to *E. coli* transformation. Recombination products amplified in *E. coli* were analyzed by restriction digestion and agarose gel electrophoresis. Figure 4(a) shows the percentages of each inter-plasmid integration product per total inter-plasmid integration products. Open rectangles indicate inter-plasmid integration products of pTargetB-LexA-Rluc and each recipient vector. Closed squares indicate inter-plasmid integration products of pTargetB-NLS-Rluc and each recipient vector. Figure 4(b) represents the ratios of pTargetB-LexA-Rluc to pTargetB-NLS-Rluc in percentages of each inter-plasmid integration product. The abscissa indicates the distance between the LexA operator and *attP* sequences in each recipient vector.

Fig. 5 Establishment of a stable cell line containing LexA operator and *attP* sequences in chromosomes

To analyze integration site-specificity in a chromosomal context, LexA operator and *attP* sequences were integrated into the chromosomes of Hela cells by *piggyBac* transposase-mediated integration. Amp^R, ampicillin-resistance gene; *attP*, *attP* sequence for phiC31integrase-mediated recombination; Kan/Neo^R, kanamycin/neomycin-resistance gene; IR, *piggyBac* inverted repeat sequence for *piggyBac* transposase-mediated integration; LexA operator, LexA operator sequence; SV40/BP, simian virus 40 promoter/bacterial promoter.

Fig. 6 Colony-forming efficiency of pTargetB-NLS-Rluc and pTargetB2-NLS-Rluc

Figure 6(a) shows the vector construction of pTargetB2-NLS-Rluc and pTargetB2(0~66x)-LexA-Rluc. *attB*, *attB* sequence for phiC31 integrase-mediated recombination; $Blast^R$, blasticidin-resistance gene; CMV, cytomegalovirus promoter; Gal4, Gal4 DNA-binding domain; hEF1, human elongation factor

1α promoter; LexA, LexA DNA-binding domain; NLS, nuclear localization signal; Rluc, *Renilla* luciferase gene; SV40/EM7, simian virus 40 promoter/bacterial EM7 promoter; ori, *E. coli* origin of replication; 0~66xUAS, 0 to 66 copies of upstream activating sequences. Figure 6(b) shows the numbers of blasticidin-resistant colonies obtained following transfection with donor and helper vectors. Hela cells were transfected with 62.5 ng of the indicated donor vector and 187.5 ng of pcDNA3.1(+) (white bar) or pCMV-int (black bar). The number of colonies was counted by methylene blue staining after 2 weeks of selection with blasticidin S. Each value represents the mean + SD (n = 4). The data were analyzed by Dunnett's multiple-comparison test. Figure 6(c) shows photographs of colonies stained with methylene blue.

Fig. 7 Effect of the DNA-binding protein on the integration site selection of phiC31 integrase in chromosomal integration

Figure 7(a) shows the design of primers used to detect targeted integration. The forward and reverse primers are designed to anneal the nearby LexA operator sequence in chromosome and *attB* in the donor vectors, respectively. PCR amplified DNA only when targeted integration products are templates. Figure 7(b) and (c) represent percentages of the donor vector integrated into target *attP* per total donor vector remaining in the cells at 21 days after transfection. The helper vector (pCMV-int) was transfected 1 (b) or 2 (c) days after donor vector transfection. Each value represents the mean + SD (n = 3).

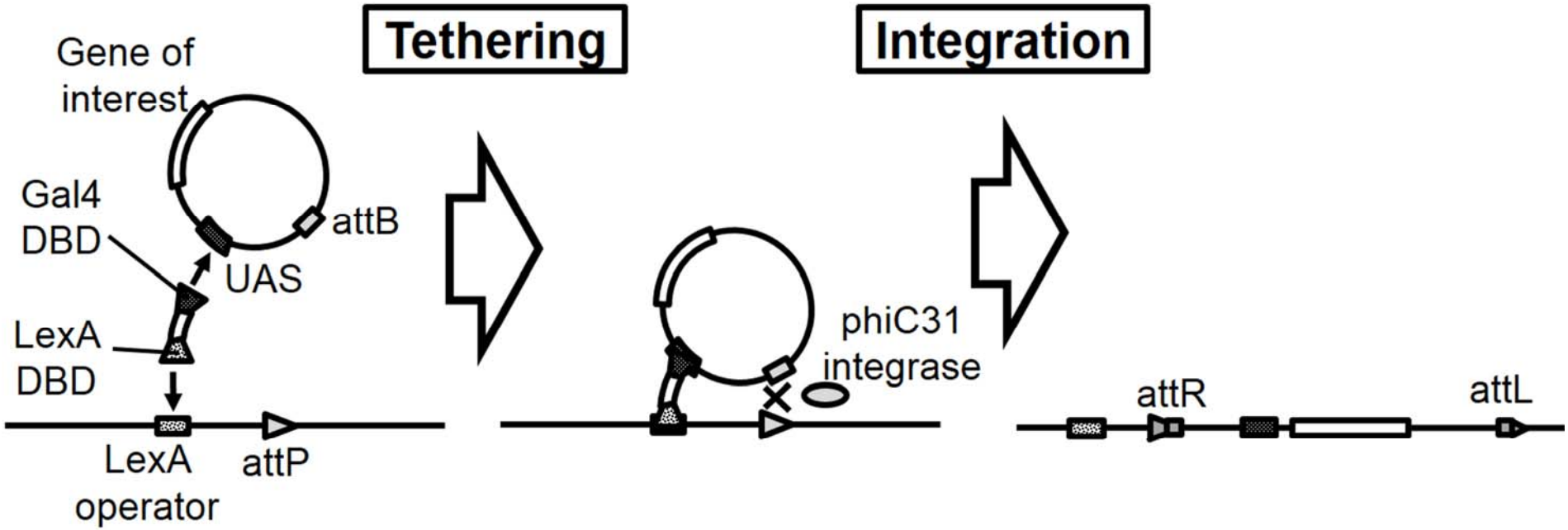

Tethering
Integration
Gene of interest
Gal4 DBD
LexA DBD
attB
UAS
LexA operator
attP
phiC31 integrase
attR
attL

