



(51) International Patent Classification:

C12N 15/861 (2006.01) C12N 7/00 (2006.01)

(21) International Application Number:

PCT/US2023/078957

(22) International Filing Date:

07 November 2023 (07.11.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/423,220 07 November 2022 (07.11.2022) US

(71) Applicant: **SALIOGEN THERAPEUTICS, INC.**
[US/US]; 75 Hayden Avenue, Lexington, MA 02421 (US).

(72) Inventors: **HIGGINS, Joseph, J.**; c/o Saliogen Therapeutics, Inc., 75 Hayden Avenue, Lexington, MA 02421 (US). **TABIBIZAR, Ray**; c/o Saliogen Therapeutics, Inc., 75 Hayden Avenue, Lexington, MA 02421 (US). **YAO, Feng**; c/o Saliogen Therapeutics, Inc., 75 Hayden Avenue, Lexington, MA 02421 (US). **ZHU, Quan, Karen**; c/o Saliogen Therapeutics, Inc., 75 Hayden Avenue, Lexington, MA 02421 (US).

(74) Agent: **ALTIERI, Stephen, L.** et al.; Morgan, Lewis & Bockius LLP, 2222 Market Street, Philadelphia, PA 19103 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

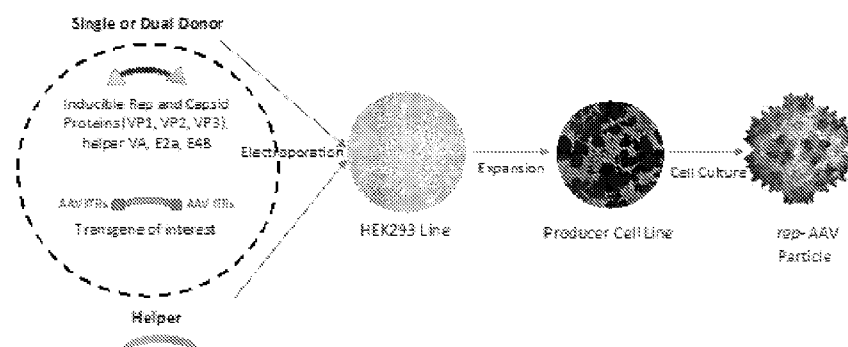
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: ADENO-ASSOCIATED VIRUS (AAV) PRODUCTION

FIG. 1A



(57) Abstract: A method of making a viral particle packaging and producer cell line is provided. The cell line is based on an adeno-associated virus (AAV) vector, the cell line is caused to express, in association with the viral particle, a desired transgene.



ADENO-ASSOCIATED VIRUS (AAV) PRODUCTION

FIELD

The present invention relates, in part, to a method of making a viral particle packaging and producer stable cell line using transposon(s) encoding genes required for AAV production and an enzyme capable of performing targeted genomic integration.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority to U.S. Provisional Patent Application No. 63/432,220, filed November 7, 2022, which is incorporated by reference herein in its entirety.

SEQUENCE LISTING

The instant application contains a Sequence Listing that has been submitted electronically in XML file format and is hereby incorporated by reference in its entirety. The Sequence Listing for this application is labeled "SAL-023PC_126933-5023_SequenceListing.XML", which was created on November 6, 2023, and is 1,045,814 bytes in size.

BACKGROUND

Adeno-associated virus (AAV), first discovered in 1965 as a small, replication-defective viruses that infect humans and some other primate species, has since become a promising tool in gene therapy for treatment of various diseases and/or disorders.

AAV is a very small (20–26 nm), icosahedral, and nonenveloped virus AAV particle containing a single-stranded DNA genome consisting of approximately 4.7 kb. The genome includes three open reading frames (ORFs) encoding for replication (non-structural) proteins (*rep*), capsid (structural) proteins (*cap*), and the assembly activating protein (*AAP*). These three genes give rise to at least nine gene products through the use of three promoters, alternative translation start sites, and differential splicing. See Naso *et al.*, *BioDrugs*. 2017;31(4):317-334. The coding sequences for the *rep*, *cap*, and *AAP* genes are flanked by inverted terminal repeats (ITRs). The 145-nt ITRs are partially paired, and they fold upon themselves to maximize base pairing and form a T-shaped hairpin structure. The *rep* gene encodes four Rep proteins (Rep78, Rep68, Rep52, and Rep40), which are required for viral genome replication and packaging. The *cap* gene expression results in viral capsid (Cap) proteins (VP; VP1/VP2/VP3), which form the outer capsid shell that protects the viral genome, as well as being actively involved in cell binding and internalization. Naso *et al.* (2017).

The AAV genome contains two viral promoters known as p5 and p19, which regulate the transcription of the four Rep proteins with molecular masses of 78, 68, 52, and 40 kDa (Rep68 and Rep40 being the splice variants of Rep78 and Rep52, respectively). Expression of the *cap* gene is driven by the P40 promoter and regulated by alternative splicing and different translation initiation sites, resulting in three cap proteins (VP1, VP2, and VP3) that form an icosahedral capsid of 3.9 kDa. The molecular ratio of these proteins (VP1:VP2:VP3) is approximately 1:1:10. The AAV genome

also encodes for the AAP in an alternative ORF of the cap gene that plays a major role for capsid assembly. See Penaud-Budloo *et al.*, *Mol Ther Methods Clin Dev* 2018;8:166-80.

AAV includes 12 different AAV serotypes. AAV belongs to the genus *Dependoparvovirus* (the *Parvoviridae* family) because it needs the presence of a helper virus for replication and assembly, e.g., adeno-, herpes-, human papilloma- or vaccinia viruses. After attachment of AAV to a cell surface receptor of the host cell, the virus gets internalized by endocytosis. Within the nucleus, the viral capsid sheds to release the single-stranded AAV genome which is then converted to double-stranded DNA. The free end of the ITR hairpin hereby acts as a primer for the DNA synthesis. Co-infection with one of the AAV helper viruses leads to the initiation of AAV gene expression, replication, and the production of AAV virions or viral particles. AAV can infect dividing or non-dividing cells.

AAV derived vectors have several advantages for viral-based gene therapy. AAV is believed to be non-pathogenic for humans, which makes it a suitable vehicle for delivery of genetic material to human cells. AAV also has low cytotoxicity and elicits a very mild immune response. Therefore, AAV vectors are well suitable for *in vivo* gene delivery. Furthermore, different AAV serotypes can target various tissues. Due to their properties, AAV derived vectors are becoming a more preferable tool as compared to adenovirus- and retrovirus-derived vectors.

Accordingly, there is an increasing demand to manufacture desired AAV serotypes in large quantities for pre-clinical and clinical trials, and various other purposes.

SUMMARY

In aspects, the present application provides methods of making viral particle packaging and producer cell lines. In embodiments, the methods comprise transfecting a cell with two or more nucleic acids. In embodiments, the nucleic acids encode one or more of: (a) an enzyme capable of performing targeted genomic integration, (b) an inducible viral replication (Rep) gene, (c) an inducible viral Capsid (cap) gene, (d) one or more adenoviral auxiliary genes (e.g., selected from one or more of E1A, E1B, E4, E2A, and VA of an AAV), (e) an insulator (e.g., selected from HS4, D4Z4), (f) one or more terminal ends recognized by the enzyme, and (g) a transgene flanked by AAV inverted terminal repeats (ITRs). In embodiments, the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof. In embodiments, the resulting transfected cell expresses the transgene in association with the viral particle.

In embodiments, the cell is transfected with two nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration, and the second nucleic acid encodes an inducible viral Rep gene, (c) an inducible viral cap gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, (f) one or more terminal ends recognized by the enzyme, and (g) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with three nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (c) an inducible viral cap gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, and (f) one or more terminal ends recognized by the enzyme; and the third nucleic acid encodes (g) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with four nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, and (f) one or more terminal ends recognized by the enzyme; the third nucleic acid encodes (c) an inducible viral cap gene; and the fourth nucleic acid encodes (g) a transgene flanked by AAV ITRs.

