# Improved site-specific recombinase-based method to produce selectable marker- and vector-backbone-free transgenic cells

Yuan Yu[1,2†], Qi Tong[1,2†], Zhongxia Li[1,2†], Jinhai Tian[1,2], Yizhi Wang[1,2], Feng Su[1,2], Yongsheng Wang[1,2], Jun Liu[1,2], Yong Zhang[1,2]*

[1]College of Veterinary Medicine, Northwest A&F University, Yangling, Shaanxi, PR China, [2]Key Laboratory of Animal Biotechnology, Ministry of Agriculture, Northwest A&F University, Yangling, Shaanxi, PR China

## SUPPLEMENTARY MATERIALS AND METHODS

### Plasmid Construction

LAR (LoxP-AttB-Rox) fragment was generated using a splicing method by overlap extension (SOE) PCR, in which LAR_F1 to LAR_R2 primers were used. LAR was digested with HindIII/SalI and then cloned into pGEM-3Z (Promega, Madison, WI); the resulting clone was labeled pGEM-LAR. The pGEM-LAR plasmid was digested with SalI/BamHI and ligated in the coding sequences of different linkers [No linker, $(Gly_4Ser)_3$ linker, or P2A peptide] generated by annealing with No_linker_F/No_linker_R, (Gly4Ser)3_F/(Gly4Ser)3_R or P2A_F/P2A_R primers. pGEM-LAR-Nolinker, pGEM-LAR-$(Gly_4Ser)_3$, or pGEM-LAR-P2A plasmid was digested with NcoI to obtain an LAR-linker fragment; this fragment was then cloned in pORF-TK (Invivogen, San Diego, CA). LA35R (LoxP-AttB35-Rox) fragment generated by SOE-PCR was digested with NcoI

and cloned in pORF-TK. Multiple TK fusion constructs were generated by cloning NheI-digested TK fusion genes in pIRES2-AcGFP1-Nuc (Clontech, Mountain View, CA).

CAGGS promoter was obtained by Ase/NheI digestion of pCAGGS [61] and cloned into pEGFP-N1 to construct the pCAG-attBrP2ATK. A BGH poly(A) terminator was obtained by PCR amplification from pcDNA3.1(+) (Invitrogen, Carlsbad, CA) and cloned downstream of the CAGGS promoter. The attBrP2ATK fusion gene was then obtained by NheI digestion and inserted between CAGGS promoter and BGH ploy(A). pBudCE4.1 (Invitrogen, Carlsbad, CA) was digested with XhoI/KpnI to obtain EF1a promoter; this promoter was then inserted upstream of the EGFP gene. The final pCAG-attBrP2ATK plasmid was generated by adding a MCS.

A fragment carrying the CMV promoter was cloned into the unique XhoI-HindIII sites upstream of the firefly luciferase gene of pGL4.10[luc2] vector (Promega), then an AscI site was introduced into BamHI-SalI sites downstream of the SV40 poly(A) signal to create pCMV-FLuc. Plasmid pCAG-attBrP2ATK-FLuc was generated by cloning XhoI-AscI fragment from pCMV-FLuc containing the CMV-FLuc expression cassette into the MCS of pCAG-attBrP2ATK.

**Immunofluorescence assay**

At 48 h post-transfection, the cells were fixed with 4% paraformaldehyde and immunostained with polyclonal antibody against HSV-1 thymidine kinase (1:500) and Cy3-labeled secondary antibody (1:500; Beyotime) in antibody dilution buffer (Beyotime). DAPI was used to counterstain the nuclei.

## Screening for site-specific integrations

At 8 d to 12 d post-transfection, G418-resistant HEK293 cell colonies were trypsinized and collected in pools. The cell pools were grown to confluency and genomic DNA was extracted. We then designed primers annealing to the inserted vector and the sequences flanking the 19q13.31 site to screen the site-specific integrations and detect the plasmid integration in both orientations (Table S1). Another set of primers was designed (Table S1) to screen the previously evaluated bovine safe genomic harbor integration of newly constructed pCAGGS-attBrP2ATK. The method of the two-step nested PCR was performed according to the following procedures. In the first round of PCR reactions, 100 ng to 400 ng of genomic DNA was used. Approximately 2 μl of the first-round PCR products were then used as a template in the second-round PCR. In both PCR rounds, 40 cycles and an annealing temperature of 55 °C to 60 °C were conducted. Second-round PCR products were subjected to agarose gel electrophoresis; the bands with expected sizes were excised, purified using a gel extraction kit (Axygen, CA, USA), and ligated to pMD19-T vector (Takara, Dalian, China) for sequencing.