Fig. 1

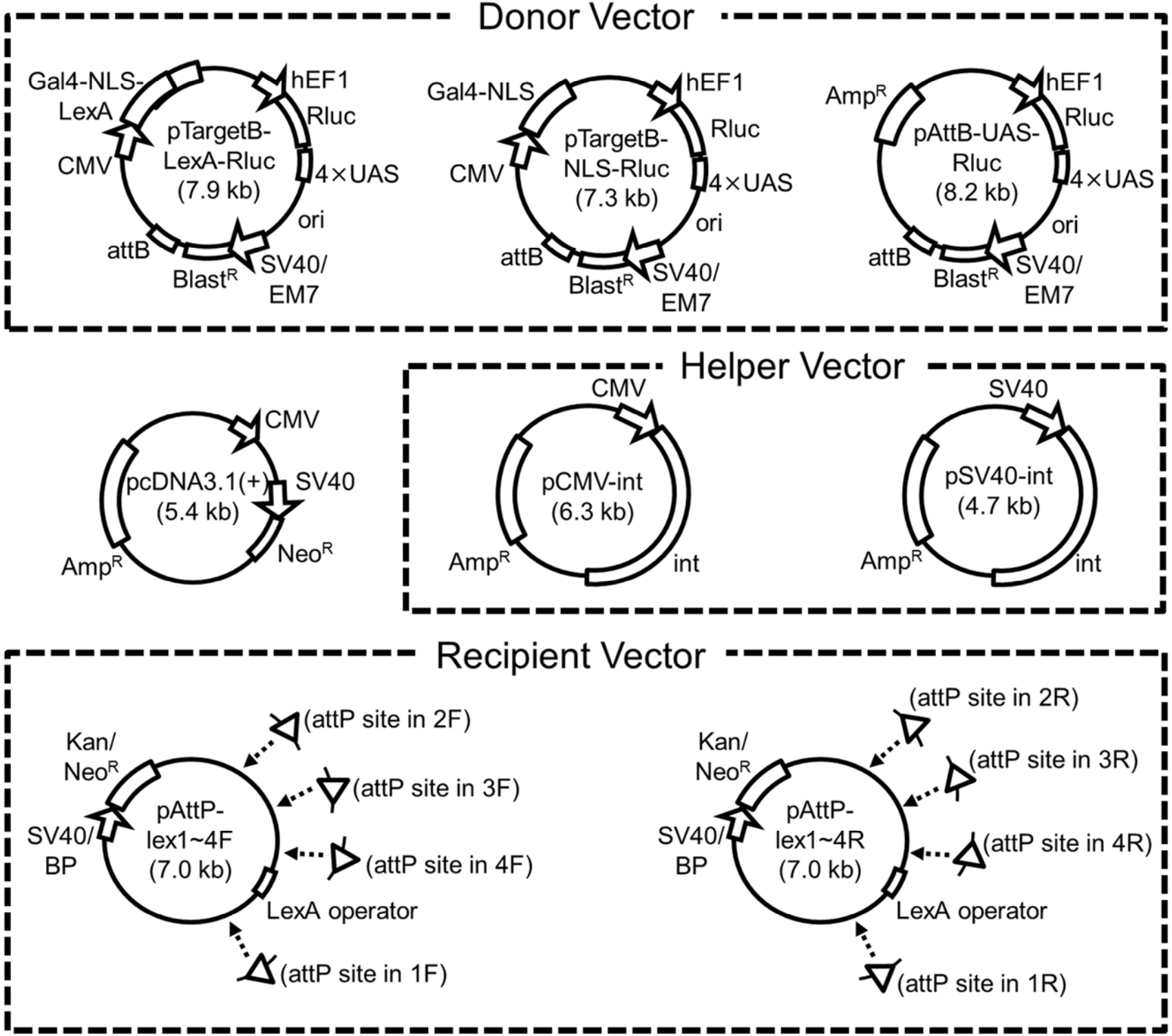

Donor Vector
Gal4-NLS-LexA
hEF1
Rluc
pTargetB-LexA-Rluc (7.9 kb)
CMV
4×UAS
ori
attB
Blast^R
SV40/EM7
Gal4-NLS
pTargetB-NLS-Rluc (7.3 kb)
Amp^R
pAttB-UAS-Rluc (8.2 kb)
Helper Vector
pcDNA3.1(+) (5.4 kb)
SV40
Neo^R
pCMV-int (6.3 kb)
int
pSV40-int (4.7 kb)
Recipient Vector
Kan/Neo^R
SV40/BP
pAttP-lex1~4F (7.0 kb)
(attP site in 2F)
(attP site in 3F)
(attP site in 4F)
(attP site in 1F)
LexA operator
pAttP-lex1~4R (7.0 kb)
(attP site in 2R)
(attP site in 3R)
(attP site in 4R)
(attP site in 1R)