In aspects, the present application provides methods of transfecting an E1A, E1B+ cell with two or more nucleic acids. In embodiments, the nucleic acids encode one or more of: (a) an enzyme capable of performing targeted genomic integration, (b) an inducible viral replication (Rep) gene, (c) an inducible viral Capsid (cap) gene, (d) an insulator (e.g., selected from HS4, D4Z4), (e) one or more terminal ends recognized by the enzyme, and (f) a transgene flanked by AAV inverted terminal repeats (ITRs). In embodiments, the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof. In embodiments, the resulting transfected cell expresses the transgene in association with the viral particle.

In embodiments, the E1A, E1B+ cell is transfected with two nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration, and the second nucleic acid encodes an inducible viral Rep gene, (c) an inducible viral cap gene, (d) an insulator, (e) one or more terminal ends recognized by the enzyme, and (f) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with three nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (c) an inducible viral cap gene, (d) an insulator, and (e) one or more terminal ends recognized by the enzyme; and the third nucleic acid encodes (f) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with four nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (d) an insulator, and (e) one or more terminal ends recognized by the enzyme; the third nucleic acid encodes (c) an inducible viral cap gene; and the fourth nucleic acid encodes (f) a transgene flanked by AAV ITRs.

In aspects, methods for an adeno-associated virus (AAV) packaging and producer cell line are provided. In embodiments, the methods make use of a technique to include, among other elements, the AAV viral replication (Rep)

and Capsid (Cap) genes under a control of an inducible promoter, in an E1A, E1B+ cell. In some embodiments, the inducible Rep and Cap genes (under a control of an inducible promoter), AAV vector DNA sequences, and essential helper genes are included in a single genetic construct such as, e.g., a donor plasmid flanked by end sequences recognized by an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme in accordance with embodiments of the present disclosure). A transgene of interest can be encoded by a separate nucleic acid, or it can be part of the genetic construct encoding the inducible Rep and Cap genes, AAV vector DNA sequences, and essential helper genes.

In aspects, methods for making a packaging and producer cell line with a dual donor/helper system are provided. In the dual donor/helper system, inducible Rep and Cap genes (sometimes collectively referred to as "Rep/Cap genes"), and a transgene of interest are encoded by respective separate nucleic acids. In embodiments, the dual donor/helper system includes, without limitation, a nucleic acid encoding inducible Rep and Cap genes, a nucleic acid (e.g., a donor expression vector, such as a helper AAV donor plasmid) encoding a transgene of interest, and a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme).

In aspects, the present application provides a method of making a viral particle packaging and producer cell line, the method comprising transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration, a second nucleic acid encoding a donor DNA comprising inducible viral replication (Rep) and Capsid (cap) genes, adenoviral auxiliary genes selected from E1A and E1B, and insulators (optionally selected from HS4, D4Z4), and comprising terminal ends recognized by the enzyme, and a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

Accordingly, in aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises transfecting an E1A, E1B+ cell with: a first nucleic acid encoding an enzyme capable of performing targeted genomic integration; a second nucleic acid encoding a donor DNA comprising inducible viral replication (Rep) and Capsid (Cap) genes, adenoviral auxiliary genes selected from E4ORF6, E2A and VA RNA, and insulators (e.g., without limitation, HS4 and/or D4Z4), and comprising terminal ends recognized by the enzyme; wherein the inducible expression of AAV viral replication (Rep) and Capsid (cap) genes, as well as the adenoviral auxiliary genes E4ORF6 and E2A, are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof, and a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

In aspects, methods for making a packaging and producer cell line with a single donor/helper system are provided. In the single donor/helper system, inducible Rep/Cap genes and a nucleic acid (e.g., a donor expression vector, such as

a helper AAV donor plasmid) encoding a transgene of interest are encoded by a same construct. In embodiments, the single donor/helper system includes, without limitation, a nucleic acid (e.g., a donor expression vector) encoding inducible Rep and Cap genes and a transgene of interest, and a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme).

Accordingly, in aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration; and a second nucleic acid encoding viral packaging genes, viral helper genes, insulators (e.g., without limitation, HS4 and D4Z4), and a transgene, and comprising terminal ends recognized by the enzyme, to thereby result in a transfected cell that expresses the transgene in association with a viral particle.

In embodiments, the E1A, E1B+ cell is, without limitation, a HEK293, E1A, E1B+ engineered CHO-K1, or Sf9 cell line.

In embodiments, the E1A, E1B+ cell is, without limitation, a Chinese hamster ovary (CHO), baby hamster kidney (BHK), human embryonic kidney (HEK293T) cells, Vero cell, or *Spodoptera frugiperda* 9 (Sf9) cell.

In embodiments, the viral helper genes comprise adenoviral auxiliary genes.

In embodiments, the adenoviral auxiliary genes are selected from one or more of E1A, E1B, E4ORF6, E2A, and VA of an AAV, optionally E1A and E1B.

In embodiments, the viral packaging genes comprise *rep* and *cap* genes of viral replication (Rep) and Capsid (Cap) proteins of an AAV.

In embodiments, the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof.

In embodiments, the tetO-containing AAV2 P40 promoter comprises a nucleotide sequence of SEQ ID NO: 816, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the P40 intron is an AAV2 P40 intron.

In embodiments, the AAV2 P40 intron comprises a C276A substitution, or a substitution at position corresponding thereto relative to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises one or more mutated translation start sites (ATGs), optionally wherein the translation start sites are mutated to one of CTG, ACG, and TTG.

In embodiments, the AAV2 P40 intron comprises substitutions at one or more positions A13, A32, T42, A61, A71, A89, A203, A246, A258, and T282, or one or more positions corresponding thereto, relative to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises substitutions at one or more positions A13C, A32C, T42C, A61C, A71T, A89C, A203T, A246C, A258C, and T282C corresponding to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 818, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 819, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 817, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified TATA box element has the nucleotide sequence of TATATAA. In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 823, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8

mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the tetO-containing P40 promoter is a tetO-containing AAV2 P40 promoter.

In embodiments, the AAV2 P40 promoter comprises the nucleotide sequence of SEQ ID NO: 820, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified P40 intron is a AAV9 P40 intron.

In embodiments, the modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 821, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified TATA box element has the nucleotide sequence of TATATAA.

In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 822, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises a nucleotide sequence of SEQ ID NO: 824, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the P40 promoter further comprises one or more enhancer elements.

In embodiments, the enhancer element comprises one or more cis-acting elements, optionally selected from an Sp1 binding site, GC rich sequence, GCGGAAC motif, TAATGARAT element, AP1 binding site, and CCAAT box element.

In embodiments, the enhancer element comprises about 1 to about 5 Sp1 binding sites, optionally about 1, or about 2, or about 3, or about 4, or about 5 Sp1 binding sites.

In embodiments, the enhancer element comprises about 1 or about 2 GC rich sequences.

In embodiments, the enhancer element is derived from an hCMV Enhancer Element-3.

In embodiments, the hCMV Enhancer Element-3 comprises the nucleotide sequence of SEQ ID NO: 830 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the expression of the Rep and Cap proteins is controlled by an inducible promoter, optionally an antibiotic-dependent promoter. In embodiments, the antibiotic-dependent promoter is tetracycline- or a variant thereof dependent promoter and a tetracycline-repressor-based (t-REx) system is used. In embodiments, the antibiotic-dependent promoter is tetracycline- or a variant thereof dependent promoter and a tetracycline-controlled transactivator (rtTA) system is used. In embodiments, the antibiotic-dependent promoter is a coumermycin/novobiocin promoter, or a variant thereof.

In embodiments, the viral particle is an AAV, and wherein the AAV is selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, and AAV9.

In embodiments, any one or more of the first, second, and third nucleic acids is encoded by a single nucleic acid.