## Quantitative RT-PCR analyses

Real-time PCR was performed using SYBR Premix Ex Taq (TaKaRa, Dalian, China) and a StepOne Plus thermocycler (Applied Biosystems) with the following parameters: 95 °C for 10 s; followed by 40 cycles at 95 °C for 5 s; and 60 °C for 30 s. For RT-PCR, total RNA was extracted from each sample by using TRIzol reagent (Invitrogen); reverse transcription was performed using a PrimeScript™ RT Reagent Kit (TaKaRa) to generate cDNA. The EGFP gene primers used in absolute or relative quantitative PCR were

5′-GCAGAAGAACGGCATCAAGGT-3′ (forward) and 5′-ACGAACTCCAGCAGGACCATG-3′ (reverse). The GAPDH gene, as a reference gene in relative quantitative RT-PCR, was amplified using the following primers: 5′-TCAACGGGAAGCTCACTGG-3′ (forward) and 5′-CCCCAGCATCGAAGGTAGA-3′ (reverse). For each DNA and cDNA sample, target and reference genes were amplified independently on the same plate and in the same experimental run in triplicate. PCR specificity was confirmed by gel electrophoresis on a 2.5% agarose gel; PCR specificity was also indicated by a single peak in the melting curve. For relative quantitative RT-PCR, the amount of target normalized to the reference was calculated by $2^{-\Delta\Delta Ct}$ method.

**Establishment of the absolute quantitative standard curve**

Generation of an absolute quantitative standard curve was necessary to determine the copy number. A series of standard samples containing 0.5, 1, 2, 4, 8, and 16 copies of the EGFP gene were prepared as described previously [62], in which the wild-type genome of a Holstein cow was mixed with the pARNG-HBD3 plasmid. To prepare a standard sample containing one copy of the EGFP gene, we calculated the quantity of the plasmid mixed with genomic DNA according to the following equation: $\frac{a \times b \times 0.5}{2.45 \times 10^9}$ ("*a*" represents the size of plasmid). The absolute quantitative standard curve was drawn by plotting Ct values against the log of EGFP gene copies of the corresponding standard samples. The parameters of the standard curve were as follows: log2N = −0.9568 Ct + 25.335 ($R^2$ = 0.9887, P < 0.001).

**Half-nested inverse PCR**

Genomic DNA (5 μg) was digested with PstI (New England Biolabs) overnight. After extraction with phenol/chloroform and ethanol precipitation, digests were ligated with a DNA Ligation Kit (TaKaRa) at 4°C overnight in a total reaction volume of 50 μL. Ligated DNA was extracted with phenol/chloroform, ethanol precipitated, and resuspended in 25 μL nuclease-free water. Primary PCR was performed with HNI_F1 and HNI_R primers that were designed against *attL* sites (Supplementary Table S1) as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, an annealing temperature gradient of 51–61°C for 45 s and 72°C for 8 min, and then 72°C for 10 min. One microliter of the PCR product was used as a template in the secondary reaction. The secondary PCR was performed with the HNI_F2 and HNI_R primers (Supplementary Table S1). The PCR conditions were the same as those for the primary PCR except for an annealing temperature gradient of 55–65°C. The location of the primers was shown in Supplementary Figure S4. PCR products were run on agarose gels and the resulting bands were excised. The purified product was cloned into a pMD19-T vector (TaKaRa) and sequenced with M13 and reverse M13 primers. Sequences were examined by BLAST searching of the bovine genome databases (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9913).

