Gene 642 (2018) 188–198

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Evaluating different DNA binding domains to modulate L1 ORF2p-driven site-specific retrotransposition events in human cells

Catherine M. Ade[a], Rebecca S. Derbes[b], Bradley J. Wagstaff[b], Sara B. Linker[c], Travis B. White[d,1], Dawn Deharo[e], Victoria P. Belancio[e], Zoltán Ivics[f], Astrid M. Roy-Engel[b,*]

[a] Department of Cellular and Molecular Biology, Tulane University, USA
[b] Tulane Cancer Center SL-66, Dept. of Epidemiology, Tulane University Health Sciences Center and LCRC, 1700 Tulane Ave., New Orleans, LA 70112, USA
[c] Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 N Torrey Pines Road, La Jolla, CA 92037-1002, USA
[d] Sloan Kettering Institute for Cancer Research, New York, NY 10065, USA
[e] Department of Structural and Cellular Biology, Tulane University School of Medicine, Tulane Cancer Center, Tulane Center for Aging, New Orleans, LA 70112, USA
[f] Division of Medical Biotechnology, Paul-Ehrlich-Institute, Langen, Germany

ARTICLE INFO

Keywords:
Alu
DNA binding domain
Fusion protein
LINE-1
ORF2
Retrotransposition
Target site

ABSTRACT

DNA binding domains (DBDs) have been used with great success to impart targeting capabilities to a variety of proteins creating highly useful genomic tools. We evaluated the ability of five types of DBDs and strategies (AAV Rep proteins, Cre, TAL effectors, zinc finger proteins, and Cas9/gRNA system) to target the L1 ORF2 protein to drive retrotransposition of Alu inserts to specific sequences in the human genome. First, we find that the L1 ORF2 protein tolerates the addition of protein domains both at the amino- and carboxy-terminus. Although in some instances retrotransposition efficiencies slightly diminished, all fusion proteins containing an intact ORF2 were capable of driving retrotransposition. Second, the stability of individual ORF2 fusion proteins varies and difficult to predict. Third, DBDs that require the formation of multimers for target recognition are unlikely to modify targeting of ORF2p-driven insertions. Fourth, the more components needed to assemble into a complex to drive targeted retrotransposition, the less likely the strategy will increase targeted insertions. Fifth, abundance of target sequences present in the genome will likely dictate the effectiveness and efficiency of targeted insertions. Lastly, the cleavage capabilities of Cas9 (or a Cas9 nickase variant) are unable to substitute for the L1 ORF2 endonuclease domain functions, suggestive that the endonuclease domain has alternate functions needed for retrotransposition. From these studies, we conclude that the most critical component for the modification of the human L1 ORF2 protein to drive targeted insertions is the selection of the DBD due to the varying functional requirements and impacts on protein stability.

1. Introduction

Different non-LTR retroelements show variations in the site insertion preference ranging from site specific to almost random insertion dispersed throughout the genome (Fujiwara, 2015; Zingler et al., 2005). For example, the R2 non-LTR retroelement from *Bombyx mori* inserts specifically into the 28S ribosomal DNA in its host genome (Burke et al., 1987). R2 has been a great model because its specific insertional preference allowed for the development of *in vitro* studies leading to deciphering many aspects of the mechanism of insertion of this element (Luan et al., 1993; Christensen et al., 2006). DNA binding is the first step in the integration reaction of retroelements and is likely key for target specific non-LTR retroelements such as the R2 retrotransposons. Although, the exact protein sequence requirements to confer site specificity are ill defined, recent studies indicate that the DNA binding domain (DBD) of the active protein of site specific elements is critical for targeting of insertions (Thompson and Christensen, 2011). Analysis of proteins of two clades of R2 retrotransposons demonstrated that many contain zinc fingers (ranging from 1 to 3) in the amino terminus upstream of the endonuclease domain. In the case of R2-A, three zinc fingers and a myb domain were shown to bind specific sequences 5′ of the target site (Thompson and Christensen, 2011).

The introduction of DBDs such as zinc fingers (ZFs) and TAL effectors has been widely successful in targeting a variety of proteins, such

Abbreviations: AAV, Adeno-Associated Virus; CRISPR, clusters of regularly interspaced short palindromic repeats; DBDs, DNA binding domains; gRNA, guide RNA; HAC, homology Arm Cassette; HRP, Horse Radish Peroxidase; PAM, Protospacer Adjacent Motif; TAL, Transcription activator-like; ZFs, Zinc fingers

* Corresponding author at: Tulane Cancer Center, Tulane University Health Sciences Center and LCRC, 1700 Tulane Ave., New Orleans, LA 70112, USA.
E-mail address: aengel@tulane.edu (A.M. Roy-Engel).
[1] Current address.

https://doi.org/10.1016/j.gene.2017.11.033
Received 6 November 2017; Accepted 11 November 2017
Available online 14 November 2017

as nucleases and transcription factors, to specific genomic locations (Straubeta and Lahaye, 2013; Klug, 2010). The CRISPR/Cas9 targeting system has been also used in targeting strategies. For example, expression of specific genes can be upregulated through the fusion of the endonuclease-deficient variant of the Cas9 protein, dCas9, to the VP64 activation domain (Maeder et al., 2013) or downregulated by using dCas9-KRAB fusion proteins (Gilbert et al., 2013).

This approach to alter targeting preference is not limited to cellular proteins, but has also been applied to viruses and non-human mobile elements (see (Kovac and Ivics, 2017) for a recent review). For example, the addition of ZF was used to modify the integration pattern of retroviruses, such as HIV (Tan et al., 2004) and MLV (Lim et al., 2010). Furthermore, insertion site preference of non-human mobile elements, such as *Sleeping Beauty*, can be modified or targeted to specific sites by the creation of chimeric proteins or fusion proteins with different DNA binding proteins, including ZFs (Yant et al., 2007; Ivics et al., 2007; Voigt et al., 2012; Ammar et al., 2012). Similarly, the addition of a Gal4 DNA binding domain or a TAL effector domain to the *piggyBac* transposase effectively biased insertion preference (Owens et al., 2012; Owens et al., 2013).

The currently active human mobile element, L1, inserts into the consensus target sequence 5′TTTT/AA3′(Jurka, 1997; Gilbert et al., 2002) using an RNA intermediate that integrates *via* a target-primed reverse transcription (TPRT) mechanism (Luan et al., 1993). The TPRT step of retrotransposition requires the target site to be cut by the endonuclease function of the L1 ORF2 protein (ORF2p). The abundance of the target sites throughout the genome contribute to the observed dispersed pattern of *de novo* L1 inserts. In addition to L1 mobilization, the L1 ORF2p can be used *in trans* to mobilize other cellular RNAs including the Alu retrotransposons (Dewannieux et al., 2003). Similarly to L1, Alu elements, that only require the L1 ORF2p for efficient retrotransposition, also insert at the AT-rich consensus target and are found throughout the genome. In addition to the nicking preference of the ORF2 endonuclease domain encoded in ORF2p, structural parameters of the DNA target site also play an important role in the integration (Repanas et al., 2007). Thus, it is considered that the L1 ORF2p is likely the main determinant of insertion site selection for both L1 and Alu elements.

Experimental alteration of target site selectivity may allow for the creation of tools that would help advance the knowledge in the integration mechanism of the currently active human mobile elements. A previous attempt to alter L1 insertion preference demonstrated that modifying the endonuclease domain of ORF2p failed to alter insertion preference of L1 (Repanas et al., 2007). This suggests that an alternate strategy may be needed to effectively promote the targeted integration of L1. Because the DNA interaction mediated by ORF2p is a critical determinant of insertion site selection, modifying the DNA binding preference of ORF2p by adding a DNA binding domain becomes a feasible alternative strategy.

1.1. Selection of DNA binding domains

We selected to evaluate five different types of DNA binding domains (DBDs) that have been previously shown to redirect binding preference when fused to proteins of interest. The five different types include: DBDs from viral proteins, a TAL effector, the bacteriophage Cre recombinase, designer ZF proteins and various adaptations of the Cas9 protein and guide RNAs from the bacterial CRISPR system. First we selected the Rep protein encoded by the Adeno Associated Virus (AAV) fused to two different oligomerization domains (LZ and TZ) that have been shown to interact with the AAVS1 site of human chromosome 19 (Owens et al., 1993; Cathomen et al., 2000), and to redirect *Sleeping Beauty* transposon insertions near Rep binding sites in the human genome (Ammar et al., 2012). In addition, a transcription activator-like (TAL) effector targeting the human AAVS1 locus (Sanjana et al., 2012) was also included. Targeting of the AAVS1 locus is ideal as it is considered a “safe harbor” for genome manipulation (Ammar et al., 2012; Hockemeyer et al., 2011) and is surrounded by multiple AT-rich regions that can serve as L1 ORF2p endonuclease cleavage sites. Our third selection was Cre recombinase that selectively binds *LoxP* sequences and is widely used for implementing and evaluating targeted mutation and integration strategies (Murray et al., 2012; Van Duyne, 2015). Fourth, due to their wide range of uses, flexibility of design and their common implementation in gene editing strategies, we selected two engineered ZF proteins that target sequences in the green fluorescent protein (GFP) (Handel et al., 2009). Our fifth and final selection consisted on the implementation of several variations of the CRISPR/Cas9 system by using the inactive dCas9 as a DBD or by substituting the ORF2p endonuclease function with the Cas9 endonuclease function.