Fig. 2

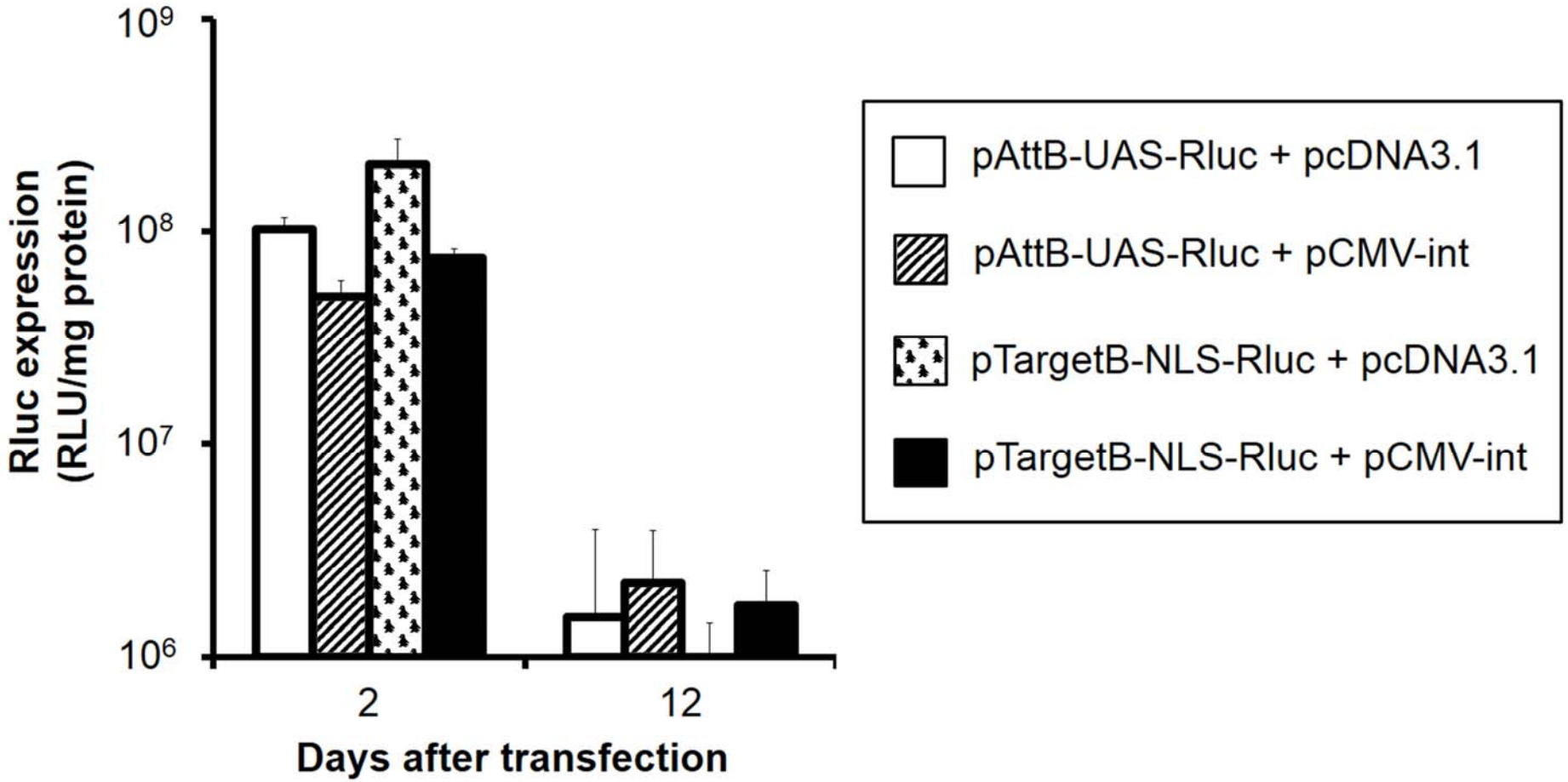

Rluc expression
(RLU/mg protein)
10^9
10^8
10^7
10^6
2
12
Days after transfection
pAttB-UAS-Rluc + pcDNA3.1
pAttB-UAS-Rluc + pCMV-int
pTargetB-NLS-Rluc + pcDNA3.1
pTargetB-NLS-Rluc + pCMV-int