In embodiments, the second nucleic acid and/or the third nucleic acid are included in a single expression vector.

In embodiments, the first and second nucleic acids are included in a single expression vector, and the third nucleic acid is included in an expression vector that is different from the expression vector including the first and second nucleic acids.

In embodiments, the single expression vector comprises a plasmid.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 825 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about

99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 826 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 841 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 842, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 843, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, one or more of the first, second, and third nucleic acids is RNA, optionally mRNA, optionally synthetic mRNA.

In embodiments, one or more of the first, second, and third nucleic acids is DNA, optionally plasmid DNA.

In embodiments, one or more of the first, second, and third nucleic acids is an expression vector, wherein the expression vector is optionally a plasmid.

In embodiments, the transfected cell generates using the method of making a viral particle packaging and producer cell line in accordance with embodiments of the present disclosure comprises a donor DNA comprising the transgene flanked by AAV inverted terminal repeats (ITRs).

In embodiments, the terminal ends or ITRs comprise the nucleotide sequence of SEQ ID NO: 831 and/or SEQ ID NO: 832, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the transgene has a size of about 200,000 base pairs or less.

The methods in accordance with the present disclosure can be used to produce viral particles carrying a transgene of a large size. For example, in embodiments, the transgene has a size of at least 200,000 base pairs. In embodiments, the transgene has a size of about 200,000 base pairs.

In embodiments, the method further comprises culturing the transfected cell in a medium that expands a population of the transfected cells to create a stably transfected packaging and producer cell line.

In embodiments, the stably transfected producer cell line is capable of producing replication-deficient viral particles in association with the transgene.

In embodiments, the transfection comprises electroporation, nucleofection, lipofection, or calcium phosphate transfection.

In embodiments, the method is helper virus-free.

In embodiments, an enzyme capable of performing targeted genomic integration causes the transgene to be inserted in a certain genomic locus and/or site (e.g., at a TA dinucleotide site or a TTAA tetranucleotide site in a genomic safe harbor site (GSHS)) of a nucleic acid molecule.

In embodiments, the enzyme capable of performing targeted genomic integration is a recombinase. In embodiments, the enzyme has one or more mutations which confer hyperactivity. In embodiments, the recombinase is an integrase.

In embodiments, the recombinase is an integrase or a mobile element enzyme. In embodiments, the integrase is a mobile element enzyme. In embodiments, the mobile element enzyme is an engineered mammalian mobile element enzyme. In embodiments, the mobile element enzyme is a mammal-derived RNA mobile element enzyme (e.g., a

helper RNA mobile element enzyme). In embodiments, the mobile element enzyme is a mammal-derived DNA mobile element enzyme. In embodiments, the mobile element enzyme is a chimeric mobile element enzyme.

In embodiments, the enzyme has gene cleavage activity (Exc+) and/or gene integration activity (Int+). In embodiments, enzyme has gene cleavage activity (Exc+) and/or a lack of gene integration activity (Int-).

In embodiments, the mobile element enzyme comprises: (a) a gene-editing system, and (b) a mobile element enzyme that is capable of inserting the donor DNA comprising a transgene at a TA dinucleotide site or a TTAA tetranucleotide site in a genomic safe harbor site (GSHS). In embodiments, the transgene encodes a complete polypeptide. In embodiments, the transgene is defective or substantially absent in a disease state.

In embodiments, the gene-editing system comprises a Cas9 enzyme guide RNA complex. In embodiments, the Cas9 enzyme guide RNA complex comprises a nuclease-deficient (or inactive, or dead) dCas9 guide RNA complex, also referred to as dCas9 guide RNA complex.

In embodiments, the nuclease-deficient dCas9 guide RNA complex comprises a guide RNA selected from: GTTTAGCTCACCCGTGAGCC (SEQ ID NO: 91), CCCAATATTATTGTTCTCTG (SEQ ID NO: 92), GGGGTGGGATAGGGGATACG (SEQ ID NO: 93), GGATCCCCCTCTACATTTAA (SEQ ID NO: 94), GTGATCTTGTACAAATCATT (SEQ ID NO: 95), CTACACAGAATCTGTTAGAA (SEQ ID NO: 96), TAAGCTAGAGAATAGATCTC (SEQ ID NO: 97), and TCAATACACTTAATGATTTA (SEQ ID NO: 98), or a variant thereof.

In embodiments, the enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

In embodiments, the enzyme is an engineered version, including but not limited to hyperactive forms, of an enzyme (e.g., without limitation, a mobile element enzyme) derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

In embodiments, the mobile element enzyme is from one or more of the Sleeping beauty, Tn1, Tn2, Tn3, Tn5, Tn7, Tn9, Tn10, Tn552, Tn903, Tn1000/Gamma-delta, Tn/O, tnsA, tnsB, tnsC, tniQ, IS10, ISS, IS911, Minos, piggyBac, Tol2, Mos1, Himar1, Hermes, Tol2, Minos, Tel, P-element, MuA, Ty1, Chapaev, transib, Tc1/mariner, or Tc3 donor DNA system, or biologically active fragments variants thereof, inclusive of hyperactive variants.

In embodiments, the enzyme is a mobile element enzyme such as a *Myotis lucifugus* mobile element enzyme (MLT mobile element enzyme), also referred to herein as an MLT mobile element enzyme. In embodiments, the MLT mobile element enzyme is a wild-type MLT mobile element enzyme. In embodiments, the MLT mobile element enzyme is a

modified MLT mobile element enzyme, also referred to herein as a corrected MLT mobile element enzyme. In embodiments, the MLT mobile element enzyme (e.g., the wild-type MLT mobile element enzyme, the corrected MLT mobile element enzyme, or a variant thereof) has one or more mutations such as hyperactive mutations.

In embodiments, the MLT mobile element enzyme is a modified MLT mobile element enzyme. In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, or a variant having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, and one or more mutations selected from L573X, E574X, and S2X, wherein X is any amino acid or no amino acid, optionally X is A, G, or a deletion. In embodiments, the mutations are L573del E574del, and S2A (SEQ ID NO: 1).

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1 and S8P and C13R mutations (SEQ ID NO: 11). In embodiments, the MLT mobile element enzyme has an amino acid sequence having mutations at positions which correspond to at least one of S8P and C13R mutations relative to the amino acid of SEQ ID NO: 1 or a functional equivalent thereof. In embodiments, the MLT mobile element enzyme has an amino acid sequence having mutations at positions which correspond to S8P and C13R mutations relative to the amino acid of SEQ ID NO: 1 or a functional equivalent thereof. In embodiments, the MLT mobile element enzyme has an amino acid sequence having mutations at positions which correspond to S8P, C13R, and N125K mutations relative to the amino acid of SEQ ID NO: 1 or a functional equivalent thereof.

In embodiments, the MLT mobile element enzyme has a nucleotide sequence of SEQ ID NO: 2 (which is codon-optimized) and an amino acid sequence SEQ ID NO: 1. In embodiments, the MLT mobile element enzyme has a nucleotide sequence of SEQ ID NO: 2, or a nucleotide sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, or a codon-optimized form thereof. In embodiments, the MLT mobile element enzyme has an amino acid sequence SEQ ID NO: 1, or an amino acid sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

In embodiments, the MLT mobile element enzyme includes a hyperactive mutation selected from **TABLE 14A** and **TABLE 14B**, e.g., about 1, or about 2, or about 3, or about 4, or about 5 hyperactive mutations selected from **TABLE 14A** or **TABLE 14B**, or combinations thereof.

In embodiments, the MLT mobile element enzyme is a wild-type mobile element enzyme that has an amino acid sequence of SEQ ID NO: 10. In embodiments, the MLT mobile element enzyme is a modified mobile element enzyme having the amino acid sequence of SEQ ID NO: 11.