**Luciferase activity assay**

The Luciferase assay (Promega, Madison, WI) was performed according to the manufacturer's protocol. In brief, cells grown on 24-well plates were washed three times with phosphate buffered saline (PBS) and incubated in 200 μl lysis buffer for 30 min on ice. Twenty microliters of cell lysate was mixed with 100 μl of luciferase reagent and

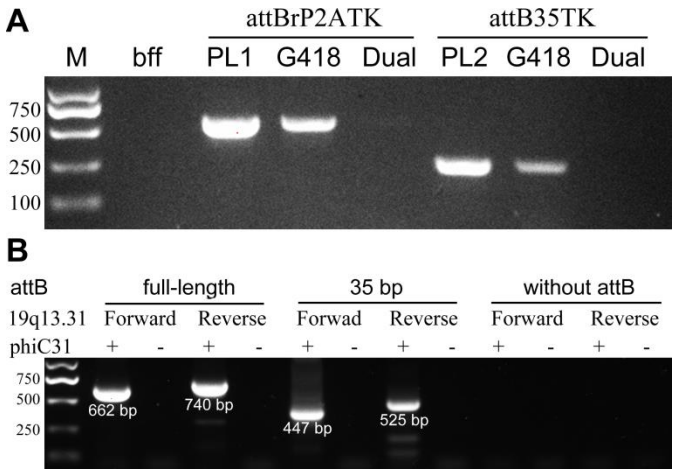
luciferase activity was read on the luminometer. Luciferase activity was normalized to total protein measured using the BCA kit (Beyotime). Results were expressed as a percentage of luciferase values of stably-transfected cells at passage 1 (to which a value of 100% was assigned).

**SCNT, activation, and culture of SCNT embryos**

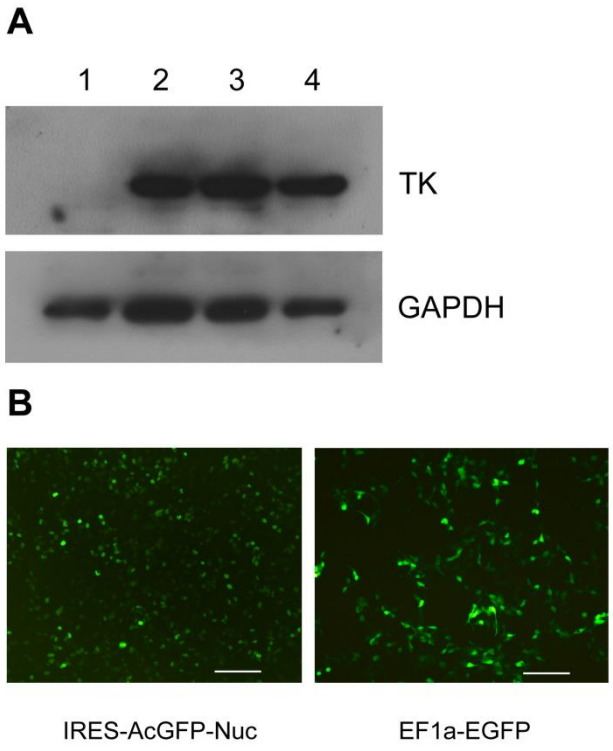
SCNT, activation of reconstructed embryos, and culture of SCNT embryos were performed as described previously [63]. In brief, matured oocytes were enucleated using a 20 mm inner diameter glass pipette to remove the first polar body and a small amount of the surrounding cytoplasm. Successful enucleation was confirmed by Hoechst 33342 staining. A single disaggregated donor cell was injected into the pre-vitelline space of an enucleated oocyte. Oocyte-cell fusion was performed using a pair of platinum electrodes connected to a micromanipulator in microdrops of Zimmermann's fusion medium at a double electrical pulse of 35 V for 10 ms. Reconstructed SCNT embryos were stored in synthetic oviductal fluid (SOFaa) containing 5 mg/mL of cytochalasin B for 2 h until activation. The mSOF medium was prepared according to a formula described previously [64] and supplemented with 8 mg/mL of bovine serum albumin, 1% MEM non-essential amino acids, and 2% BME essential amino acids. The reconstructed embryos were activated in 5 mM ionomycin for 4 min and exposed to 1.9 mM dimethynopyridine in SOFaa for 4 h. After activation, the embryos were cultured in G1.3/G2.3 sequential media (Vitrolife AB, Gothenburg, Sweden). Droplets of 150 μL of G1.3 were prepared in a 35 mm cell culture dish under mineral oil and equilibrated for 2 h before the embryos were loaded (20 embryos/microdrop). The embryos were then