Fig. 3

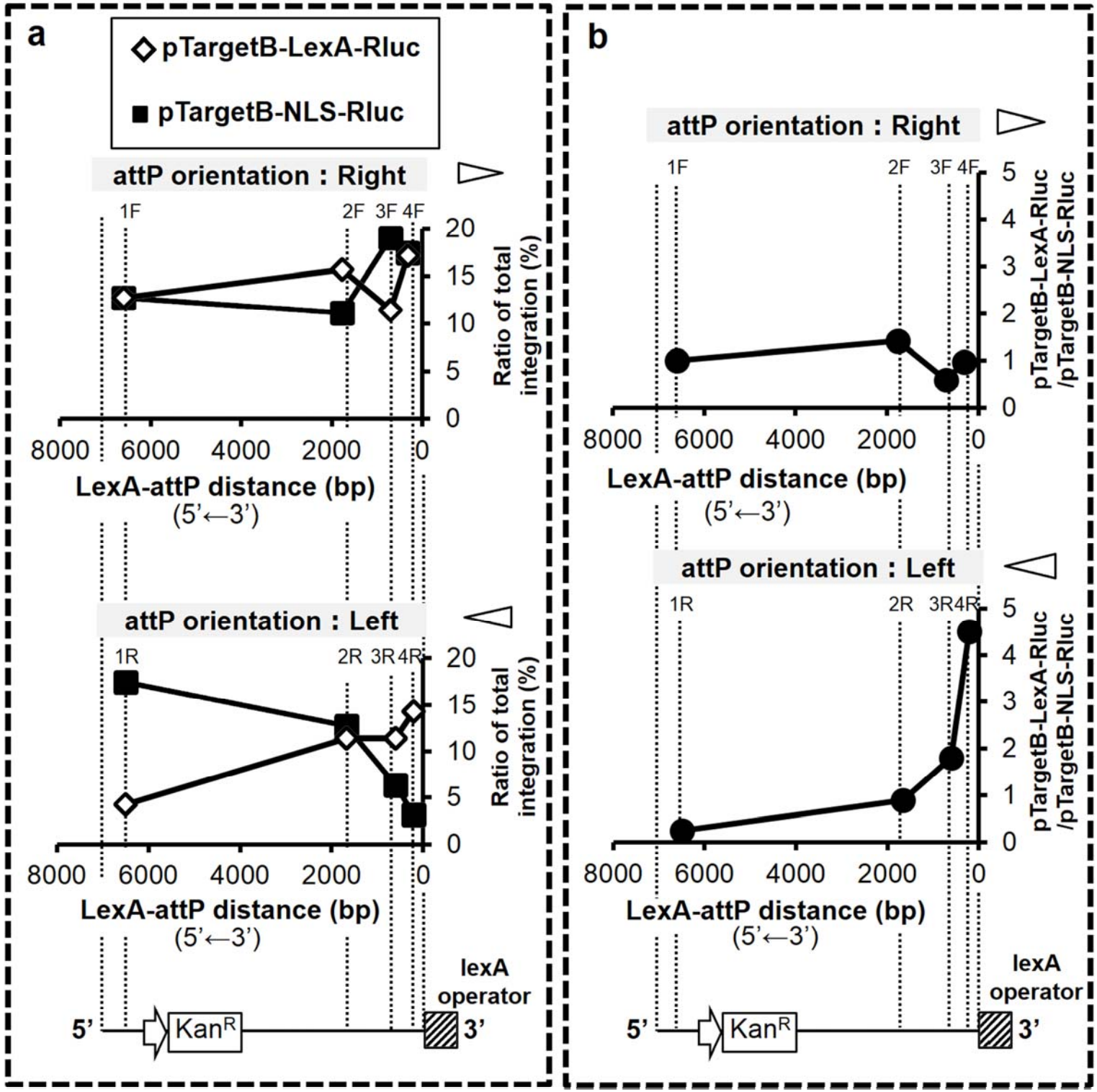

a
pTargetB-LexA-Rluc
pTargetB-NLS-Rluc
attP orientation : Right
1F
2F
3F
4F
Ratio of total integration (%)
LexA-attP distance (bp)
(5'←3')
attP orientation : Left
1R
2R
3R
4R
lexA operator
5'
Kan^R
3'
b
pTargetB-LexA-Rluc /pTargetB-NLS-Rluc

Fig. 4

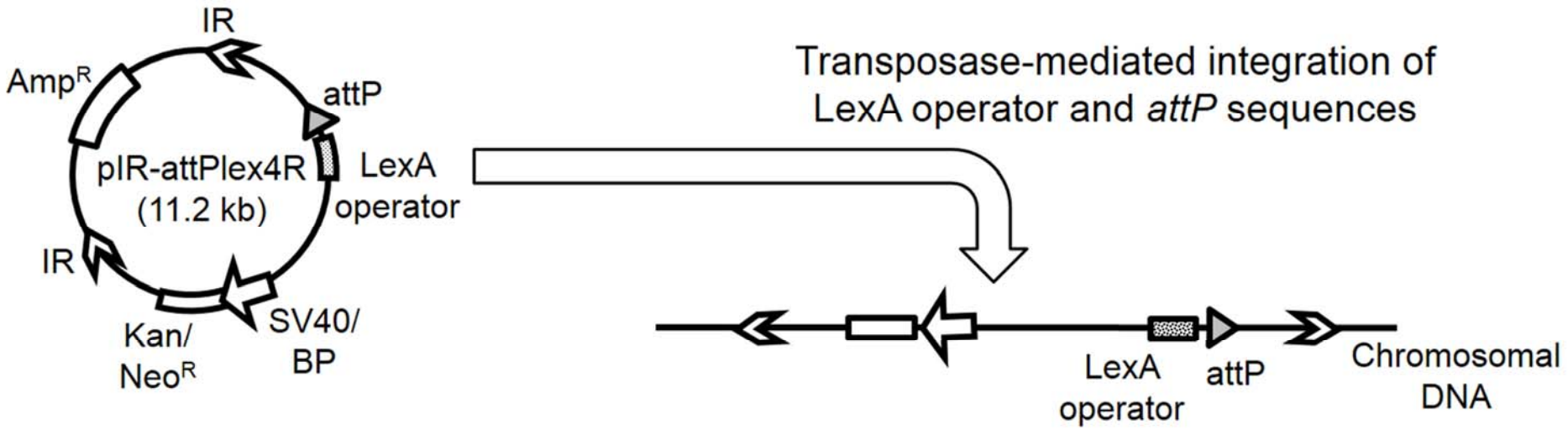

IR
Amp^R
attP
pIR-attPlex4R
(11.2 kb)
LexA
operator
IR
Kan/
Neo^R
SV40/
BP
Transposase-mediated integration of
LexA operator and attP sequences
LexA
operator
attP
Chromosomal
DNA