In embodiments, at least one of the first, second, and third nucleic acids is in the form of a lipid nanoparticle (LNP). In embodiments, a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme) and a nucleic acid encoding a donor DNA are in the form of the same LNP, optionally in a co-formulation. In embodiments, a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme) and a nucleic acid encoding a donor DNA are in the same mixture with an LNP.

In embodiments, a method of producing an AAV bearing a gene of interest is provided, to produce the AAV bearing the gene of interest.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with REP expression or substantially enhanced REP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression and REP expression or substantially enhanced REP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with REP expression or substantially enhanced VP1 expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression.

In embodiments, a cell for gene therapy is provided that is generated by a method in accordance with embodiments of the present disclosure. In some embodiments, a pharmaceutical composition comprising the cell is provided.

In embodiments, a method for treating a disease or condition using gene therapy is provided, the method comprising administering to a subject in need thereof a transfecting cell generated using a method in accordance with embodiments of the present disclosure. In embodiments, the disease or condition comprises cancer.

In embodiments, the disease or condition comprises an infectious disease. In embodiments, the infectious disease is a coronavirus infection, optionally selected from infection with SAR-CoV, MERS-CoV, and SARS-CoV-2, or variants thereof.

The details of the invention are set forth in the accompanying description below. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, illustrative methods and materials are now described. Other features, objects, and advantages of the invention will be apparent

from the description and from the claims. In the specification and the appended claims, the singular forms also include the plural unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1A is a non-limiting representation of an AAV production strategy in accordance with embodiments of the present disclosure, using a donor vector to create a HEK293 producer cell line or another cell line expressing the AdV E1a and E1b genes (e.g., engineered CHO-K1 or Vero cell line with E1a, E1b) that forms a replication deficient AAV particle containing a transgene of interest. The single or dual donor comprising a transgene of interest is incorporated (e.g., by transfection such as electroporation) into a HEK293 cell line or another cell line expressing the E1a and E1b genes. The HEK293 cells are expanded to create a producer cell line and culturing the producer cells creates replication deficient (*rep-*) AAV particles with the transgene of interest.

FIG. 1B is a representation of an example of an inducible Rep/Cap and helper AAV donor plasmid construct (Kana^r), encoding E2A, E4ORF6, and VA RNA helper genes flanked by insulators and mobile element enzyme recognition ends, and used with a helper RNA or DNA to create an AAV producer cell line, in accordance with embodiments of the present disclosure.

FIG. 1C is a representation of an example of a nucleic acid (plasmid) encoding a transgene (gene of interest (GOI)) included between AAV inverted terminal repeats (ITRs), in accordance with embodiments of the present disclosure.

FIGs. 2A-E depict non-limiting representations of chimeric, monomer or head-to-tail dimer mobile element enzymes that are designed to target human GSHS using TALE and Cas9/guide RNA DNA binders. **FIG. 2A.** TALEs include nuclear localization signals (NLS) and an activation domain (AD) to function as transcriptional activators. The DNA binding domain has approximately 16.5 repeats of 33-34 amino acids with a residual variable di-residue (RVD) at position 12-13. **FIG. 2B.** RVDs are shown that have specificity for one or several nucleotides. Only bases of the DNA leading strand are shown. **FIG. 2C.** A chimeric mobile element enzyme construct comprising a TALE DNA-binding protein fused thereto by a linker that is greater than 23 amino acids in length (top) and a chimeric mobile element enzyme construct comprising dCas9 linked to one or more guide RNAs (bottom). **FIG. 2D** is a non-limiting representation of a system in accordance with embodiments of the present disclosure comprising a nucleic acid (e.g., helper RNA) encoding an enzyme capable of performing targeted genomic integration and a nucleic acid encoding a mobile element enzyme (donor DNA). The helper RNA is translated into a bioengineered enzyme (e.g., integrase, recombinase, or mobile element enzyme) that recognizes specific ends and seamlessly inserts the donor DNA into the human genome in a site-specific manner without a footprint. Chimeric mobile element enzymes form dimers or tetramers at open chromatin to insert donor DNA at TTAA recognition sites near DNA binding regions targeted by

dCas9/gRNA or TALEs. Binding of the TALE or Cas9/gRNA to GSHS physically sequesters the mobile element enzyme as a monomer or dimer to the same location and promotes transposition to the nearby TTAA sequences. All RVDs are preceded by a thymine (T) to bind to the NTR shown in **FIG. 2A**). **FIG. 2E** is a non-limiting representation of a system in accordance with embodiments of the present disclosure comprising a nucleic acid (e.g., helper RNA) encoding an enzyme capable of performing targeted genomic integration and a nucleic acid encoding a mobile element enzyme (donor DNA). The helper RNA is translated into a bioengineered enzyme (e.g., integrase, recombinase, or mobile element enzyme) that recognizes specific ends and seamlessly inserts the donor DNA into the human genome in a site-specific manner without a footprint.

FIGs. 3A-3E depict construct templates. **FIG. 3A** depicts a plasmid construct template that transcribes helper RNA that is later processed with a 5'-m7G cap and pseudouridine substitution. **FIG. 3B** depicts a donor DNA construct template with the transgene VLDLR. **FIG. 3C** depicts a chimeric mobile element enzyme construct template with a TALE DNA binder. Other TALEs and mobile element enzymes can be substituted. **FIG. 3D** depicts a chimeric mobile element enzyme construct template with a dCas9/gRNA DNA binder. **FIG. 3E** depicts a system comprising donor DNA (panel A) and helper RNA (panel B): panel **(A)** The donor DNA can be any gene of interest (GOI) including a gene that replaces, inactivates, or provides suicide or helper functions. The GOI can be driven by a predetermined promoter and flanked by insulators to prevent gene silencing. The internal terminal repeats (ITRs) can be specific for a mammal-derived mobile element enzyme. Panel **(B)** The helper RNA is 5'-m7G capped (cap 0, or cap1, or cap 2) with flanking globin 5'- and 3'-UTRs, an N-terminus nuclear localization signal (SV40 or nucleoplasmin), a 34 polyalanine tail region, and pseudouridine modification. The mobile element enzyme (e.g., hyperactive mobile element enzyme) is bioengineered to insert the donor DNA in a site- and/or locus specific manner in a human genome.

FIG. 4 depicts a non-limiting representation of a conventional rAAV production system. Production in adenovirus complementation systems is usually performed as plasmid transfection processes, where AAV Rep/Cap genes, the ITR-flanked gene of interest (GOI), as well as Ad-helper genes are provided as three separate plasmids, respectively, to a E1a/E1b containing HEK293 cell line or other engineered cell line that contains E1a/E1b (e.g., engineered CHO-K1 or Vero cells).

FIG. 5 depicts a non-limiting schematic of an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter with AAV2_REP52/40 ORFs and CAPs (VP1/2/3) under their native P19 or P40 promoter, respectively, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-2 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

FIG. 6 depicts a non-limiting schematic of an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter and AAV2_REP52/40 ORFs under the native P19

promoter, 2) AAV2 CAPs under the tetO-containing P40 promoter with AAV9 P40 intron, 3) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 4) VA RNAs under the control of HSV-2 ICP4 promoter, and 5) hygro-B resistant gene under the control of TK promoter.

FIG. 7 depicts a non-limiting schematic of an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP 78/68 ORFs under the control of P5TO1 promoter and AAV2_REP52/40 ORFs under the native P19 promoter, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-2 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter. Note that the nucleotide sequence between the end of native AAV Rep ORFs and native poly A was deleted in the construct to create an AAV Cap VP1/2/3 deletion phenotype.

FIG. 8 depicts a non-limiting schematic of a construct encoding AAV2 CAP under the control of the tetO-containing hCMV-derived enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_1 (a at position 276 substituted for C) and WPRE/modified ICP27 poly A, followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with ITRs for Sleeping Beauty (SB) transposase.