transferred to G2.3 droplets at 3 d of culture (0 d corresponded to the day of SCNT). Fresh senven-day-old blastocysts were nonsurgically transferred to the uterine horn ipsilateral to the corpus luteum in Red Angus recipients at day of standing estrus. Pregnancy was detected by rectal palpation at 90 d of gestation.

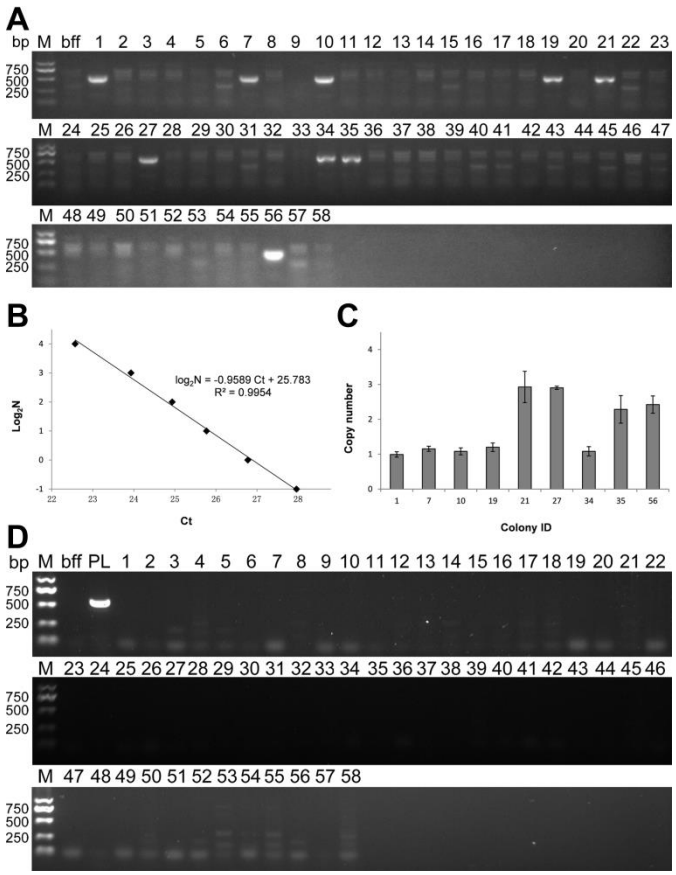
**Supplementary Figures**




**Supplementary Figure S1. PCR assays for site-specific integration on HEK293 cells.** (A) PCR assay to detect non-recombined *attB* in cell pools derived by G418 screening or cell pools derived by G418/GCV dual selection. bff: untransfected bovine fetal fibroblast cells; PL1: attBrP2ATK plasmid; PL2: attB35TK plasmid. (B) PCR assay to screen the 19q13.31 site integration of TK constructs. Only the cells co-transfected with both attB-containing donor vector and functional phiC31 integrase showed an appropriately sized PCR product.