2. Materials and methods

2.1. Constructs

The plasmids *p*AluYa5-*neo*^{TET} (Kroutter et al., 2009), pBS-Ya5r-escue-A70D-SH (Wagstaff et al., 2012) and pBS-L1PA1$_{CH}$notag, pBudORF2$_{CH}$ (Wagstaff et al., 2011) have been previously described. pBudORF2$_{CH}$ was used as the base plasmid utilized to create all the ORF2 fusion constructs (except for the pcDNALZ-ORF2 and pcDNATZ-ORF2). This plasmid contains the fully codon optimized ORF2p from L1$_{RP}$, cloned into the expression vector pBudCE4.1 (Invitrogen/Thermo Scientific) under control of the CMV promoter. Additionally, a glycine helical peptide linker (GHL) (KLGGGAPAVGGGPKAADK) commercially synthesized by Genscript, was cloned into the *Pst*I and *Bam*HI sites creating the new plasmid pBudO2$_{CH}$ + L$_{GHL}$. The restriction sites *Hin*dIII, *Pst*I, and *Sal*I were added to the 5′ end of the GHL linker to allow N-terminal cloning procedures. Table 1 summarizes the ORF2p fusion constructs evaluated in this study.

Table 1
ORF2p fusion protein constructs.

Plasmid	DNA binding domain	Target sequence
pcDNALZ-ORF2	AAVS1 REP binding leucine zipper	5′-CAGCGAGCGAGCGAGC-3′ (RBE)
pcDNATZ-ORF2	AAVS1 REP binding leucine zipper	5′-CAGCGAGCGAGCGAGC-3′ (RBE)
pBud-N-Cre-ORF2	Cre	5′-ATAACTTCGTATAgcatacatTATACGAAGTTAT-3′
pBud-C-Cre-ORF2	Cre	5′-ATAACTTCGTATAgcatacatTATACGAAGTTAT-3′
pBud TAL-ORF2	TAL effector	5′-TGTCCCCTCCACCCCACA-3′
pBud-2.18-ORF2	Zn finger	EGFP: 5′-GAGGACGGC-3′
pBud-2.17-ORF2	Zn finger	EGFP: 5′-ATCCGCCAC-3′
pBud Cas9-ORF2endo–	CRISPR system	Determined by the gRNA
pBud Cas9-RTCYS	CRISPR system	Determined by the gRNA
pBud nickase-ORF2endo–	CRISPR system	Determined by the gRNA
pBud nickase-RTCYS-	CRISPR system	Determined by the gRNA

2.1.1. Construction of the AAVS1 Rep proteins and TAL effector–ORF2p fusion plasmids

pcDNALZ-ORF2 and pcDNATZ-ORF2 contain the fully codon optimized ORF2p from $L1_{RP}$ cloned in frame downstream of the LZ (leucine zipper) or TZ (engineered leucine zipper) Rep binding motifs of the Adeno-Associated virus (AAV) and the transcriptional activation domain of VP16 (AD). The ORF2 coding sequence was PCR amplified with primers that would add 5′ *Apa*I and 3′ *Age*I restriction sites for cloning into pcDNARepLZAD and pcDNARepTZAD (Cathomen et al., 2000).

pBud TAL-ORF2 contains the TAL effector targeting the AAV site fused upstream of the ORF2 coding sequence. Complementary oligonucleotides designed to contain a *Bsm*BI site (which provides compatible ends with *Sal*I) were cloned into the *Ear*I site of the hAAVS1 1 L TALEN plasmid (Addgene# 35431 (Sanjana et al., 2012)). The TAL effector was excised from the modified plasmid using the *Bsm*BI and *Sac*I sites and cloned into the. *Sac*I and *Sal*I of $pBudO2_{CH}$ + L_{GHL} to introduce the TAL effector upstream of the flexible peptide linker GHL and downstream of the CMV promoter.

2.1.2. Construction of the Cre–ORF2p fusion plasmids

pBud-N-Cre-ORF2 and pBud-C-Cre-ORF2 contain Cre either at the N-terminus or the C-terminus of the ORF2p. Cre was PCR amplified from pBS185 CMV-Cre (Addgene plasmid#11916 (Sauer and Henderson, 1990)) with the following primers containing compatible ends for cloning the Cre recombinase protein upstream or downstream of the L1 ORF2p into the $pBudO2_{CH}$ + L_{GHL} plasmid: F*Pst*I: 5′-GAACGGATCTGCAGGATCAGGATCAGGCATGTCCAATTTACTGAC -3′ and R*Sal*I: 5′- GCAGCTCAGTCGACCCTAATCGCCATCTTCCAG -3′. For the N-terminally fused construct 5′ *Pst*I and 3′ *Sal*I were used and for cloning the C-terminally fused Cre, 5′ the *Bam*HI and 3′ *Eco*RI were used for cloning into pBud- $ORF2_{CH}$.

2.1.3. Construction of the zinc finger-ORF2p fusion plasmids

pBud $ZF_{2.17}$-ORF2, and pBud $ZF_{2.18}$-ORF2 were created by introducing the zinc fingers upstream of the L1 ORF2p. These zinc fingers target specific sequences present in the engineered green fluorescence protein (EGFP) gene (Handel et al., 2009). The zinc fingers were PCR amplified and cloned in frame into the *Sal*I site of $pBudO2_{CH}$ + L_{GHL} plasmid using primers that introduced flanking *Sal*I sites.

2.1.4. Construction of the CRISPR/Cas9 fusion protein plasmids

The Cas9 and the nickase proteins were fused upstream of an L1 ORF2p lacking endonuclease activity and the catalytic inactive dCas9 was fused to the active ORF2p. To be able to fuse these proteins upstream of the ORF2p additional restriction sites were added to $pBudO2_{CH}$ + L_{GHL} plasmid. Complementary oligonucleotides containing the *Cla*I, *Pac*I, and *Bst*EII restriction sites were cloned into the *Hind*III and *Sal*I sites of $pBudO2_{CH}$ + L_{GHL}, creating $pBudO2_{CH}$ + CL_{GHL}. The oligonucleotides used were as follows:

Top Hind newRE Sal: 5′- Phos -AGCTTatcgattaattaagaattccacc atgctggtaaccG -3′ and Bot Hind newRE Sal: 5′- Phos-TCGACggttacc agcatggtggaattcttaattaatcgatA -3′.

The pBudRTCYS + CL_{GHL} construct was created by PCR amplifying the reverse transcriptase and Cys domains (RTCYS) of the L1 ORF2p with added *Eco*RV and *Bam*HI sites using the following two primers: F EcoRV RTCysCH: 5′- AAGATATCATGAAGGCCGAGATCAAGATGTT CTTC -3′ and R BamHI ORF2CH: 5′- GGATCCGTTGCCGCCGATCAGGC -3. The RTCYS PCR product was used to substitute the full length ORF2 coding sequence of the $pBudO2_{CH}$ + CL_{GHL} plasmid.

To generate the ORF2p endonuclease double mutant constructs, the first 500 bp of the ORF2 coding sequence of the $pBudO2_{CH}$ + CL_{GHL} plasmid was exchanged with the sequence of the endonuclease double mutant to create the pBud L1 endo– + CL_{GHL} construct. The 500 bp PCR product from the endonuclease double mutant (ORF2 D205, H230) (Wallace et al., 2008) was amplified with the following primers containing *Xba*I and *Eco*RV sequences: F EcoRV ORF2ch: 5′-ATCATGACCGGCAGCACCAGCC -3′ and R XbaI endo (−) ORF2ch: 5′ CTCTAGATTCTCCAGCTTGTTGGCGTACAGG -3′.

All of the following constructs were built using the Gibson Assembly kit (NEB) following the manufacturer's recommended protocol:

pBud Cas9-RTCYS, pBud Cas9- L1 endo–, pBud Nickase- L1 endo– and, pBud Nickase- RTCYS-constructs were created by PCR amplifying the Cas9 nuclease from the spCas9 plasmid (Addgene# 48137, (Cong et al., 2013)) and the D10A Cas9 nickase protein from the hCas9_D10A plasmid (Addgene# 41816, (Mali et al., 2013)) using the following primers: F-Cas9/nickaseGA 5′-ACCCAAGCTTatcgattaatCGACCATGGA TAAAAAGTATTC-3′ and R Cas9 GA: 5′-CCAGCTTGTCGACGGTTACCT TGACTTTCCTCTTCTTCTTG-3′. The PCR products were assembled with a *Pac*I and *Bst*EII digest of the appropriate plasmid, either pBud L1 endo– + CL_{GHL} construct or pBudRTCYS + CL_{GHL}.