FIG. 9 depicts a non-limiting schematic of a construct encoding AAV2 CAP under the control of the tetO-containing hCMV enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and MODIFIED AAV2 P40 intron_2 and WPRE/modified ICP27 poly A, followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

FIG. 10A-10B depicts GFP images 120 hours post-transduction of AAVH clone 17+ (**FIG. 10A**) and ddPCR analyses for quantification of the rAAV2-EGFP genome copy number per nanogram (ng) of total genomic DNA (**FIG. 10B**).

FIG. 11A – 11B depicts GFP images 120 hours post-transduction of AAVH clone 13+ (**FIG. 10A**) and ddPCR analyses for quantification of the rAAV2-EGFP genome copy number per nanogram (ng) of total genomic DNA (**FIG. 10B**).

FIG. 12A – 12B depicts expression of functional AAV2 Rep proteins in AAV2-HB (Plasmid shown in FIG. 6) clones as determined by western blot and ddPCR. **FIG. 12A** depicts a western blot analysis illustrating cells from selected AAV2-HB cell clones HB-36, HB-42, HB-118, and HB-120 were treated with 10ng/ml doxycycline for 72 hours followed by extraction with RIPA buffer in the presence of proteinase inhibitor. **FIG. 12B** shows the results of a ddPCR analysis for AAV Genome Amplification / Rep function Assay.

FIG. 13A – 13B depicts expression of functional AAV2 Rep proteins in AAV-CDH (Plasmid shown in FIG. 7) clones as determined by western blot (**FIG. 13A**) and ddPCR (**FIG. 13B**).

FIG. 14A – 14B depicts establishment of AAV2 full producer stable cell pools, CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB, with sleeping beauty transposase-mediated integration of AAV2 CAPs-expressing cassettes (Plasmid shown in FIG. 7) into AAV-CDH Clone 24, 48, 129 serotype adaptable (Cap minus or Cap-) stable cell lines.

FIG. 14A shows ddPCR analyses of SB-mediated AAV2 Cap transposon integration in each Zeocin-selected stable pools. **FIG. 14B** shows western blot analysis of SB-mediated AAV2 Cap transposon integration in each Zeocin-selected stable pools.

FIG. 15A – 15B depicts expression of functional AAV2 Rep proteins in CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB pools as determined by western blot and ddPCR. **FIG. 15A** shows ddPCR results of rAAV Rep proteins being expressed and **FIG. 15B** shows expressed AAV2 Rep proteins are functionally active in supporting AAV genome amplification.

FIG. 16 depicts expression of AAV2 Cap and Rep in CDH-48/Cap-SB stable pool increased with doxycycline dosages and over time. Cells from CDH-48/Cap-SB stable pools were incubated either in the presence of 0, 5, 10, 25, 50, or 100ng/ml doxycycline and harvest at 72 hours post induction; or incubated with 50ng/ml doxycycline and harvested at 24, 48, 72, or 96 hours post induction (**FIG. 16**). Cell extraction with RIPA buffer containing proteinase inhibitors, western blot, and ddPCR analyses of the CDH-48/Cap-SB (AAV2-capsid minus stable lines Clone CDH-48) samples were carried out as described previously. The results show that expression of AAV2 Cap and Rep proteins increased over time in the presence of 50ng/ml doxycycline.

FIG. 17A – 17B depicts production of competent rAAV2 particles from stably engineered CDH-48/Cap-SB cell pools in response to doxycycline. **FIG. 17A** show a graph depicting rAAV2 physical titers from CDH-48/CAP-SB stable pool harvests. **FIG. 17B** shows a graph depicting transduction of CDH-48 cells (AAV2-CAPSID MINUS STABLE LINES CLONE CDH-48) with unconcentrated rAAV supernatants.

FIG. 18A – 18B depicts characterization of Zeocin-resistant single clones selected from the CDH-48/Cap-SB stable pool. **FIG. 18A** depicts ddPCR analysis for SB-mediated AAV2-Cap cassette integration. **FIG. 18B** depicts western blot analysis with antibodies specific for AAV2 CAP.

FIG. 19A – 19B depicts characterization of Zeocin-resistant single clones selected from the CDH-48/Cap-SB stable pool. **FIG. 19A** depicts western blot analysis with antibodies specific for AAV2 REP. **FIG. 19B** shows ddPCR assays of AAV genome amplification following transduction with low VGC/cell of AAV2-EGFP viral particles.

DETAILED DESCRIPTION

The present invention is based, in part, on the discovery that a mobile element enzyme-mediated site-specific genomic integration can be used to produce a stable AAV-based cell line. The cell line is produced by insertion of inducible viral replication (*rep*) and Capsid (*cap*) genes and helper genes (E4ORF6, E2A, and VA) in specific genomic locations of a cell such as, e.g., a HEK293 (E1+) cell or a genetically engineered CHO-K1 or Vero cell line. An advantage of the approach in accordance with embodiments of the present disclosure is that it does not require DNA homology for

recombination, and it is independent of the size of the donor or recipient DNA molecules. Also, the described method is performed in a brief enzyme-catalyzed reaction using an enzyme (e.g., without limitation, a mobile element enzyme) delivered as either DNA or RNA. In embodiments, this protocol allows for the production of stable expression human cell pools in about 4 weeks or in less than 4 weeks.

In aspects, the present application provides methods of making viral particle packaging and producer cell lines. In embodiments, the methods comprise transfecting a cell with two or more nucleic acids. In embodiments, the nucleic acids encode one or more of: (a) an enzyme capable of performing targeted genomic integration, (b) an inducible viral replication (Rep) gene, (c) an inducible viral Capsid (cap) gene, (d) one or more adenoviral auxiliary genes (e.g., selected from one or more of E1A, E1B, E4, E2A, and VA of an AdV), (e) an insulator (e.g., selected from HS4, D4Z4), (f) one or more terminal ends recognized by the enzyme, and (g) a transgene flanked by AAV inverted terminal repeats (ITRs). In embodiments, the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof. In embodiments, the resulting transfected cell expresses the transgene in association with the viral particle.

In embodiments, the cell is transfected with two nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration, and the second nucleic acid encodes an inducible viral Rep gene, (c) an inducible viral cap gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, (f) one or more terminal ends recognized by the enzyme, and (g) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with three nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (c) an inducible viral cap gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, and (f) one or more terminal ends recognized by the enzyme; and the third nucleic acid encodes (g) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with four nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (d) one or more adenoviral auxiliary genes, (e) an insulator, and (f) one or more terminal ends recognized by the enzyme; the third nucleic acid encodes (c) an inducible viral cap gene; and the fourth nucleic acid encodes (g) a transgene flanked by AAV ITRs.

In aspects, the present application provides methods of transfecting an E1A, E1B+ cell with two or more nucleic acids. In embodiments, the nucleic acids encode one or more of: (a) an enzyme capable of performing targeted genomic integration, (b) an inducible viral replication (Rep) gene, (c) an inducible viral Capsid (cap) gene, (d) an insulator (e.g., selected from HS4, D4Z4), (e) one or more terminal ends recognized by the enzyme, and (f) a transgene flanked by AAV inverted terminal repeats (ITRs). In embodiments, the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a

functional variant or fragment thereof. In embodiments, the resulting transfected cell expresses the transgene in association with the viral particle.

In embodiments, the E1A, E1B+ cell is transfected with two nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration, and the second nucleic acid encodes an inducible viral Rep gene, (c) an inducible viral cap gene, (d) an insulator, (e) one or more terminal ends recognized by the enzyme, and (f) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with three nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (c) an inducible viral cap gene, (d) an insulator, and (e) one or more terminal ends recognized by the enzyme; and the third nucleic acid encodes (f) a transgene flanked by AAV ITRs.

In embodiments, the cell is transfected with four nucleic acids, and the first nucleic acid encodes (a) an enzyme capable of performing targeted genomic integration; the second nucleic acid encodes (b) an inducible viral Rep gene, (d) an insulator, and (e) one or more terminal ends recognized by the enzyme; the third nucleic acid encodes (c) an inducible viral cap gene; and the fourth nucleic acid encodes (f) a transgene flanked by AAV ITRs.