Fig. 5

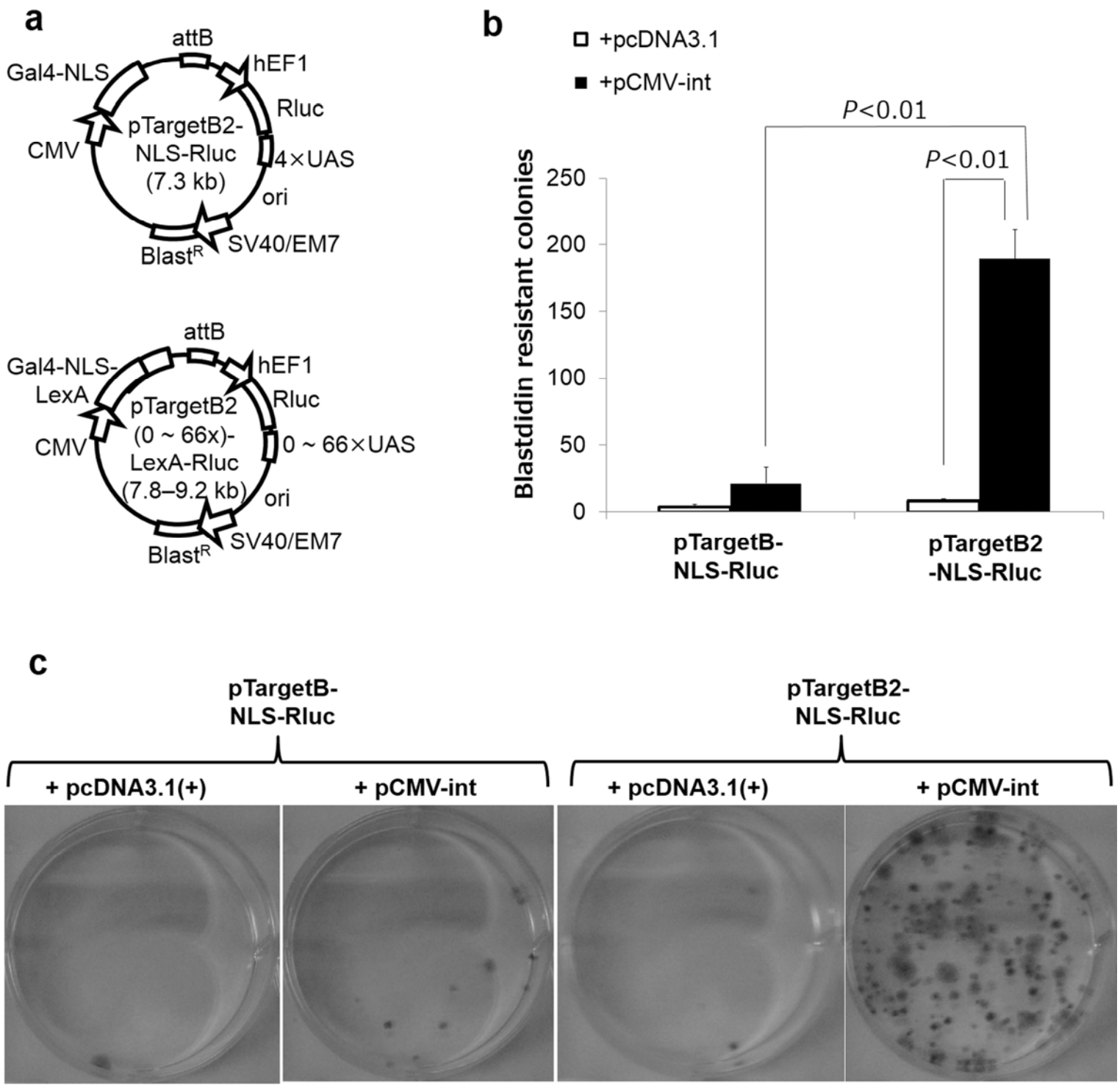
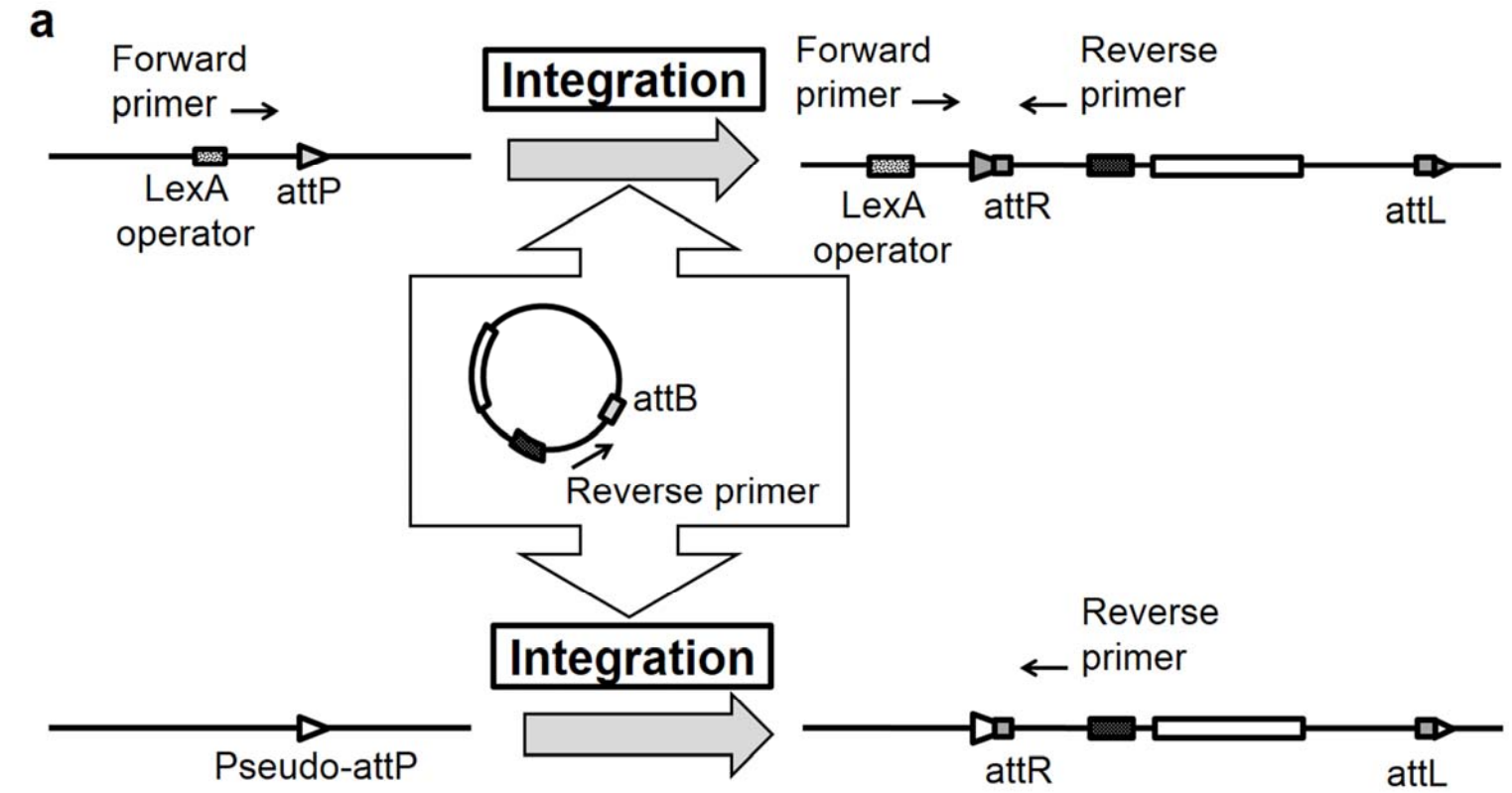
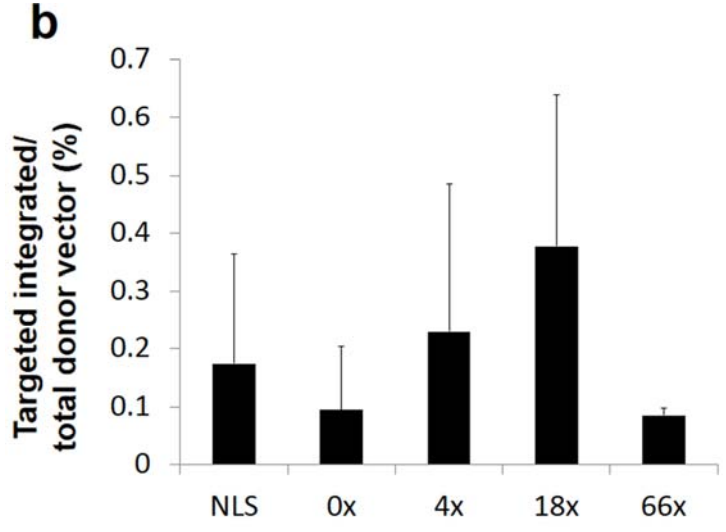