To build the pBud dCas9-ORF2, the dCas9 protein (catalytically inactive D10A, H841 double mutant) was PCR amplified from the plasmid pcDNA dCas9-VP64 (Addgene #61422, (Konermann et al., 2015)) using the following primers: F-dCas9-VP64-GA 5′-ACCCAAGCT TatcgattaatAACGAGATGGCCAAGGTGGACGA-3′ and R dCas9VP64 GA: 5′-ccagcttGTCGACggttaccttCTTGTACAGCTCGTCCATGCC-3′. The PCR products were assembled with a *Pac*I and *Bst*EII digest of the $pBudO2_{CH}$ + CL_{GHL} plasmid.

To build the plasmid pBud MS2-ORF2, the MS2 protein sequence from the MS2-HB plasmid (Addgene# 35573, (Konermann et al., 2015)) was PCR amplified using the following primers: F-MS2 GA: 5′–ACCCAAGCTTatcgattaattCGACCATGGCTTCTAACTTTAC-3′and R MS2 GA: 5′- CCAGCTTGTCGACGGTTACCTTGTTAATTAAGGAGTT TGCTGCG -3′.

2.1.5. Construction of the gRNA plasmids

Two guide RNA (gRNA) backbones were used for the CRISPR/Cas9 gRNA plasmids: MLM 3636 (Addgene# 43860) and sgRNA MS2 (Addgene# 61424) (Konermann et al., 2015). The sgRNA MS2 backbone was used in order to create the gRNAs that utilized the MS2 protein, as this construct contains two MS2 binding loops encoded in the gRNA sequence. Five gRNAs were created using the MLM 3636 backbone, and three gRNAs were designed to clone into the sgRNA MS2 backbone:

gRNA2, was designed to target a unique site on chromosome 8 (5′-CTGATAAATAGTCAGTTAAA-3′) using the available website www.ZiFit.partners.org. This sequence contains the appropriate sequences to both direct site-specific Cas9 cleavage and provide the sequences for A-tail annealing during TPRT.

Three gRNAs were designed to the 5′ UTR of genomic L1 elements: gRNA 551, gRNA 765, and gRNA 892 (kind gift Prescott Deininger). These three gRNAs targeted the sequences: 5′-GCCTCTGTAGGCTCCA CCTC-3′, 5′-AGCAGGGGCACACTGACACC-3′ and 5′-GTAGATAAAACC ACAAAGAT-3′, respectively. An additional gRNA was designed to the 3′ end of genomic L1 elements, with the target sequence of 5′-GTGGGT GCAGCGCACCAGCA-3′. The guide RNAs have been previously tested as a pool and observed to target Cas9 to the desired genomic location (Prescott Deininger, *personal communication*).

2.2. Evaluation of the targeted cleavage capability of Cas9-fusion proteins

The Homology Arm Cassette (HAC, details provided in Supplemental Data S1) contains a left arm with 354 bp and a right arm with 505 bp regions with sequence homology to intron 20, exon 21and intron 21 of the *WRN* gene (Werner syndrome RecQ-like helicase). The homology arms flank a splice acceptor upstream of a blasticidin resistance gene. Cleavage of *WRN* exon 21 by gRNA directed Cas9 endonuclease stimulated recombination to integrate the HAC into *WRN* and confer blasticidin resistance. HeLa cells were transiently transfected with 1 μg Cas9, nickase or the Cas9-fusion protein constructs, 1 μg of the gRNA construct plus 300 ng of the HAC using Lipofectamine plus (Life Technologies) following the manufacturer's recommended

protocol. The following day, the transfected cells were switched to media with blasticidin (2 μg/mL) and grown under selection for 12–14 days until individual colonies were observed. To confirm recombination of the HAC into *WRN*, individual blasticidin-resistant colonies were evaluated by semi-nested PCR using combinations of the following primers:

Red set: FWRN Nest1: 5′-GAGCCATGTAGTATATTATGGC-3′/FWRN Nest3: 5′-ATGTTTCATCCCACCATCTTTAATGAG-3′ and R106Blast: 5′-ATCTCATGCTGGAGTTCTTCGC-3′.

Green set: F107Blast: 5′- ATGGGGATGCTGTTGATTGTAGCCG-3′ and RWRN Nest1: 5′-GCCAACAAACTACTTGTTGAGTAC-3/RWRN Nest3: 5′-ATAGTAAACAAGAGTCAAATAGGGA-3′.

The guide RNA was designed to target a unique site (5′-GTCATAG CTACCATAGCTTT-3′) in exon 21 of the WRN gene using the available website http://zifit.partners.org.

2.3. Alu retrotransposition assay, Alu rescue assay and insert analysis

Alu retrotransposition and Alu rescue assays were performed in HeLa cells as previously described (Krouter et al., 2009; Wagstaff et al., 2012). The Alu plasmid *p*AluYa5-*neo*^{TET} is an engineered Alu tagged with a retrotransposition indicator cassette for evaluating the Alu retrotransposition in culture. This cassette is designed so that the selectable marker (neomycin) expresses only following an RNA-mediated retrotransposition event of the spliced tagged Alu RNA. To evaluate insertion preference, the Alu Rescue assay using the modified *p*AluYa5-*neo*^{TET} construct, pBS-Ya5rescue-A70D-SH. Recovery of Alu inserts followed our previously published detailed protocol (Ade and Roy-Engel, 2016). For the Alu rescue assay, neomycin resistant colonies from individual flasks were pooled separately and processed independently to be able to distinguish between potentially "identical" independent insertions from those representing the recovery of the same insertion multiple times. The plasmids containing the recovered Alu inserts were sent for sequencing to Elim Biopharmaceuticals, Inc., Hayward, California. DNA Star Lasergene 10 software was utilized to analyze the flanking genomic sequences. The genomic position of each rescued Alu insertion was determined using BLAT (http://genome.ucsc.edu) search using the human genome reference (GRCh37hg19).

2.4. Requirements of a "targeted insertion"

To consider an Alu insertion as a potential "targeted insertion" the following characteristics need to be met: 1- presence of the target sequence within 2 kb (4 kb window) of the Alu insert and 2- the identified target sequence can only differ from the canonical target sequence by two nucleotides (2 mismatches) with no gaps tolerated.

2.5. Determination of target site frequency in the genome

The identification of instances and respective locations of sequence motifs was implemented in python using the regex library. The sequence of interest and the reverse complement were identified by searching the human hg19 reference against the sequence of interest with a maximum of 2 errors (*i.e.* mismatches) (see Table 5).

2.6. Motif Analysis

Motifs were identified near the integration sites for each DNA binding domain using the hg19 reference genome for sequence information. A directed query was performed by scanning regions within 4 kb (2 kb upstream and downstream) or within 40 bp (20 bp upstream - 20 bp downstream) for the expected motif which was converted to a motif table using the seq2profile.pl script in HOMER (Heinz et al., 2010) using two mismatches. An unbiased analysis was then performed using the findMotifsGenome.pl script in HOMER to search for all motifs present within either 200 or 20 bp surrounding the insertion site. The matrix file from the top motif identified within the sequence was then imported into STAMP to visualize the sequence logo.

2.7. Creation of a HeLa-LoxP cell line and HeLa-EGFP cell line

To evaluate the ORF2-Cre fusion proteins, we generated the HeLa-*Lox*P cell line by stably integrating the *LoxP* target sequence using a previously developed *Sleeping beauty* strategy and grown under blasticidin selection. To create the HeLa-EGFP cell line, the pExT plasmid (Addgene# 36889 (Tasic et al., 2012)) was stably integrated into HeLa cells and kept under hygromycin selection.

2.8. Western blot analysis

Two to four T75 flasks of HeLa (4×10^6/flask) were transiently transfected with 5 μg of plasmid per flask using Lipofectamine plus (Life Technologies) following the manufacturer's recommended protocol. Cells were harvested 24 or 48 h post-transfection using total lysis buffer, (50 mM Tris, 150 mM NaCl, 10 mM EDTA, 0.5% Triton × 0.5%, pH = 7.2). Protein samples were sonicated three times for 10 s using a Microson Ultrasonic Cell Disrupter, and incubated on ice between each sonication. The protein concentrations of the lysates were determined using the Bradford protein assay (BioRad #500–0006) using a Bovine Serum Albumin (BSA) standard curve. Equivalent amounts of protein for each sample were mixed with Laemmli buffer plus 2-β-mercaptoethanol and boiled for 15 min. A total of 5 μg of the protein extracts were electrophoresed on NuPage Bis-Tris gels (Invitrogen/Thermo Scientific) and transferred to a nitrocellulose membrane using the iBlot gel transfer system (Invitrogen/Thermo Scientific) for 8 min, using the P3 setting. Membranes were blocked for at least 1 h in PBS pH 7.4, 0.05% Tween 20, 5% non-fat dry milk (BioRad) at room temperature. The membrane was incubated with primary antibody overnight at 4 °C. Custom polyclonal rabbit antibodies were generated against amino acids 159–172 of the mouse L1 ORF2 endonuclease domain (deHaro et al., 2014). The secondary HRP-donkey anti-rabbit (Santa Cruz Biotechnology Inc.; sc-2317) was diluted to 1:5000 in PBS pH 7.4, 0.05% Tween 20, 3% non-fat dry milk (Biorad) and incubated for at least 1 h at room temperature. Signals were detected using the SuperSignalWest Pico Chemiluminescent substrate (Pierce, Rockford, IL). The protein standard was visualized using the precision protein StrepTactin-HRP conjugate (BioRad).