In aspects, the present application provides a method of making a viral particle packaging and producer cell line, the method comprising transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration, a second nucleic acid encoding a donor DNA comprising inducible viral replication (Rep) and Capsid (cap) genes, adenoviral auxiliary genes selected from E1A and E1B, and insulators (optionally selected from HS4, D4Z4), and comprising terminal ends recognized by the enzyme, and a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

In aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises: transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration; a second nucleic acid encoding a donor DNA comprising inducible viral replication (*rep*) and Capsid (*cap*) genes, adenoviral auxiliary genes selected from E2A, E4ORF6, and VA, and insulators (optionally selected from HS4, D4Z4), and comprising terminal ends recognized by the enzyme, wherein the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof; and a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

In aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises: transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration, and a second nucleic acid encoding viral packaging genes, viral helper genes, insulators (optionally selected from HS4 and D4Z4), and a transgene, and comprising terminal ends recognized by the enzyme to thereby result in a transfected cell that expresses the transgene in association with a viral particle.

In embodiments, the inducible viral replication (*Rep*) and Capsid (*cap*) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof.

In embodiments, a mammal-derived, helper RNA mobile element enzyme and donor DNA system is used to produce purified recombinant adeno-associated virus (rAAV) vector stocks. rAAV stocks, comprising an ITR-flanked therapeutic transgene or a gene of interest (GOI), are generated by a one plasmid or two-plasmid system, which creates a single producer cell line or separate packaging and producer cell lines.

In mammalian cell-based production systems, the assembly of rAAV vectors typically requires (1) a recombinant vector genome comprising a gene of interest (GOI) and the regulation elements for the GOI expression in target cells (e.g., a promoter, poly A, introns, etc.) flanked by AAV ITRs, (2) the AAV *rep* and *cap* genes provided in *trans*, and (3) Helper functions from adeno-, herpes-, human papilloma- or vaccinia viruses, for replication and rescue of the recombinant genome. A conventional method for production of a vector is co-transfection of a vector plasmid containing the vector genome and a helper plasmid encoding the *rep* and *cap* genes into E1A-transformed human embryonic kidney cells (HEK293 cells) infected with Ad (see FIG. 4). Aponte-Ubillus *et al.*, *Appl Microbiol Biotechnol* 2018;102:1045-54. By this method, 100 to 1,000 vector particles per cell, or 10^8 to 10^9 particles per ml of crude stock lysate, can usually be prepared. Large-scale transfections are typically desired, and vector virions are required to be purified from liters of crude lysate. Although several modifications designed to increase DNA transfer efficiency and AAV helper gene expression have been developed, the existing protocols still require transfection, electroporation, and preparation of complicated conjugates. At the same time, it would be advantageous to develop a stable packaging cell line that can produce vector particles without a transfection step, thus allowing efficient large-scale stock preparation.

Furthermore, AAV vector packaging strategies that rely on overexpression of the viral gene products face challenges such as issues associated with the cellular toxicity of Rep proteins, and the requirement that *rep* and *cap* expression levels be tightly regulated for cell viability hence maximal virion production. One strategy to increase viral gene expression during vector production is to use an inducible promoter (e.g., tetracycline (Tet), cumate, coumermycin/novobiocin, etc. see Zhao *et al.*, *Hum Gene Ther* 2003;14:1619-29) to mimic the state that occurs during wild type AAV infection by amplifying intact fragments of the AAV genome containing the *rep* and *cap* genes.

The most established method for production of rAAV vectors is the plasmid transfection of human embryonic HEK293 cells. Typically, HEK293 cells are simultaneously transfected by a vector plasmid (containing the GOI) and one or two

helper plasmids. The helper plasmid(s) allow the expression of the four Rep proteins, the three AAV structural proteins VP1, VP2, and VP3, the AAP, and the adenoviral auxiliary functions *E2A*, *E4ORF6*, and VA RNA. The additional adenoviral E1A/E1B co-factors necessary for rAAV replication are expressed in HEK293 producer cells. See Qiao *et al.*, *J Virol* 2002;76:1904-13; Yuan *et al.*, *Hum Gene Ther* 2011;22:613-24; Lock *et al.*, *Hum Gene Ther* 2010;21:1259-71. *Rep/cap* genes and adenoviral helper sequences are either cloned on two separate plasmids or combined on one plasmid, hence evolving from a triple plasmid system to transfection with only two plasmids. Grimm *et al.* *Hum Gene Ther* 1998;9:2745-60. The triple plasmid protocol provides versatility with a *cap* gene that can be switched from one serotype to another. The plasmids are usually produced by conventional techniques in *E. coli* using bacterial origin and antibiotic-resistance gene or by minicircle (MC) technology. Schnodt *et al.*, *Mol Ther Nucleic Acids* 2016;5:e355. Although transient transfection in adherent HEK293 cells has been used for large-scale manufacturing of rAAV vectors, it typically requires multiple production batches to fulfill the needs of clinical trials, resulting in lengthy and costly production campaigns. HEK293 cells have been adapted to suspension conditions to be economically viable in the long term. Grieger *et al.*, *Mol Ther* 2016;24:287-97.

Generation of AAV Vector Packaging Cell Lines.

In embodiments, a method of making a viral particle packaging and producer cell line (e.g., the AAV cell line) is provided that comprises transfecting an E1A, E1B+ cell with (1) a first nucleic acid encoding an enzyme capable of performing targeted genomic integration, (2) a second nucleic acid encoding a donor DNA comprising inducible viral replication (*rep*) and Capsid (*cap*) genes, adenoviral auxiliary genes selected from *E2A*, *E4ORF6*, and VA, and insulators (optionally selected from HS4 and D4Z4), and comprising terminal ends recognized by the enzyme, wherein the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof, and (3) a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

In embodiments, a method of making a viral particle packaging and producer cell line is provided that comprises transfecting an E1A, E1B+ cell with (1) a first nucleic acid encoding an enzyme capable of performing targeted genomic integration, and (2) a second nucleic acid encoding viral packaging genes, viral helper genes, insulators (optionally selected from HS4 and D4Z4), and a transgene, and comprising terminal ends recognized by the enzyme, to thereby result in a transfected cell that expresses the transgene in association with a viral particle.

FIG. 1A illustrates a non-limiting example of a method for generating AAV, using a single or dual donor DNA to create a producer cell line from a E1+ cell line (e.g., without limitation, an HEK293 cell line) that forms a replication deficient (*rep*-) AAV particle containing a transgene of interest. As shown in **FIG. 1A**, a donor DNA construct encodes inducible Rep and Cap ("Capsid") proteins (VP1, VP2, and VP3) and helper genes (*E4*, *E2A*, and VA) having end sequences

recognized by an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme), as well as a transgene (or gene) of interest flanked by AAV ITRs. In embodiments, the donor DNA construct is a dual donor DNA comprising a nucleic acid encoding the inducible Rep and Cap proteins and helper genes that is separate from a nucleic acid encoding transgene. In embodiments, the donor DNA construct is a single donor DNA comprising a nucleic acid encoding the inducible Rep and Cap proteins, helper genes, and a transgene. **FIG. 1B** shows an example of an inducible Rep/Cap and helper AAV donor plasmid construct (Kana^r), encoding helper E2A, E4ORF6 and VA genes flanked by insulators and mobile element enzyme recognition ends, and used with a helper RNA or DNA (e.g., a mobile element enzyme) to create an AAV producer cell line. **FIG. 1C** shows an example of a plasmid encoding a transgene included between AAV ITRs. The plasmids shown in **FIGs. 1B** and **1C** can be combined or can be used separately for transfection into an E1+ cell line to produce an AAV particle comprising the transgene (or gene of interest (GOI)).