a
attB
Gal4-NLS
hEF1
Rluc
pTargetB2-
NLS-Rluc
(7.3 kb)
CMV
4×UAS
ori
Blast^R
SV40/EM7
attB
Gal4-NLS-
LexA
hEF1
Rluc
pTargetB2
(0 ~ 66x)-
LexA-Rluc
(7.8–9.2 kb)
CMV
0 ~ 66×UAS
ori
Blast^R
SV40/EM7
b
+pcDNA3.1
+pCMV-int
P<0.01
P<0.01
Blastdidin resistant colonies
250
200
150
100
50
0
pTargetB-
NLS-Rluc
pTargetB2
-NLS-Rluc
c
pTargetB-
NLS-Rluc
pTargetB2-
NLS-Rluc
+ pcDNA3.1(+)
+ pCMV-int
+ pcDNA3.1(+)
+ pCMV-int

Fig. 6


a
Forward primer
Integration
Forward primer
Reverse primer
LexA operator
attP
LexA operator
attR
attL
attB
Reverse primer
Integration
Reverse primer
Pseudo-attP
attR
attL


b
Targeted integrated/ total donor vector (%)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
NLS
0x
4x
18x
66x

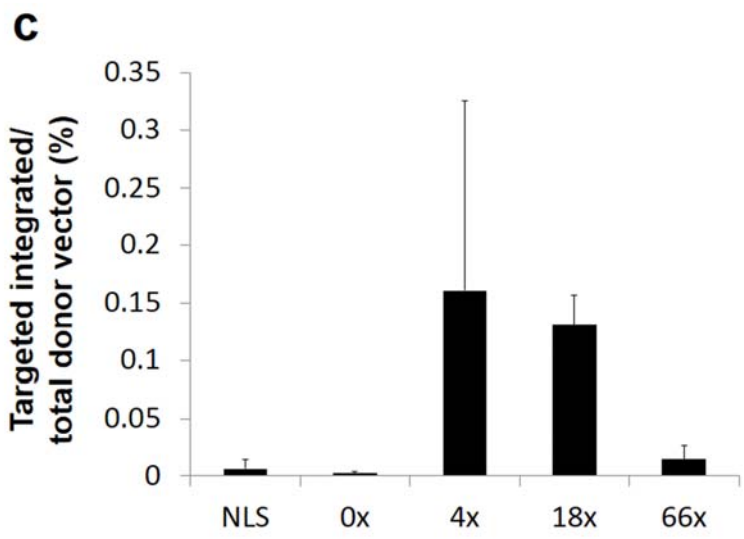

c
Targeted integrated/ total donor vector (%)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
NLS
0x
4x
18x
66x