2.9. Supplemental information

Supplemental information (Tables S1–S6) and can be found with this article online.

3. Results

3.1. Most DBD-ORF2 fusion proteins support efficient retrotransposition of a tagged Alu

One concern is that the addition of sequences to the ORF2p may interfere with its function. Because Alu mobilization only requires expression of the L1 ORF2p (Dewannieux et al., 2003), the Alu retrotransposition assay was selected to evaluate the function of the ORF2 fusion proteins. We cotransfected HeLa cells with the tagged Alu construct in addition to either the ORF2 or the selected DBD-ORF2 fusion protein construct. In our tissue culture assay system, all fusion proteins that contained a full-length functional ORF2 supported Alu retrotransposition with efficiencies either comparable to the ORF2p or reduced by about 20 to 30% (Table 2) but none significantly different from ORF2p (Unpaired Student *t*-test). However, Alu retrotransposition significantly diminished by several orders of magnitude or was completely abolished when the fusion protein contained an endonuclease mutant or truncated ORF2 (Table 2).

Table 2
Retrotransposition efficiency of Alu driven by the ORF2p fusion proteins.

ORF2p fusion	Cell line	Retrotransposition rate ($\times 10^4$)	Relative to ORF2p (%)
ORF2 (HeLa control)	HeLa	8.93 ± 0.13	100
LZ-ORF2	HeLa	9.46 ± 0.15	106.0 ± 1.6
TZ-ORF2	HeLa	9.49 ± 0.11	106.3 ± 1.2
TAL-ORF2	HeLa	8.89 ± 0.22	99.6 ± 2.4
ORF2 (HeLa *LoxP* control)	HeLa *LoxP*	9.68 ± 0.19	100
Cre-ORF2	HeLa *LoxP*	9.84 ± 0.13	101.7 ± 1.3
ORF2-Cre	HeLa *LoxP*	9.67 ± 0.53	99.9 ± 5.5
ORF2 (HeLaEGFP control)	HeLa EGFP	7.13 ± 0.40	100
2.18-ORF2	HeLa EGFP	6.45 ± 0.25	90.5 ± 3.9
2.17-ORF2	HeLa EGFP	8.55 ± 2.03	119.9 ± 23.8
ORF2 (HeLa control)	HeLa	8.93 ± 0.13	100
dCas9-ORF2[a]	HeLa	7.32 ± 0.14	82.00 ± 2.33[b]
dCas9 + MS2-ORF2[c]	HeLa	6.38 ± 0.12	71.44 ± 9.64[b]
Cas9 Endo–	HeLa	0	0
Cas9 RTCYS	HeLa	0	0
Nickase Endo–	HeLa	0.0073 ± 0.0073	0.082 ± 0.082[b]
Nickase RTCYS	HeLa	0.0109 ± 0.0109	0.122 ± 0.122[b]

[a] Average from combined assays using different gRNAs (551, 765, 892, and 3′L1).
[b] Significantly different from ORF2p control, $p < 0.00001$ (Unpaired Student *t*-test).
[c] Average from combined assays using different gRNAs (551_{MS2}, 765_{MS2}, and 892_{MS2}).

3.2. Targeting the AAVS1 locus by fusing the AAV REP binding domains or a TAL effector to ORF2p

We fused the adeno-associated viral (AAV) REP proteins which contain either the dimerization domain LZ or tetramerization domain TZ (Cathomen et al., 2000; Ammar et al., 2012), as well as a previously validated TAL effector (designed to also target the AAVS1) upstream of the ORF2p. Our expectation was that these fusion proteins would drive retrotransposition of Alu inserts to the AAVS1 locus located on the q arm of chromosome 19 (Kotin et al., 1990). One concern when creating recombinant proteins is the possibility that the fusion protein becomes unstable (*i.e.* degraded), or partially processed by cellular enzymes and thus limits the ability to drive targeted Alu insertions. Western blot analyses were performed to determine the expression of ORF2-fusion proteins and evaluate the expected molecular weight. Expression analyses determined that proteins of the predicted sizes were being expressed from the TZ-ORF2, LZ-ORF2, and TAL-ORF2 fusion constructs (149 kDa for ORF2p, 190 kDa for both Rep-TZ-ORF2p and Rep-LZ-ORF2p and 250 kDa for TAL-ORF2p; Fig. 1A).

To evaluate targeting capability of these fusion proteins, we used our previously described Alu rescue assay to recover and determine the location of the *de novo* Alu insertions, (Wagstaff et al., 2012; Ade and Roy-Engel, 2016). We recovered 74 Alu insertions driven by Rep-LZ-ORF2p, 15 insertions driven by Rep-TZ-ORF2p, and 43 insertions driven by TAL-ORF2p (Fig. 1B). All insertions showed the signature characteristics of Alu retrotransposition events (direct repeats of average length 14 bp, an A-tail and insertions occurred at the canonical endonuclease site (TTTT/AA). Only one of the Alu inserts driven by Rep-TZ-ORF2p inserted into chromosome 19, but nowhere remotely near the known AAVS1 target sequence. Motif analysis of the insertion sites showed there was no significant enrichment of the expected target sequence (motif) and the top consensus sequences for the DNA binding domain showed no similarity with the expected DBD (Table 3 and Supplemental Table S1). Detailed genomic locations are listed in Supplemental Table S2. Interestingly, two recovered Alu insertions driven by TAL-ORF2p shared between 60 and 70% sequences similarity to the AAVS1 locus (Table 4). These two Alus inserted 38 bp and 140 bp away from the mismatched target site. However, due to the high number of mismatches and gaps in the alignment, we consider that these two events did not meet our requirements of a targeted insertion. The TAL-ORF2p fusion had no substantial effect on altering insertion preference of the tagged Alu. Overall, the analysis was negative indicates that these

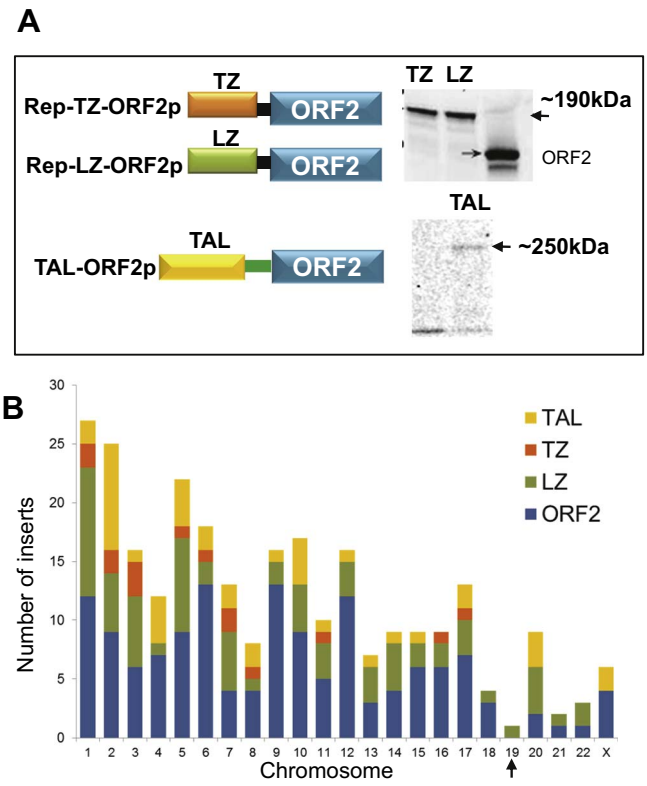

Fig. 1. The Rep-TZ-ORF2p, Rep-LZ-ORF2p and Rep-TAL-ORF2p fusion proteins. A. A schematic of the fusion proteins are shown next to Western blot analysis of protein extracts from HeLa cells transiently transfected with the corresponding plasmids. Bands of the expected sizes: ~190 kDa for the full-length Rep-TZ-ORF2p and Rep-LZ-ORF2p fusion proteins and ~250 kDa for the TAL-ORF2p fusion protein were observed. The ORF2p was visualized at approximately 149 kDa. B. Histogram of the chromosomal distribution of genomic locations of the recovered of Alu inserts driven by the TAL-ORF2p (yellow) Rep-TZ-ORF2p (orange) and Rep-LZ-ORF2p (green) fusion proteins and by the ORF2p (blue). The X axis corresponds to the chromosome number in the human reference genome, while the Y axis represents how many Alu insertions landed in each chromosome. Arrow indicates the two recovered Alu insertions landed in chromosome 19, but localized to completely unrelated locations relative to the AAVS1 locus target sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Motif analysis 2 kb distance (4 kb window).