In embodiments, the method depicted in **FIG. 1A** is performed using a system in accordance with embodiments of the present disclosure shown in **FIG. 2D**. The system of **FIG. 2D** is an integrative, non-viral donor DNA system for site-specific, stable genomic integration. The system allows producing any recombinant AAV serotype by the insertion of inducible AAV rep/cap genes and helper genes (E4, E2A, and VA) in specific genomic locations in a cell, such as, e.g., HEK293 (E1+) cells, a genetically engineered CHO-K1 cell line, or in other cells. The system of **FIG. 2D** comprises comprising a nucleic acid (e.g., helper RNA [which is different from helper genes]) encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme) and a nucleic acid encoding a transgene of interest (donor DNA). The helper RNA is translated into a bioengineered enzyme (e.g., integrase, recombinase, or mobile element enzyme) that recognizes specific ends and seamlessly inserts the transgene of interest into the human genome in a site-specific manner without a footprint.

The present disclosure provides, in embodiments, methods for producing AAV packaging cell lines capable of regulated gene amplification. In embodiments, the methods of the present disclosure employ tetracycline repressor-controlled (tetR) tetO containing promoters (Yao F et al., Hum Gene Ther., 9:1939-1950, 1998) or reverse tetracycline-controlled transactivator (rtTA)-controlled gene expression and SV40 origin replication to regulate the expression of the AAV *rep* and *cap* genes, and the Adv5 E2A and E4ORF6. In embodiments, binding of doxycycline to t-REx or rtTA will initiate transcription from the tetO-containing promoters or SV40 T-antigen gene, leading to expression of the AAV *rep* and *cap* genes, and Adv5, E2A, and E4ORF6 genes (**FIG. 1B**). In embodiments, the construct shown in **FIG. 1B** is permanently integrated into a E1a/E1b containing cell line (e.g., without limitations, an HEK293 cell line stably expressing tetR gene). Transfection in media with an inducer (e.g., without limitation, doxycycline) of inducible *rep* and *cap* genes results in high-level production of Rep and Cap proteins and AAV virion assembly. Because Rep protein expression is inducible, this strategy avoids the toxicity of Rep proteins. In embodiments, producer cells can be generated from these packaging lines by the addition of a vector construct comprising a transgene (or a GOI) flanked

by AAV ITRs that can excise, replicate, and be packaged into virions once *rep* and *cap* are expressed. Alternatively, in embodiments, the transgene construct is included in the same plasmid as the donor DNA, in a "single plasmid" protocol.

In embodiments, methods for an AAV packaging and producer cell line are provided. In embodiments, the AAV *rep* and *cap* genes are under control of an inducible promoter (e.g., t-Rex or rtTA), such that the *rep* and *cap* genes, the AAV vector DNA sequences, and the essential helper genes are encoded on a single donor plasmid flanked by mobile element enzyme recognition sequences (**FIG. 1B**). In embodiments, a transgene is encoded separately from the donor plasmid (see, e.g., **FIG. 1C**). In embodiments, the donor plasmid comprises a transgene such that a one-plasmid (e.g., without limitation, a minicircle) is created and used to generate a transgene-specific cell line (e.g., an E1A, E1B+ cell such as, without limitation, a HEK293 cell line) that is able to produce a clinical grade product for human gene therapy.

In embodiments, methods for making a packaging and producer cell line with a dual donor/helper system are provided. In the dual donor/helper system, inducible Rep and Cap genes (sometimes collectively referred to as "Rep/Cap genes"), and a transgene of interest are encoded by respective separate nucleic acids. In embodiments, the dual donor/helper system includes, without limitation, a nucleic acid encoding inducible Rep and Cap genes, a nucleic acid (e.g., a donor expression vector, such as a helper AAV donor plasmid) encoding a transgene of interest, and a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme).

Accordingly, in aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises transfecting an E1A, E1B+ cell with: a first nucleic acid encoding an enzyme capable of performing targeted genomic integration; a second nucleic acid encoding a donor DNA comprising inducible viral replication (Rep) and Capsid (Cap) genes, adenoviral auxiliary genes selected from E2A, E4ORF6, and VA, and insulators (e.g., without limitation, HS4 or D4Z4), and comprising terminal ends recognized by the enzyme; and a third nucleic acid encoding a donor DNA comprising a transgene flanked by AAV inverted terminal repeats (ITRs), to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

In aspects, methods for making a packaging and producer cell line with a single donor/helper system are provided. In the single donor/helper system, inducible Rep/Cap genes and a nucleic acid (e.g., a donor expression vector, such as a helper AAV donor plasmid) encoding a transgene of interest are encoded by a same construct. In embodiments, the single donor/helper system includes, without limitation, a nucleic acid (e.g., a donor expression vector) encoding inducible Rep and Cap genes and a transgene of interest, and a nucleic acid encoding an enzyme capable of performing targeted genomic integration (e.g., without limitation, a mobile element enzyme).

Accordingly, in aspects, a method of making a viral particle packaging and producer cell line is provided. In embodiments, the method comprises transfecting an E1A, E1B+ cell with a first nucleic acid encoding an enzyme capable of performing targeted genomic integration; and a second nucleic acid encoding viral packaging genes, viral

helper genes, insulators (e.g., without limitation, HS4, D4Z4), and a transgene, and comprising terminal ends recognized by the enzyme, to thereby result in a transfected cell that expresses the transgene in association with a viral particle.

In embodiments, the E1A, E1B+ cell is, without limitation, HEK293, E1A, E1B+ engineered CHO-K1, or Sf9 cell line.

In embodiments, the viral helper genes comprise adenoviral auxiliary genes. In embodiments, the adenoviral auxiliary genes are selected from one or more of E1A, E1B, E4, E2A, and VA of an AAV, optionally E1A and E1B.

In embodiments, the viral packaging genes comprise *rep* and *cap* genes of viral replication (Rep) and Capsid (Cap) proteins of an AAV. In embodiments, the expression of the Rep and Cap proteins is controlled by an inducible promoter, optionally an antibiotic-dependent promoter. In embodiments, the antibiotic-dependent promoter is tetracycline- or a variant thereof dependent promoter and a tetracycline repressor-controlled (tetR) or tetracycline-controlled transactivator (rtTA) system is used. In embodiments, the antibiotic-dependent promoter is a coumermycin/novobiocin promoter, or a variant thereof.

In embodiments, the inducible viral replication (Rep) and Capsid (*cap*) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof.

In embodiments, the tetO-containing AAV2 P40 promoter comprises a nucleotide sequence of SEQ ID NO: 816, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the P40 intron is an AAV2 P40 intron.

In embodiments, the AAV2 P40 intron comprises a C276A substitution, or a substitution at position corresponding thereto relative to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises one or more mutated translation start sites (ATGs), optionally wherein the translation start sites are mutated to one of CTG, ACG, and TTG.

In embodiments, the AAV2 P40 intron comprises substitutions at one or more positions A13, A32, T42, A61, A71, A89, A203, A246, A258, and T282, or one or more positions corresponding thereto, relative to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises substitutions at one or more positions A13C, A32C, T42C, A61C, A71T, A89C, A203T, A246C, A258C, and T282C corresponding to SEQ ID NO: 818.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 818, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 819, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 817, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified TATA box element has the nucleotide sequence of TATATAA. In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 823, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the tetO-containing P40 promoter is a tetO-containing AAV2 P40 promoter.

In embodiments, the AAV2 P40 promoter comprises the nucleotide sequence of SEQ ID NO: 820, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8

mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified P40 intron is a AAV9 P40 intron.

In embodiments, the modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 821, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the modified TATA box element has the nucleotide sequence of TATATAA.

In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 822, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises a nucleotide sequence of SEQ ID NO: 824, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the P40 promoter further comprises one or more enhancer elements.

In embodiments, the enhancer element comprises one or more cis-acting elements, optionally selected from an Sp1 binding site, GC rich sequence, GCGGAAC motif, TAATGARAT element, AP1 binding site, and CCAAT box element.