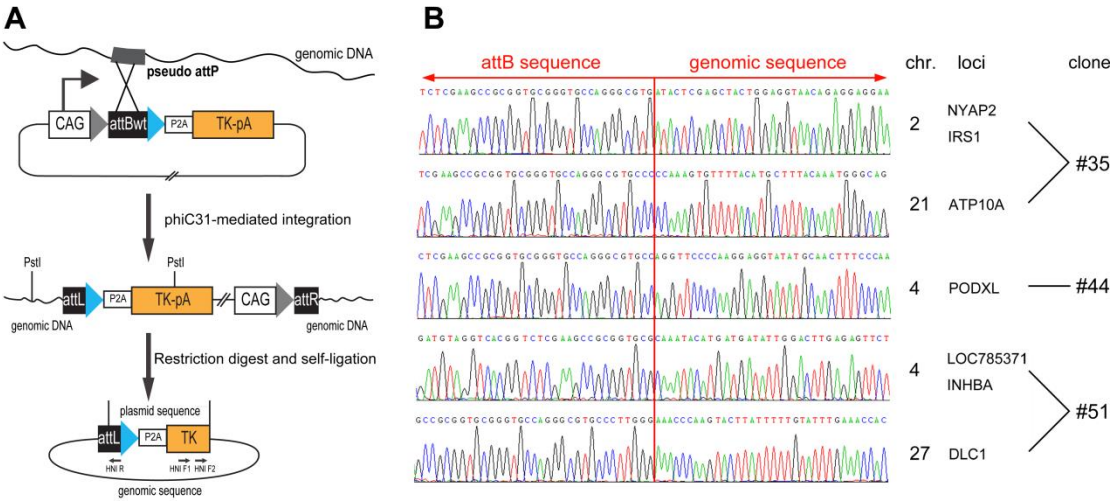



**Supplementary Figure S2. Functional verification of the designs in pCAG-attBrP2ATK vector.** (A) TK protein expression was detected by western blot at 48 h after transfection. Lanes 1 to 4 indicated HEK293 cells transfected with pEGFP-N1 (empty vector control), attBrP2ATK construct (CMV promoter), pCAG-attBrP2ATK (CAGGS promoter), and pORF-HSVtk (positive control), respectively. The robust expression of TK protein was detected with CMV and CAGGS promoters. HEK293 cells transfected with the control vector showed no significant signal. (B) GFP protein expression was observed by fluorescence microscopy at 48 h post-transfection with attBrP2ATK construct (IRES-AcGFP-Nuc) or pCAG-attBrP2ATK (EF1a-EGFP). Scale bars = 10 μm.

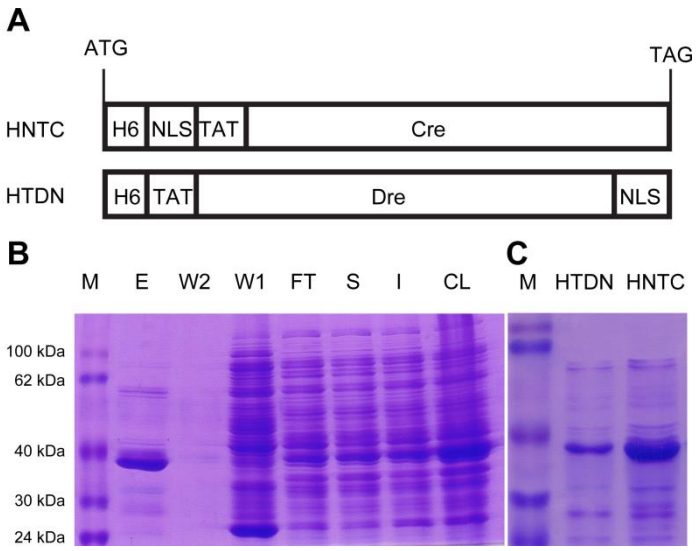



**Supplementary Figure S3. Analysis of stably-transfected bovine fetal fibroblast clones obtained by G418/GCV dual selection.** (A) Verification of the safe harbor integration by a nested PCR analysis using vector- and locus-specific primers. Only cell colonies with the safe harbor integration showed the specific band of 491 bp. Lane M: DNA ladder; Lanes 1–58: second-round PCR products derived from 58 selected bovine fetal fibroblast colonies; Lane bff: PCR control using bovine fetal fibroblast cells as template. (B) The absolute quantitative standard curve was drawn by plotting Ct values against the log of EGFP gene copies of the corresponding standard samples. The parameters of the standard curve were as follows: log2N = –0.9589 Ct + 25.783 ($R^2$ = 0.9954, P < 0.001). (C) Copy number assay of the nine safe harbor-integrated colonies by absolute quantitative PCR. Five out of nine colonies revealed single-copy integration. Error bars denote SEM. (D) Detection of random integration in the stably-transfected clones by