DBD (loci)	Expected motif	Expected motifs identified within 4 kb of insert, count	Expected motifs identified within 4 kb of ORF2 insert, count	Actual consensus	Enrichment of actual consensus, *p*-value	Sequence Logo
ORF2 (140)	TTAAAA	84,026	84,026	AGGGTCAGCG	1e-09	
TZ ORF2 (15)	CAGCGAGCGAGCGAGC	0	0	AATGCWMAGA	1e-05	
LZ ORF2 (74)	CAGCGAGCGAGCGAGC	0	0	GGGACAACTG	1e-07	
TAL ORF2 (41)	TGTGGGGTGGAGGGGACA	0	1	GAGAAATTGT	1e-08	
NCreORF2 (6)	ATAACTTCGTATAGCATACATTATACGAAGTTAT	0	0	AGAGAAAGGG	1e-04	
CCreORF2 (11)	ATAACTTCGTATAGCATACATTATACGAAGTTAT	0	0	GGGCCACGTC	1e-07	
ZF 2.17 (14)	ATCCGCCAC	83	955	CACTTAGGTT	1e-05	
ZF 2.18 (11)	GAGGACGGC	61	891	GGAAAAACCC	1e-06	
gRNA 551 dCas9(37)	GAGGTGGAGCCTACAGAGGCAGGC	0	0	CRATGCCACG	1e-08	
gRNA 765 dCas9(28)	GGTGTCAGTGTGCCCCTGCTGGG	0	0	TCATCCAACG	1e-07	
gRNA 892 dCas9(38)	GTAGATAAAACCACAAAGATGGG	1	0	TGGTTCCCGG	1e-07	
gRNA 3′-L1 dCas9(8)	GTGGGTGCAGCGCACCAGCATGG	0	0	TCTATGAAGG	1e-04	
gRNA 551 MS2 (37)	GAGGTGGAGCCTACAGAGGCAGGC	0	0	CCACTTCTAA	1e-08	
gRNA 765 MS2 (28)	GGTGTCAGTGTGCCCCTGCTGGG	0	0	ACTATGATAT	1e-07	
gRNA 892 MS2 (38)	GTAGATAAAACCACAAAGATGGG	0	0	CTGTGAGCAA	1e-07	

Table 4
Two potential Alu targeting events when Alu is driven by TAL-ORF2.

	Chromosome	Potential target	% Match to target	Distance away from Alu
1	1	AA TGGG TTGGAGG AG-CA	72.2%	38 bp
2	20	TGTGGG -TGGAG TGAGGG	66.7%	140 bp

The TAL effector target sequence: 5′-TGTGGGGTGGAGGGGACA-3′.
Gray highlight indicates nucleotides matching the target sequence. Dashes represent a missing base pair.

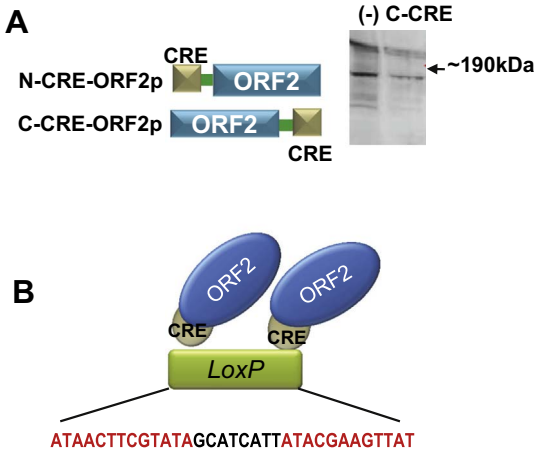

Fig. 2. The Cre-ORF2p fusion proteins. A. Schematic of the fusion proteins are shown next to Western blot analysis of protein extracts from HeLa cells transiently transfected with the empty vector (−) or the C-Cre-ORF2 plasmid. Arrow indicates the expected 190 kDa location of the fusion protein which was not detected even after multiple attempts. Other bands represent non-specific bands. B. Predicted model of the expect interaction of two Cre-ORF2p fusion proteins binding to the palindromic repeats within a *LoxP* sequence.

three fusion proteins did not meet our criteria of being able to drive Alu insertions to a specific, endogenous, single locus in the human genome.

3.3. Targeting the LoxP site by fusing the Cre recombinase to ORF2p

One of the most developed approaches for gene manipulation is the Cre-recombinase system from the P1 bacteriophage (reviewed in (Nagy, 2000)). The Cre protein recognizes a 34 bp recognition sequence, known as *LoxP* (Hamilton and Abremski, 1984). Cre was cloned upstream and downstream of the ORF2p to create C-Cre-ORF2p and N-Cre-ORF2p fusion proteins. We were unable to detect N- and C-terminal Cre ORF2p fusion proteins at their expected size, ~190 kDa by Western blot analysis with only background bands were observed (Fig. 2A shows C-Cre ORF2p as an example). However, the constructs are able to drive Alu retrotransposition in culture (Table 2). This suggests that these fusion proteins may not be completely degraded, but instead processed products retaining ORF2 activity may exist or that very little of the expressed full-length fusion protein is sufficient to efficiently drive Alu retrotransposition.

We next performed the Alu rescue assay in the created HeLa-*LoxP* cell line (see Materials and Methods) to ensure the presence of the target sequence in the human genome. This cell line contains the stable integration of the standard *LoxP* sequence, which is composed of two 13 bp palindromic Cre binding sites, separated by an 8 bp spacer sequence. Thus in this system, we expect that the integrated *LoxP* site could potentially bind two Cre-ORF2 fusion proteins (Fig. 2B). All 11 of the recovered Alu inserts showed the signature characteristics of retrotransposition. However, further analysis of the genomic sequence flanking the insertion verified that the *LoxP* or a similar sequence (allowing for 2 mismatches) was absent at the site of insertion or near its vicinity (within 700 bp, the average length of a sequencing read). Again, motif analysis of the insertion sites showed there was no significant enrichment of the expected target sequence (Table 3). Detailed genomic locations are listed in Supplemental Table S3. It is possible that the close proximity of the two fusion proteins interfered with targeted retrotransposition.

3.4. Targeting EGFP sequences by fusing engineered zinc finger proteins to ORF2p

Gene editing technologies based on designer ZF proteins were one of the first techniques for genome engineering employed by researchers.

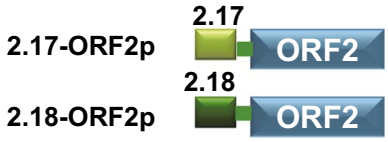

Fig. 3. The EGFP targeting ZF-ORF2p fusion proteins. Schematic of the fusion proteins are shown. $ZF_{2.17}$-ORF2p and $ZF_{2.18}$-ORF2p each contain a three-fingered zinc finger that targets a 9 bp sequence in engineered green fluorescent protein (EGFP).

Most ZFs are engineered to have from three to six fingers to effectively target a unique DNA sequence in a genome (Beerli et al., 2000; Beerli et al., 1998; Gaj et al., 2013). Due to their wide range of uses and flexibility of design, we decided to utilize two ZFs, $ZF_{2.17}$ and $ZF_{2.18}$ previously designed to target a sequence in the engineered green fluorescence protein (EGFP) (Handel et al., 2009). Each zinc finger is deigned to target 9 bp in the EGFP gene.

Two fusions were created: $ZF_{2.17}$-ORF2, and $ZF_{2.18}$-ORF2 (Fig. 3). Although the Alu retrotransposition assay using the created HeLa-EGFP cell line (see Materials and Methods) showed that the fusion proteins supported retrotransposition (Table 2), multiple attempts failed to detect the fusion proteins by Western blot analysis (data not shown). We recovered 14 Alu insertions driven by $ZF_{2.17}$-ORF2 and12 insertions driven by the $ZF_{2.18}$-ORF2 fusion protein (detailed data in Supplemental Table S4). There was no significant enrichment of the expected target sequence (motif) and the top consensus sequences for the DNA binding domain showed no similarity with the expected DNA binding domain (Table 3, Supplemental Table S1). From these results, we determined that fusing these specific zinc fingers to the N-terminus of the L1 ORF2 protein retained the ability to drive Alu retrotransposition events, but the resulting fusion proteins might not be stable or capable of driving targeted Alu insertions.

3.5. Adapting the CRISPR system to target ORF2p-driven retrotransposition events to specific sites

Components of the clustered regularly interspaced short palindromic repeats (CRISPR) bacterial immune system have been successfully adapted to generate a variety of site-specific gene editing tools. These components include the endonuclease Cas9 and a modified RNA component referred to as “guide RNA” or gRNA that confers targeting specificity of the system through base pairing to the DNA and determining the location of the Cas9 cleavage. Sequence modifications of the guide RNA allows to redirect Cas9 to selected target sequences of interest.

Because of the previous success of Cas9 fusion and selected gRNAs to redirect proteins, we selected to test this system to modify the targeting capability of the L1 ORF2p. However, for the approach to function, the Cas9 cleaving should not compete with the ORF2 endonuclease activity. Thus, for our purposes, Cas9-ORF2 fusions should only have one functional endonuclease. In order to create a fusion protein with only one functional endonuclease present in the final protein, we have two alternatives: 1)- use an endonuclease-deficient Cas9, or 2)- use an ORF2p lacking endonuclease activity.