Fig. 7

Table 1

Sequence and locations of phiC31 integrase-mediated integration of pTargetB2(4x)-LexA-Rluc

Sequences	Locations
CGCGCCCGGGGAGCCCAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC	320 bp upstream of CMV promoter in pCMV-int
CGCGCCCGGGGAGCCAAAGGTTACCCCAGTTGGGGCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATTAAATGTC	*attP* located 219 bp from LexA operator sequence

Columns 1 and 2 show sequences of *attR* and flanked sequences and locations of integration sites. Sequences with underlines refers pTargetB2(4x)-LexA-Rluc derived sequences.

Supplementary Methods

pDNA construction

1. Construction of pIR-UAS-Rluc

pIR-blastHGF was constructed as previously described (Nakanishi et al., 2010, Mol. Ther., 18, 707-714). The UAS insert was created by PCR using the primers (AAATCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCG TCGCGACGGAGTACTGTCC, CTATATTACCCTGTTATCCCTAGCGTAACTCGGAGGA CAGTACTCCGGTCGGAGGACAGTACTCCGTCGCGA). To construct pIR-UAS-HGF, pIR-blastHGF was digested with the restriction enzyme SwaI, and ligated with the UAS insert.

The Rluc cDNA insert was amplified by PCR using pRL-SV40 (Promega, Tokyo, Japan) as a template and the primers (CACCGGTATGACTTCGAAAGTTTATGATCCAGA, GCTCTAGAATCGATGAATTATTGTTCATTTTTGAGAACTCG). The Rluc cDNA insert was digested with restriction enzymes AgeI and XbaI. To construct pIR-UASRluc, pIR-UASHGF was digested with restriction enzymes AgeI and NheI, and ligated with the digested Rluc cDNA insert.

2. Construction of pTarget-NLS-Rluc

The NLS insert was created by PCR using the primers (GAAGATCTGGCGATCGCC GATCCTAAGAAGAAGCGCAAGGTGGGCGACCCGAAAAAGAAACGCAAAGTT, TTCCAATGCATTGGCTGCAGGTTTAAACGGAATTCttctcACCGGTACCAACTTTGCGTTTCT TTTTCGGG). To construct pET-NLS, the NLS insert and pET-42b (+) (Merk KGaA, Darmstadt, Germany) were digested with restriction enzymes BglII and PstI, and ligated.

To obtain the DNA fragment containing NLS, pET-NLS was digested with restriction enzymes SgfI and PmeI. To construct pBIND-NLS, the DNA fragments containing NLS and pFN11A (BIND) (Promega) digested with SgfI and PmeI were ligated.

pIR-UASRluc was digested with restriction enzymes SalI and MluI, and pBIND-NLS was digested with restriction enzymes BglII and FspI, respectively. To construct pTarget-NLS-Rluc-pre, these two digested pDNAs were blunted and ligated.

Two DNA fragments were created by PCR using pTarget-NLS-Rluc-pre as a template and two primer sets (ATGACTTCGAAAGTTTATGATCCAgaacaa + AGGTTTAGTTCCTGGTGTACTTGAgggggatg, GGCCGGCCAATTCGCTAGAGGGCCCTATTCTATAGTGTCA + CCGCGGATCTCAGGTAGGCGCCGGTCA). To construct pTarget-NLS-Rluc, these two DNA fragments were ligated.

3. Construction of pTarget-LexA-Rluc

The LexA DNA-binding domain (LexA-DBD) insert was amplified by PCR using pLexA-C bait vector (Dualsystems Biotech, Schlieren, Switzerland) as a template and primers (TCCCCCCGGGcgaaccagttggatgaaagcgtta, CGGAATTCtcacagccagtcgccgttgc). pTarget-NLS-Rluc was digested with restriction enzymes AgeI and EcoRI, and the LexA-DBD insert was digested with restriction enzymes XmaI and EcoRI. To construct pTarget-LexA-Rluc, the digested pTarget-NLS-Rluc and LexA-DBD insert were ligated.

4. Construction of pAttB-UAS-Rluc, pTargetB-NLS-Rluc, pTargetB-LexA-Rluc

The *attB* insert was amplified by PCR using pORF-luc-attB as template and primers (CGCGCACGTACGAAACCGAAGCGAATTTCGAGGTG, CGCGCACGTACGGCCGCTCGAGGCATCAAGCTAAT). pIR-UAS-Rluc, pTarget-NLS-Rluc, pTarget-LexA-Rluc and the *attB* insert were digested with restriction enzyme BsiWI. To construct pAttB-UAS-Rluc, pTargetB-NLS-Rluc and pTargetB-LexA-Rluc, the digested *attB* insert was ligated with digested pIR-UAS-Rluc, pTarget-NLS-Rluc and pTarget-LexA-Rluc, respectively.

5. Construction of pTargetB2(4x)-NLS-Rluc

The *attB* insert was prepared as described above. The *attB* insert was phosphorylated, and ligated with pTarget-NLS-Rluc digested with restriction enzyme BmgBI.