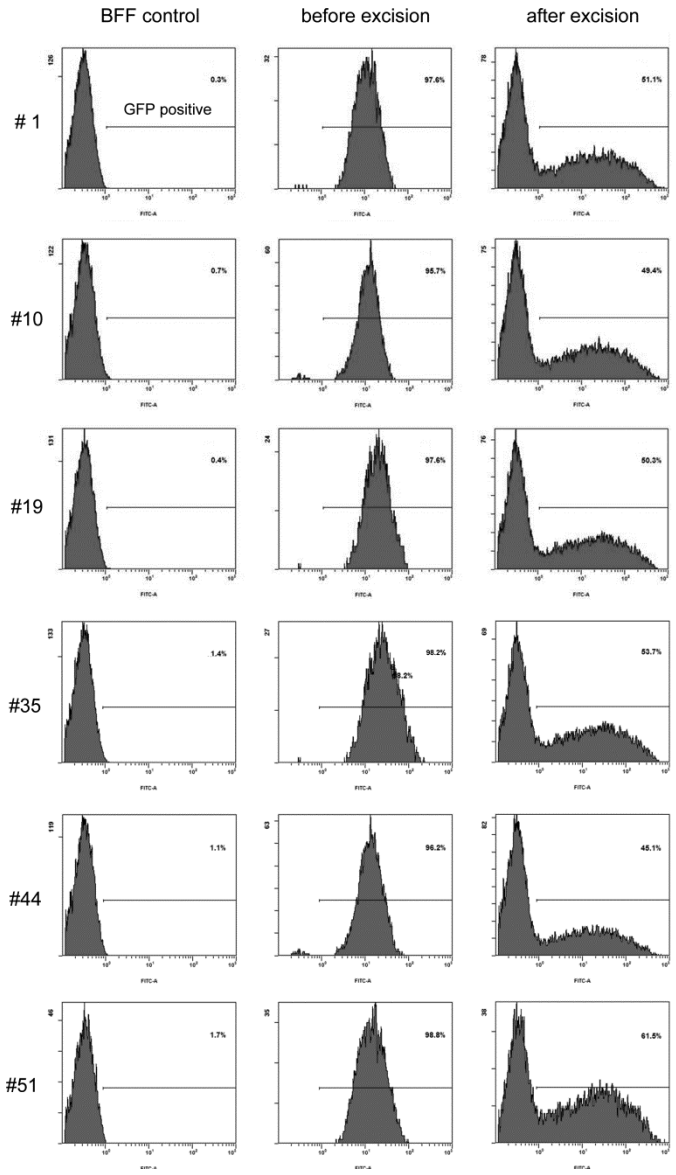
using PCR for the non-recombined *attB* site. Cell clones derived from random integration should show the specific band of 587 bp. Lane PL: PCR control using plasmid pCAG-attBrP2ATK as template.