3.6. Use of a catalytically deficient Cas9 as a DNA binding domain

Our first approach was to fuse an endonuclease defective Cas9 (dCas9) with the functional ORF2 protein (Fig. 4A). Although the dCas9 protein is a catalytically inactive, it retains its ability to interact with a gRNA and targeting capability (Qi et al., 2013); thus serving as a DBD. As we rely on the ORF2 endonuclease for DNA cleavage, we selected gRNAs that would target sequences proximal to potential ORF2 endonuclease sites, *i.e.* AT-rich. A second variation of this approach was also evaluated where the ORF2p was instead fused to the bacteriophage MS2 coat protein and targeted using modified guide RNAs tethered

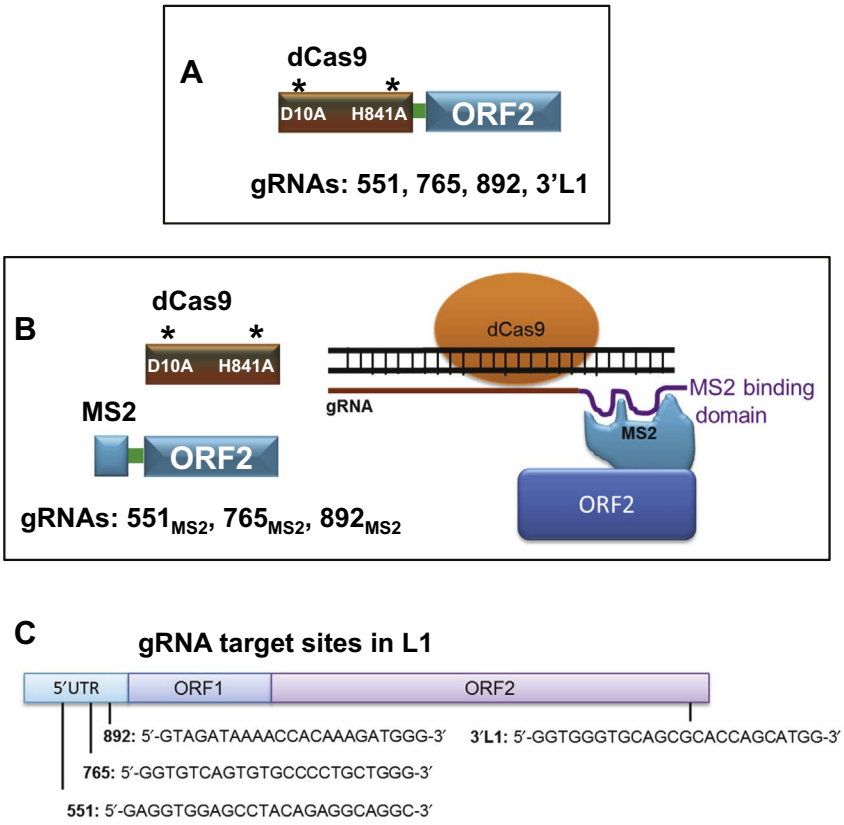

Fig. 4. Using dCas9 as a DNA binding domain. A. Schematic of the catalytic defective dCas9 mutant (D10A, H841A) -ORF2p fusion protein. The four guide RNAs (gRNAs) used with this fusion are indicated. B. Schematic of the approach using the dCas9 + MS2-ORF2p fusion protein and the MS2 tagged gRNAs. The rationale is that the dCas9 and MS2-tagged gRNA complex will bind the MS2-ORF2p fusion protein and bring it into proximity of the target sequence. C. Schematic of an L1 sequence is shown with the location and sequence of the target sites of the gRNAs used with this approach.

with the MS2 stem-loop structures (See Fig. 4B). To increase the abundance of potential target sites, we selected three gRNAs that target different sequences in the 5'UTR of the consensus sequence of the L1Hs: 551, 765, and 892 (#s denote position in the 5'UTR of the repetitive element L1, see Fig. 4C). Thus, instead of relying on only one target site per genome there will be several thousand potential sites for insertions to occur. Genome analysis evaluating the frequency of selected the targets sites (allowing for one mismatch) confirms the abundance of these sites in the genome (Table 5). Both the dCas9-ORF2p fusion protein and Cas9 plus the MS2-ORF2p fusion protein were able to effectively drive Alu retrotransposition events (Table 2). In this assay, we co-transfected our dCas9-ORF2 fusion protein with the 551, 765, 892 or 3'L1 gRNA to ensure the gRNA did not inhibit retrotransposition (data in Table 2 represent an average of all conditions). However, after analysis of a total of 157 recovered Alu inserts, we observed no significant enrichment of the expected target sequence (motif) and the top consensus sequences for the gRNA target site showed no similarity with the expected DNA binding domain (Table 3 and Supplemental Table S1) (insert details in Supplemental Tables S5–8).

3.7. Use of Cas9-or Nickase- endonuclease defective ORF2p fusion proteins

The second "CRISPR" approach was to create a fusion protein containing an ORF2p devoid of endonuclease activity. In this case, we selected to fuse the wildtype (WT) Cas9 protein or the D10A nickase variant (Cong et al., 2013) to the N-terminus of either an ORF2 endonuclease double mutant (endo–) or to a truncated ORF2p lacking the endonuclease domain *i.e.* containing only the reverse transcriptase and cys domain (see Fig. 5A). In addition, for effective insertion of Alu the 3′end of the RNA needs to base pair with the cleaved DNA for TPRT to occur (Fig. 5B). Thus, we designed a gRNA using the available website www.ZiFit.partners.org to target a genomic location that when cleaved by Cas9 or the nickase, it would provide the necessary T-rich sequence for TPRT priming of the Alu A tail (Fig. 5B). The selected guide RNA (gRNA2) targets a unique site on chromosome 8:2587720–2587745 (hg19) that contains the appropriate sequences to both direct site-specific Cas9 cleavage and provide a T-rich sequence for A-tail annealing of the Alu RNA.

Both the Cas9 and the L1 ORF2p are relatively large proteins (~150 kDa each) and a fusion of the two may compromise the Cas9 protein/gRNA interaction. To evaluate the Cas9 targeting and cleavage capabilities of the fusion proteins, we utilized a gene-targeting assay at the *WRN* (Werner syndrome RecQ-like helicase) locus. In this assay, Cas9 stimulates homologous recombination with a homology arm cassette (HAC) that contains sequences homologous to the specific genomic location targeted by the gRNA flanking a blasticidin-resistance gene. Targeted insertions of the HAC can be identified through PCR analysis of DNA from blasticidin-resistant colonies (see Fig. 5C). Using this assay, we evaluated the wild-type Cas9, the nickase (Cas9 D10A), the Cas9-ORF2endo–, Cas9-RTCYS, nickase-ORF2endo– and the nickase-RTCYS fusion proteins. PCR analysis demonstrated that three out of the four fusion proteins promoted the successful integration of the HAC into *WRN* (Fig. 5C) indicative of proper function of these Cas9/nickase fusion proteins. Although multiple efforts were made, no positive PCR results were observed from blasticidin-resistant colonies generated by the nickase-ORF2endo– fusion.

Evaluation of retrotransposition efficiency demonstrated that although three of the fusion proteins appear to retain Cas9 or nickase cleavage, these fusion proteins were either unable or severely hindered in their capacity to drive Alu retrotransposition (Table 2). Furthermore, sequence analysis of seven rescued Alu inserts from the Cas9 Endo– and the Cas9 RTCYS fusions with gRNA2 revealed no targeting as none of the flanking sequences of the rescued Alus mapped to chromosome 8 (Supplemental Table S9). Furthermore, no PAM sequences were observed near the predicted cleavage site of the insertions, suggestive that the Cas9/nickase did not contribute to the insertion process. Again, we did not observe any significant enrichment of the expected target sequence (motif) and the top consensus sequences for the DNA binding

Table 5
Genomic frequency of the selected gRNA target sites to L1.

gRNA	Target sequence	Instances (allowing for max 1 mismatch)
551	5′-GAGGTGGAGCCTACAGAGGCAGGC-3′	3266
765	5′-GGTGTCAGTGTGCCCCTGCTGGG-3′	1346
892	5′-GTAGATAAAACCACAAAGATGGG-3′	4739
3′L1	5′-GTGGGTGCAGCGCACCAGCATGG-3′	4706

domain showed no similarity with the expected target of the gRNA (Table 3, Supplemental Table S1).

Overall, these data demonstrate that using components of the CRISPR/Cas9 system is not an efficient approach for redirecting ORF2p-driven retrotransposition events to specific locations.