6. Construction of pTargetB2(4x)-LexA-Rluc

pTargetB2(4x)-NLS-Rluc was digested with restriction enzymes BssHII and ApaLI, and the 2.8 kb DNA fragment containing the *attB* sequence was purified. pTarget-LexA-Rluc was digested with restriction enzymes BssHII and ApaLI, and the 5.1 kb DNA fragment containing the

gene of the DNA-binding protein was purified. To obtain pTargetB2(4x)-LexA-Rluc, these purified DNA fragments were ligated.

7. Construction of pTargetB2(6, 10, 18, 34 and 66x)-LexA-Rluc

pTargetB2(4x)-LexA-Rluc was digested with restriction enzymes SwaI and AvrII, and the 1.2 kb DNA fragment containing 4xUAS was purified. pTargetB2(4x)-LexA-Rluc was also digested with restriction enzymes NruI and AvrII, and the 6.7 kb DNA fragment containing 2xUAS was purified. To obtain pTargetB2(6x)-LexA-Rluc, these purified DNA fragments were ligated.

pTargetB2(10, 18, 34 and 66x)-LexA-Rluc was constructed by repeating the same procedure.

8. Construction of pTargetB2(0x)-LexA-Rluc

pTargetB2(18x)-LexA-Rluc was digested with restriction enzymes SwaI and AvrII, and the 6.7 kb DNA fragment containing no UAS was purified. pIR-blastHGF was also digested with restriction enzymes SwaI and AvrII, and the 1.1 kb DNA fragment was purified. To obtain pTargetB2(0x)-LexA-Rluc, these DNA fragments were ligated.

9. Construction of pSV40-int

pCMV-int (Addgene, Cambridge, MA, USA) was digested with restriction enzymes SpeI and NheI, and the DNA fragment containing the phiC31 integrase gene was purified. pRL-SV40 was digested with restriction enzymes NheI and XbaI, and the DNA fragment containing the SV40 promoter was purified. To construct pSV40-int, these DNA fragments were ligated.

10. Construction of pVITRO1-lexluc

pCMV-luc was constructed as previously described (Nomura et al., 1999, Gene Ther., 6, 121-129). The firefly luciferase gene flanked with the lexA operator sequence was amplified by PCR using pCMV-luc as a template and primers (CGGGATCCCTGTATATATATACAGATGGAAGACGCCAAAAACATAA, AAACGTACGCTAGTTACACGGCGATCTTTCC). To construct pVITRO1-lexluc, the PCR product and pVITRO1-neo-mcs (Invivogen, San Diego, CA, USA) were digested with restriction enzymes BamHI and BsiWI, and ligated.

11. Construction of pAttP-DsRed express

The whole DNA of pCMV-DsRed express (Takara Bio, Otsu, Japan) flanked with the *attP* sequence was amplified by PCR using pCMV-DsRed express as a template and primers (AGTTCTCTCAGTTGGGGGGCATTATGCCCAGTACATGACCTT, caaAGGTTACCCCAGTTGGGGcaggcgggccatttaccgtcatt). To construct pAttP-DsRed express, the PCR product was phosphorylated and self-ligated.

12. Construction of pAttP-lex1F

pAttP-DsRed express was digested with restriction enzymes ApaLI and NdeI. The DNA fragment containing the *attP* sequence was purified and blunted. pVITRO1-lexluc was digested with restriction enzyme PsiI and the DNA fragment containing the LexA operator sequence was purified. To construct pAttP-lex1F, the DNA fragments derived from pAttP-DsRed express and pVITRO1-lexluc were ligated.

13. Construction of pAttP-lex1R

pAttP-lex1F was digested with restriction enzymes ApaLI and NdeI. The DNA fragments were purified, blunted, and ligated. To obtain pAttP-lex1R, the pDNA with the *attP* sequence in a reverse orientation compared to pAttP-lex1F was selected from the ligation products.

14. Construction of pAttP-lex2F, 2R, 3F, 3R, 4F, 4R

To obtain pAttP0-lex, pAttP-lex1F was digested with restriction enzymes AfeI and ScaI, and the DNA fragment containing the LexA operator sequence was purified and self-ligated. pAttP-lex1F was digested with restriction enzymes AfeI and ScaI, and the DNA fragment containing the *attP* sequence was purified. To construct pAttP-lex2F, 2R, 3F, 3R, 4F and 4R, the DNA fragment was ligated with pAttP0-lex digested with restriction enzyme SnaBI, SmaI or BstZ17I, respectively.

15. Construction of pIR-attPlex4Rt

The DNA containing the 3’ and 5’ inverted repeat sequences of *piggyBac* transposon was amplified by PCR using p3E1.2 (a gift from Prof. Hajime Mori, Kyoto Institute of Technology, Kyoto, Japan) as a template, and primers (AGAACTACCCATTTTATTATATATTAGTCACGA, AATACAACATGACTGTTTTTAAAGTACAAAAT). To construct pIR-attPlex4R, pAttP-lex4R was

digested with restriction enzyme HindIII, phosphorylated, and ligated with the PCR product containing inverted repeat sequences of *piggyBac* transposon.

16. Construction of pVITRO1-neo-RnasePfragment

The DNA containing a part of the RNaseP gene was amplified by PCR using the Hela-attPlex4R genome as a template and primers (AGATTTGGACCTGCGAGCG, GAGCGGCTGTCTCCACAAGT). To obtain pVITRO1-neo-RNasePfragment, the PCR product was purified, phosphorylated, and ligated with pVITRO1-neo-mcs digested with restriction enzyme BstZ17I.