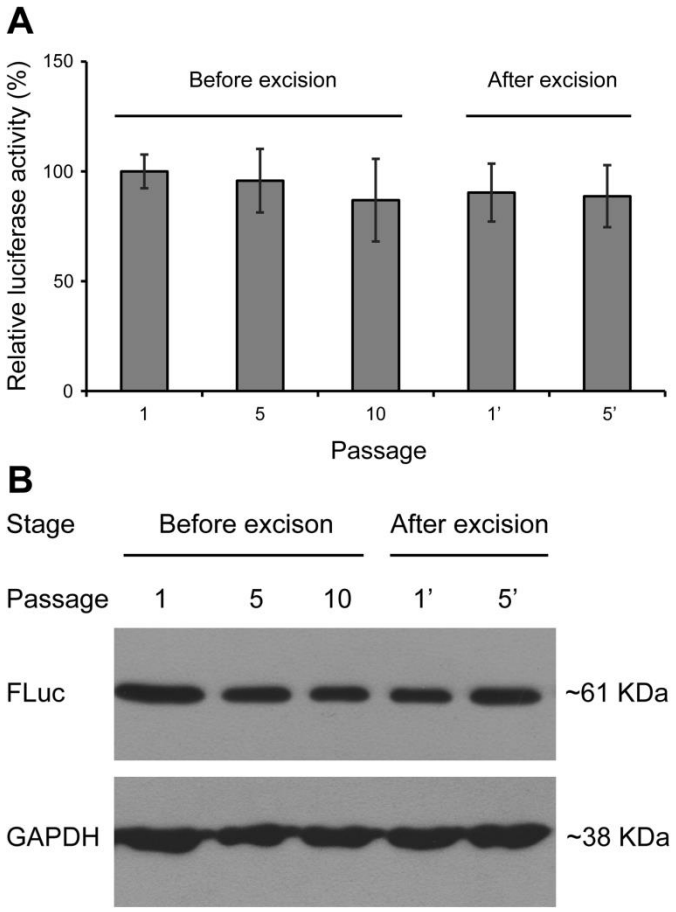

**Supplementary Figure S4. Site-specific integration mediated by phiC31 integrase and the identification of pseudo attP sites.** (A) In the context of bovine cells, phiC31 integrase targets the plasmid pCAG-attBrP2ATK into native pseudo *attP* sites. Genomic DNA with an integration event is digested by restriction enzyme PstI and the digests are self ligated. Half-nested inverse PCR was performed to identify the integration sites. The *attL* junction was rescued using the vector-specific primers HNI_F1, HNI_F2, and HNI_R. The PCR products were then purified and sequenced. (B) Pseudo-*attP* sites in the bovine genome were identified by half-nested PCR. The sequences begin in the *attB* region of the transgene plasmid and join with bovine genomic DNA in different loci. The identified pseudo-*attP* sites of three representative clones were located in the intergenic regions or within an intron.




**Supplementary Figure S5. Recombinant-modified Cre and Dre protein mediated recombination.** (A) Schematic of cell-permeant His-NLS-TAT-Cre (HNTC) and His-TAT-Dre-NLS (HTDN) fusion proteins. H6: 6× His tag; NLS: nuclear localization sequence derived from SV40 large T antigen; TAT: basic protein translocation peptide derived from HIV-TAT. (B) Purification of recombinant HTDN protein from bacteria as analyzed by Coomassie blue staining of an SDS-PAGE gel. CL: Cleared lysate; I: Insoluble; S: Supernatant; FT: Flow-through; W: Washing; E: Eluate; M: Marker. (C) Comparison of purified HTDN and HNTC fusion proteins on a Coomassie-stained gel.




**Supplementary Figure S6. FACS analysis after cell-permeant Cre and Dre protein transduction into three stably-transfected cells lines.** At 5 d after protein transduction, cells were trypsinized and resuspended in PBS containing 10% FBS. The percentage of GFP-expressing cells was then analyzed by flow cytometry. Approximately 50% of the transduced cells lost their fluorescence as shown by flow cytometry. Untransfected bovine fetal fibroblast cells were used as untransfected controls.




**Supplementary Figure S7. Effect of site-specific integration and excision on expression of fire luciferase (FLuc) gene in the pre-defined safe harbor.** (A) Luciferase activities were measured on stably-transfected bovine fetal fibroblast cells at different passages before and after Cre/Dre-mediated excision. Results were expressed as a percentage of luciferase values of stably-transfected cells at passage 1. The decline of luciferase activity during multiple passages before and after excision of selection markers was not significant ($P > 0.05$). Error bars denote SEM. (B) Western blot analysis of stably-transfected bovine fetal fibroblast cells at different passages before and after Cre/Dre-mediated excision. The blot was probed with polyclonal antibodies to firefly luciferase or GAPDH.

**Supplementary Table S1. Primers used in SOE-PCR, gene amplification, screening for site-specific integrations, and half-nested inverse PCR.**