4. Discussion

One of the more surprising observations from our studies is the ability of the ORF2p to remain functional with the addition of a variety of DNA binding domains both at the N- and C- terminus. Previous reports have suggested that GFP fused to the N-terminus of ORF2 changed expression and localization of the ORF2 in tissue culture experiments (Dai et al., 2014; Goodier et al., 2004). Conversely, it has been shown that a 3 × flag tag, when fused to the C-terminus of the ORF2p, is able to recover ORF2 stability and function that is lost with the single GFP fusion (Dai et al., 2014). In our experiments, the ORF2p fusion constructs retained the ability to drive Alu retrotransposition with the addition of protein domains ranging from 20 kDa to 100 kDa to the N terminus (Table 2). However, retrotransposition efficiency drastically diminished or failed when an endonuclease double mutant of ORF2p or the N-terminally truncated version (RTCYS) were included in the Cas9 or nickase fusion proteins. These results are suggestive that the cleavage capabilities of Cas9 or the nickase are unable to substitute for the ORF2 endonuclease. Furthermore, it implies that the endonuclease domain may have other functions needed for retrotransposition in addition to DNA cleavage.

There are a variety of reasons why the evaluated DBD-ORF2 fusion proteins did not appear to modify insertion preference of the retrotransposed Alu. For example, a requisite for proper binding of the Rep protein is the formation of multimers (dimer for LZ, tetramer for TZ) (Waterman et al., 1996). Both LZ- and TZ-tagged Rep were shown to

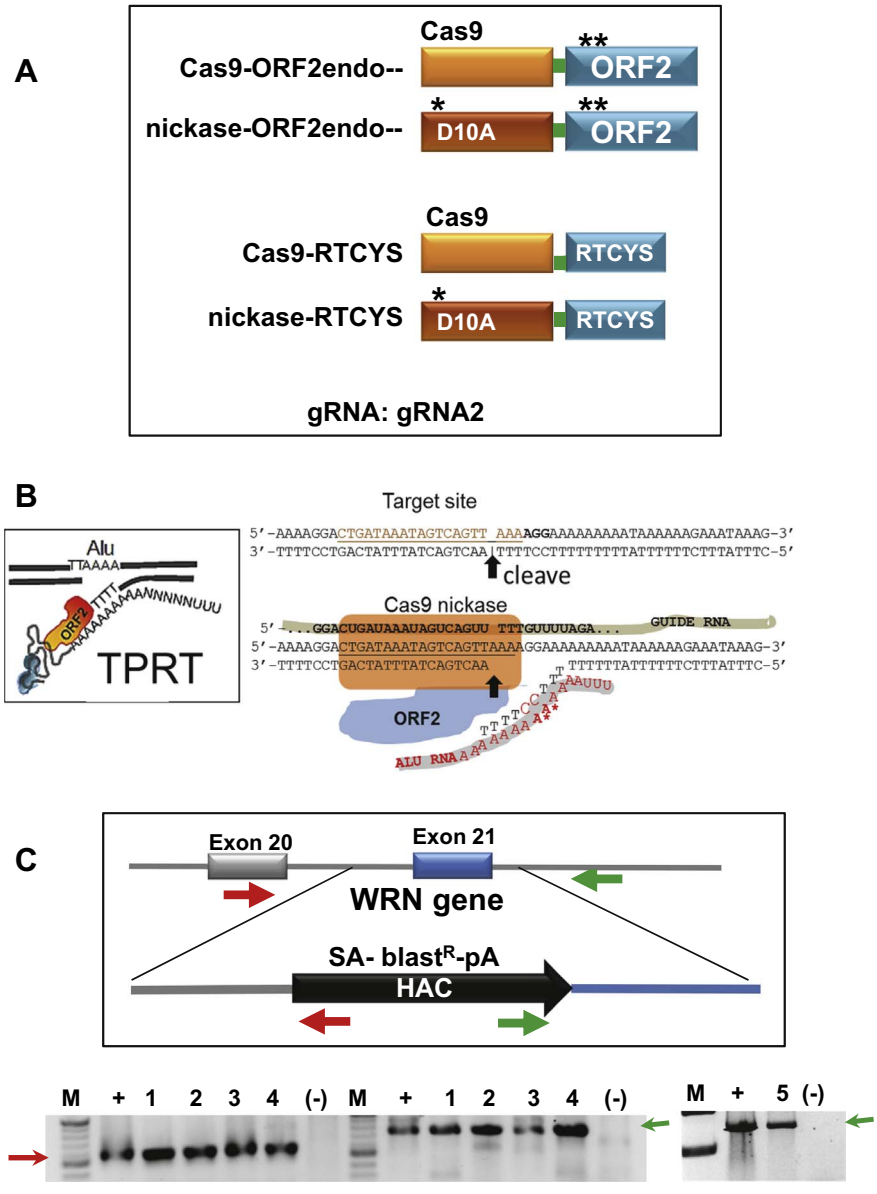

Fig. 5. Using the active Cas9 as the source for the cleavage site for TPRT. A. Schematic of the active Cas9 or nickase and the ORF2p endonuclease double mutant (**) or the N-terminally truncated ORF2p (RTCYS which lacks the endonuclease domain). The gRNAs used with these fusions is indicated. B. Proposed model of the expected Cas9 directed TPRT of Alu inserts. Inset shows the expected TPRT interaction. The ORF2p endonuclease cleaves at AT-rich regions exposing a T-rich DNA strand that can base pair with the A-tail of the Alu, which is needed for TPRT to occur. The T-rich DNA strand provides the priming site for the reverse transcription of the RNA to generate the cDNA (*i.e.* new insert). The strategy for the active Cas9 approach is to select a guide RNA that will target the Cas9 (double strand cut) or nickase (single cut) to cleave at an AT-rich site so that it will expose a T-rich DNA strand comparable to the ORF2p endonuclease cleavage. An advantage of using Alu RNA is that the A-tail RNA sequence can be manipulated so that it is perfectly complementary to the exposed DNA strand (indicated by *). C. Evaluating Cas9 and nickase efficiency. Top inset shows the schematic of the WRN gene, the location of the cleavage target site and the expected Cas9-driven insertion of the homology arm cassette (HAC). Targeted cleavage by the Cas9 will promote recombination of the HAC and generate blasticidin resistance. PCR of the region will provide verification of insertion of the HAC. Red and green arrow show the annealing locations for the primer sets used for PCR analysis. PCR results from individual blasticidin resistant colonies generated by: 1- Cas9, 2- D10A nickase, 3- Cas9-ORF2endo-, 4- Cas9-RTCYS and the 5- nickase-RTCYS fusion proteins. Red and green arrows show the corresponding expected size of the PCR product for each primer pair. The (+) represents a PCR product that was previously generated and used as a positive control for insertion size, (−) represent results from cells transfected with empty vector control and M = size marker. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

redirect *Sleeping Beauty* insertions (Ammar et al., 2012). Although it is unclear if ORF2p functions as a monomer or a dimer, to date no one has been able to show ORF2p dimers or ORF2p-ORF2p interaction. Thus, we speculate that the reason that we were unable to see the Rep target sequence motif near Alu insertions is that the multimerization of the Rep proteins interfered with the ability of the ORF2p to drive retrotransposition. Conversely, in the monomeric form ORF2p was "free" to function properly, however the Rep proteins LZ and TZ were unable to form their required multimeric structure, losing their targeting capability. *Re*-targeting *Sleeping Beauty* might have been successful, because the transposase also functions as a multimer (Izsvak et al., 2002). Similarly, the Cre recombinase fused to both the N and C terminus of ORF2p showed no enrichment of Alu inserts near target the sequence. It is possible that Cre was limited in enriching Alu insertions near the target sequence for the same reason. Therefore, multimerization requirements of both the DNA binding domain and mobile element need to be taken into consideration when designing these fusion proteins.

Interestingly, the TAL effector evaluated that also targets the AAVS1 site but functions as a monomer, did not enrich for targeted Alu insertions near the site when fused to the N terminus of the ORF2p. Most of the fusion proteins tested targeted unique sites, which can limit efficiency of detection.

Another reason for the lack of targeting may be that these fusion proteins are not stable (lack of detection by Western blot analyses). This suggests that some engineered fusion proteins might be more susceptible to processing, losing their ability to successfully confer targeting. Although we were not able to detect the fusion protein, the transfected construct was able to drive retrotransposition. It is possible that the fusion protein was partially processed or needed in very small amount retaining the ability to drive Alu retrotransposition events, but the resulting protein lost targeting capabilities. Therefore, the stability of these fusion proteins needs to be taken into consideration when engineering designer fusion proteins.

Additionally, strategies that relied on the inactive dCas9 protein as a DBD were also unable to drive targeted Alu insertions. These mechanisms likely involved too many components that needed to assemble at a very specific time and place at the site of insertion. Our experience suggests that the more components required to assemble in a complex, the less likely it will work. Therefore, strategies that are complex, either regarding the insertion and targeting machinery requirements, timing, or precision, are more likely to be less effective than simple strategies that only rely on fusing a DNA binding domain to the ORF2p.

Previous data showed that the interchange of the endonuclease domain of the telomere-specific LINEs changed the target site specificity (Takahashi and Fujiwara, 2002). However, when this strategy was applied to the human LINE-1 by adding the site specific endonuclease, insertions recovered were randomly spread in the genome and showed the continued use of the 5′-TTAAAA-3′ consensus (Repanas et al., 2007). Similarly, our strategy to use the Cas9 to provide the cleavage instead of the ORF2 endonuclease also failed to support retrotransposition. The data are indicative that additional factors other than the DNA nicking specificity of endonuclease cleavage were needed for the targeted integration of the human LINE-1 element.

The details governing the ability of DBDs to target ORF2p driven insertions is highly complex. Although adding DBDs to a selected protein through genetic engineering is relatively simple, altering biological properties of the human non-LTRs still remains complex due to the intrinsic details governing the ability of the ORF2p to drive retrotransposition.

Acknowledgments

This research was supported by National Institutes of Health (NIH) R01GM079709A, R01GME079709-05S1 (AMR-E and Z.I.) and P20GM103518/P20RR020152 (AMR-E and VPB.). Competitive Advantage Funds from the Louisiana Cancer Research Consortium (LCRC) and Office of Research of Tulane University were also awarded to AMR-E. VPB is supported by R21AG055387 NIA and by Schlieder Foundation. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. GFP-specific zinc finger proteins were kindly provided by T. Cathomen.

Conflict of interest statement

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2017.11.033.

References

Ade, C., Roy-Engel, A.M., 2016. SINE Retrotransposition: evaluation of alu activity and recovery of de novo inserts. Methods Mol. Biol. 1400, 183–201.
Ammar, I., Gogol-Doring, A., Miskey, C., Chen, W., Cathomen, T., Izsvak, Z., Ivics, Z., 2012. Retargeting transposon insertions by the adeno-associated virus Rep protein. Nucleic Acids Res. 40, 6693–6712.
Beerli, R.R., Segal, D.J., Dreier, B., Barbas III, C.F., 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. U. S. A. 95, 14628–14633.
Beerli, R.R., Dreier, B., Barbas III, C.F., 2000. Positive and negative regulation of endogenous genes by designed transcription factors. Proc. Natl. Acad. Sci. U. S. A. 97, 1495–1500.
Burke, W.D., Calalang, C.C., Eickbush, T.H., 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. 7, 2221–2230.
Cathomen, T., Collete, D., Weitzman, M.D., 2000. A chimeric protein containing the N terminus of the adeno-associated virus Rep protein recognizes its target site in an in vivo assay. J. Virol. 74, 2372–2382.
Christensen, S.M., Ye, J., Eickbush, T.H., 2006. RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc. Natl. Acad. Sci. U. S. A. 103, 17602–17607.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.
Dai, L., La Cava, J., Taylor, M.S., Boeke, J.D., 2014. Expression and detection of LINE-1 ORF-encoded proteins. Mob. Genet. Elements 4, e29319.
deHaro, D., Kines, K.J., Sokolowski, M., Dauchy, R.T., Streva, V.A., Hill, S.M., Hanifin, J.P., Brainard, G.C., Blask, D.E., Belancio, V.P., 2014. Regulation of L1 expression and retrotransposition by melatonin and its receptor: implications for cancer risk associated with light exposure at night. Nucleic Acids Res. 42, 7694–7707.
Dewannieux, M., Esnault, C., Heidmann, T., 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35, 41–48.
Fujiwara, H., 2015. Site-specific non-LTR retrotransposons. Microbiol. Spectr. 3 (MDNA3-2014).
Gaj, T., Gersbach, C.A., Barbas III, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405.
Gilbert, N., Lutz-Prigge, S., Moran, J.V., 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell 110, 315–325.
Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451.
Goodier, J.L., Ostertag, E.M., Engleka, K.A., Seleme, M.C., Kazazian Jr., H.H., 2004. A potential role for the nucleolus in L1 retrotransposition. Hum. Mol. Genet. 13, 1041–1048.
Hamilton, D.L., Abremski, K., 1984. Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. J. Mol. Biol. 178, 481–486.
Handel, E.M., Alwin, S., Cathomen, T., 2009. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol. Ther. 17, 104–111.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.
Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., Zeitler, B., Cherone, J.M., Meng, X., Hinkley, S.J., Rebar, E.J., Gregory, P.D., Urnov, F.D., Jaenisch, R., 2011. Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. 29, 731–734.
Ivics, Z., Katzer, A., Stuwe, E.E., Fiedler, D., Knespel, S., Izsvak, Z., 2007. Targeted sleeping beauty transposition in human cells. Mol. Ther. 15, 1137–1144.
Izsvak, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R.H., Ivics, Z., 2002. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. J. Biol. Chem. 277, 34581–34588.

Jurka, J., 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. U. S. A. 94, 1872–1877.
Klug, A., 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu. Rev. Biochem. 79, 213–231.
Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O., Zhang, F., 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588.
Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A., McLaughlin, S., Muzyczka, N., Rocchi, M., Berns, K.I., 1990. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. U. S. A. 87, 2211–2215.
Kovac, A., Ivics, Z., 2017. Specifically integrating vectors for targeted gene delivery: progress and prospects. Cell Gene Ther. Insights 3, 103–123 (Ref Type: Journal (Full)).
Kroutter, E.N., Belancio, V.P., Wagstaff, B.J., Roy-Engel, A.M., 2009. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE Retrotransposition. PLoS Genet. 5, e1000458.
Lim, K.I., Klimczak, R., Yu, J.H., Schaffer, D.V., 2010. Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties. Proc. Natl. Acad. Sci. U. S. A. 107, 12475–12480.
Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72, 595–605.
Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K., 2013. CRISPR RNA-guided activation of endogenous human genes. Nat. Methods 10, 977–979.
Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., Church, G.M., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838.
Murray, S.A., Eppig, J.T., Smedley, D., Simpson, E.M., Rosenthal, N., 2012. Beyond knockouts: cre resources for conditional mutagenesis. Mamm. Genome 23, 587–599.
Nagy, A., 2000. Cre recombinase: the universal reagent for genome tailoring. Genesis 26, 99–109.
Owens, R.A., Weitzman, M.D., Kyostio, S.R., Carter, B.J., 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J. Virol. 67, 997–1005.
Owens, J.B., Urschitz, J., Stoytchev, I., Dang, N.C., Stoytcheva, Z., Belcaid, M., Maragathavally, K.J., Coates, C.J., Segal, D.J., Moisyadi, S., 2012. Chimeric piggyBac transposases for genomic targeting in human cells. Nucleic Acids Res. 40, 6978–6991.
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.
Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183.
Repanas, K., Zingler, N., Layer, L.E., Schumann, G.G., Perrakis, A., Weichenrieder, O., 2007. Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. Nucleic Acids Res. 35, 4914–4926.
Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., Zhang, F., 2012. A transcription activator-like effector toolbox for genome engineering. Nat. Protoc. 7, 171–192.
Sauer, B., Henderson, N., 1990. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol. 2, 441–449.
Straubeta, A., Lahaye, T., 2013. Zinc fingers, TAL effectors, or Cas9-based DNA binding proteins: what's best for targeting desired genome loci? Mol. Plant 6, 1384–1387.
Takahashi, H., Fujiwara, H., 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. EMBO J. 21, 408–417.
Tan, W., Zhu, K., Segal, D.J., Barbas III, C.F., 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.
Tasic, B., Miyamichi, K., Hippenmeyer, S., Dani, V.S., Zeng, H., Joo, W., Zong, H., Chen-Tsai, Y., Luo, L., 2012. Extensions of MADM (mosaic analysis with double markers) in mice. PLoS One e33332, 7.
Thompson, B.K., Christensen, S.M., 2011. Independently derived targeting of 28S rDNA by A- and D-clade R2 retrotransposons: plasticity of integration mechanism. Mob. Genet. Elements 1, 29–37.
Van Duyne, G.D., 2015. Cre Recombinase. Microbiol. Spectr. 3 (MDNA3-2014).
Voigt, K., Gogol-Doring, A., Miskey, C., Chen, W., Cathomen, T., Izsvak, Z., Ivics, Z., 2012. Retargeting sleeping beauty transposon insertions by engineered zinc finger DNA-binding domains. Mol. Ther. 20, 1852–1862.
Wagstaff, B.J., Barnerssoi, M., Roy-Engel, A.M., 2011. Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. PLoS One 6, e19672.
Wagstaff, B.J., Hedges, D.J., Derbes, R.S., Campos, S.R., Chiaromonte, F., Makova, K.D., Roy-Engel, A.M., 2012. Rescuing Alu: recovery of new inserts shows LINE-1 preserves alu activity through a-tail expansion. PLoS Genet. 8, e1002842.
Wallace, N.A., Belancio, V.P., Deininger, P.L., 2008. L1 mobile element expression causes multiple types of toxicity. Gene 419, 75–81.
Waterman, M.J., Waterman, J.L., Halazonetis, T.D., 1996. An engineered four-stranded coiled coil substitutes for the tetramerization domain of wild-type p53 and alleviates transdominant inhibition by tumor-derived p53 mutants. Cancer Res. 56, 158–163.
Yant, S.R., Huang, Y., Akache, B., Kay, M.A., 2007. Site-directed transposon integration in human cells. Nucleic Acids Res. e50, 35.
Zingler, N., Weichenrieder, O., Schumann, G.G., 2005. APE-type non-LTR retrotransposons: determinants involved in target site recognition. Cytogenet. Genome Res. 110, 250–268.