| Primer | Sequence | Restriction site | Template |
|---|---|---|---|
| LAR_F1 | CCAAGCTTCCATGGTAGCTAGCATAACTTCGTATAGCATACATTATACGAAGTTATAGG | HindIII, NcoI, NheI | |
| LAR_R1 | GGTGGCCCTATAACTTCGTATAATGTATGCTATACGAAGTTATGCTAGCTACCATGGAAGC | | |
| LAR_F2 | TTATAGGGCCACCATGCCCGCCGTGACCG | | |
| LAR_R2 | CGACGTCGACTAACTTTAAATAATTGGCATTATTTAAAGTTAAGATGTAGGTCACGGTC | SalI | |
| No_linker_F | TCGACTCCATGGCGCG | SalI | |
| No_linker_R | GATCCGCGCCATGGAG | NcoI, BamHI | |
| (Gly4Ser)3_F | TCGACGGTGGCGGTGGCTCGGGCGGTGGTGGGTCGGGTGGCGGCGGTTCCATG | SalI | |
| (Gly4Ser)3_R | GATCCATGGAACCGCCGCCACCCGACCCACCACCGCCCGAGCCACCGCCACCG | NcoI, BamHI | |
| P2A_F | TCGACGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCCATG | SalI | |
| P2A_R | GATCCATGGGTCCAGGGTTCTCCTCCACGTCTCCAGCCTGCTTCAGCAGGCTGAAGTTAGTAGCTCCGCTTCCG | NcoI, BamHI | |
| LA35R_F1 | CATGCCATGGTAGCTAGCATAACTTCGTATAGCATACATTATACGAAGTTATCGCGCCC | NcoI, NheI | |
| LA35R_R1 | CCTTGGGCTCCCCGGGCGCGATAACTTCGTATAATGTATGCTATACGAAGTTATGCT | | |
| LA35R_F2 | GGAGCCCAAGGGCACGCCCTGGCACCTAACTTTAAATAATGCCAATTATTTAAAGTTAG | | |
| LA35R_R2 | CATGCCATGGTGGCTAACTTTAAATAATTGGCATTATTTAAAGTTAGGTGCCAGGGCG | NcoI | |
| BGHpA_F | CTAGCTAGCTGTGCCTTCTAGTTGC | NheI | pcDNA3.1(+) |
| BGHpA_R | TATCTCGAGCCATAGAGCCCAC | XhoI | pcDNA3.1(+) |
| MCS_F1 | CCATCTTAAGTAACTTTAAATAATGCCAATTATTTAAAGTTATCGCGATCGCGACCGCT | AflII, PvuI | |
| MCS_R1 | TGGGTTGGCGCGCCAACCTTAATTAAGGCTAGACTAGTCTAGCCGCTCGAGCGGTCGCG | XhoI, SpeI, PacI, AscI | |
| MCS_R2 | GCAGCTTAAGATAACTTCGTATAATGTATGCTATACGAAGTTATCCCAAGCTTGGGTTGGCGC | HindIII, AflII | |
| Chr19_vec1 | ATGGGCGTGGATAGCGGTTTG | | HEK293 genomic DNA |
| Chr19_vec2 | TAGGCGTGTACGGTGGG | | HEK293 genomic DNA |
| Chr19_fwd1 | CCAAGCGTCATCAGAAGTCCAACGAC | | HEK293 genomic DNA |
| Chr19_fwd2 | ATAGTCCCAGCGACAGTGAGCAATTC | | HEK293 genomic DNA |
| Chr19_rev1 | TTGTGTGATTCTGCTGCCTTACACCA | | HEK293 genomic DNA |
| Chr19_rev2 | GCAGGGTAGCAGATGGAACACTTAGC | | HEK293 genomic DNA |
| BFF2_vec1 | GCAGTAGCGTGGGCATTT | | BFF gemonic DNA |
| BFF2_vec2 | GATGTAGGTCACGGTCTCGAAGCC | | BFF gemonic DNA |
| BFF2_locus1 | GGGGTTTGTTGTTGTTTAGTC | | BFF gemonic DNA |
| BFF2_locus2 | GTCGGACACGACTGAGCGACTTC | | BFF gemonic DNA |
| Excision_F | GAGGTGGCGGTAGTTGAT | | BFF genomic DNA |
| Excision_R | CGAGACCGTGACCTACAT | | BFF genomic DNA |
| HNI_F1 | CGTTCTGGCTCCTCATA | | BFF gemonic DNA |
| HNI_F2 | CAACATCGTGCTTGGG | | BFF gemonic DNA |
| HNI_R | TGTAGGTCACGGTCTCG | | BFF gemonic DNA |