In embodiments, the enhancer element comprises about 1 to about 5 Sp1 binding sites, optionally about 1, or about 2, or about 3, or about 4, or about 5 Sp1 binding sites.

In embodiments, the enhancer element comprises about 1 or about 2 GC rich sequences.

In embodiments, the enhancer element is derived from an hCMV Enhancer Element-3.

In embodiments, the hCMV Enhancer Element-3 comprises the nucleotide sequence of SEQ ID NO: 830 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the viral particle is an AAV of a suitable serotype. In embodiments, the AAV is selected from AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9. In some embodiments, the AAV serotype is a natural serotype or a synthetically engineered serotype. In embodiments, the AAV is AAV2. In embodiments, the AAV is AAV9. In embodiments, the viral particle is an AAV of any of the AAV serotypes. Today, 12 AAV serotypes are known, and more than 100 variants have been identified. Different serotype capsids can infect different tissues or culture cells in different ways, which depend on the primary receptor and co-receptors on the cell surface or the intracellular trafficking pathway itself.

In embodiments, any one or more of the first, second, and third nucleic acids is encoded by a single nucleic acid. In embodiments, the second nucleic acid and the third nucleic acid are included in a single expression vector. In embodiments, the first and second nucleic acids are included in a single expression vector, and the third nucleic acid is included in an expression vector that is different from the expression vector including the first and second nucleic acids.

In embodiments, the single expression vector comprises a plasmid.

In embodiments, the plasmid comprises an AAV2 CAP under the control of the tetO-containing hCMV-derived enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_1 (a at position 276 substituted for C) and WPRE/modified ICP27 poly A followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 825 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an AAV2 CAP under the control of the tetO-containing hCMV enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and MODIFIED AAV2 P40 intron_2 and WPRE/modified ICP27 poly A followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 826 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP 78/68 ORFs under the control of P5TO1 promoter, REP 52/40 under the control of P19 promoter, 2) Ad5 E2A_IRES_Ad5 E4 OR F6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-2 ICP4 promoter, and 4) hygromycin-B resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 841 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter, REP 52/40 under the control of P19 promoter, and CAPs under the native P40 promoter, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygromycin-B resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 842, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter, REP 52/40 under the control of P19 promoter, 2) AAV2 CAPs under the tetO-containing P40 promoter with AAV9 P40 intron, 3) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 4) VA RNAs under the control of HSV-1 ICP4 promoter, and 5) hygromycin-resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 843, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, any one of the first, second, and third nucleic acids is RNA, optionally mRNA, optionally synthetic mRNA. In embodiments, any one of the first, second, and third nucleic acids is DNA, optionally plasmid DNA. In embodiments, any one of the first, second, and third nucleic acids is an expression vector, wherein the expression vector is optionally a plasmid.

In embodiments, the transfected cell generates using the method of making a viral particle packaging and producer cell line in accordance with embodiments of the present disclosure comprises a donor DNA comprising the transgene flanked by AAV inverted terminal repeats (ITRs).

In embodiments, the method further comprises culturing the transfected cell in a medium that expands a population of the transfected cells to create a stably transfected packaging and producer cell line.

In embodiments, the stably transfected producer cell line is capable of producing replication-deficient viral particles in association with the transgene. In embodiments, the transfection comprises electroporation, nucleofection, lipofection, or calcium phosphate transfection. In embodiments, the method is helper virus-free.

In embodiments, the cell is human embryonic kidney (HEK293), Chinese hamster ovary (CHO) E1A, E1B+ engineered CHO-K1, or *Spodoptera frugiperda* (Sf9) cell line, baby hamster kidney (BHK), vero cell.

In embodiments, the viral helper genes comprise adenoviral auxiliary genes. In embodiments, the adenoviral auxiliary genes are selected from one or more of E1A, E1B, E4, E2A, and VA of an AAV, optionally E1A and E1B.

In embodiments, the viral packaging genes comprise *rep* and *cap* genes of viral replication (Rep) and Capsid (Cap) proteins of an AAV. In embodiments, the expression of the Rep and Cap proteins is controlled by an inducible promoter. In embodiments, the inducible promoter is an antibiotic-dependent promoter, optionally a tetracycline-dependent promoter or a variant thereof, or a cumate or coumermycin/novobiocin promoter or a variant thereof.

In embodiments, the viral particle is an AAV, and optionally wherein the AAV is selected from AAV1, AAV5, AAV2, AAV6, AAV7, AAV8, and AAV9.

In embodiments, any one or more of the first, second, and third nucleic acids is encoded by a single nucleic acid. In embodiments, the second nucleic acid and the third nucleic acid are included in a single expression vector. In embodiments, the first and second nucleic acids are included in a single expression vector, and the third nucleic acid is included in an expression vector that is different from the expression vector including the first and second nucleic acids. In embodiments, the expression vector is or comprises a plasmid.

In embodiments, the plasmid comprises an AAV2 CAP under the control of the tetO-containing hCMV-derived enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_1 (a at position 276 substituted for C) and WPRE/modified ICP27 poly A followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 825 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an AAV2 CAP under the control of the tetO-containing hCMV enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and MODIFIED AAV2 P40 intron_2 and WPRE/modified ICP27 poly A followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 826 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP 78/68 ORFs under the control of P5TO1 promoter, REP 52/40 under the control of P19 promoter, 2) Ad5 E2A_IRES_Ad5 E4 OR F6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 841 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter, REP 52/40 under the control of P19 promoter, and CAPs under the native P40 promoter, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 842, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the plasmid comprises an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter and REP 52/40 under the control of P19 promoter, 2) AAV2 CAPs under the tetO-containing P40 promoter with AAV9 P40 intron, 3) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 4) VA RNAs under the control of HSV-1 ICP4 promoter, and 5) hygro-B resistant gene under the control of TK promoter.

In embodiments, the plasmid comprises the nucleotide sequence of SEQ ID NO: 843, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, any one of the first, second, and third nucleic acids is or comprises RNA, optionally mRNA, optionally synthetic mRNA or modified mRNA. In embodiments, any one of the first, second, and third nucleic acids is DNA, optionally plasmid DNA. In embodiments, any one of the first, second, and third nucleic acids is an expression vector, wherein the expression vector is optionally a plasmid.

In embodiments, the transgene is flanked by AAV inverted terminal repeats (ITRs).

In embodiments, the transgene is flanked by Sleeping beauty ITRs.

SEQ ID NO: 831: Sleeping Beauty LE ITR Sequence (IR/DR(L) Lmut44) (231 bp including TATA):

5'-

tatacagttgaagtcggaagttacatacactaagttggagtcattaaaactcgitttcaactactccacaaattctgttaacaaacaatagtttggcaagtcagtiag
gacatctactttgtcatgacacaagtcattttccaacaattgttiacagacagatttttcaactataattcactgtatcacaaitccagtggtgcagaagttacatacact
aa-- 3'

SEQ ID NO: 832: Sleeping Beauty RE ITR Sequence (IR/DR (R) Rmut13, Δ130, 143, 150) (232 bp including TATA):

5'-

ttgagtgtaagtaaaactctgaccactgggaatgtgatgaagaaataaaaagctgaaatgaatcattctctactattattctgattttcacattctaaaataaagtggt
gaicctaactgacctagacaggaattttactaggattaaatgtcaggaattgtgaaaaagtgagtttaaatgtatttggctaagggtgatgtaaacttcgactcaac
tgtata -- 3'

In embodiments, the ITRs comprise the nucleotide sequence of SEQ ID NO: 831 and/or SEQ ID NO: 832, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the transgene or transposon has a size of about 200,000 base pairs or less. In embodiments, the transgene or transposon has a size of at least 200,000 base pairs. In embodiments, the transgene or transposon has a size of about 200,000 base pairs. In embodiments, the transgene or transposon has a size of about 150,000 base pairs or less. In embodiments, the transgene or transposon has a size of at least 150,000 base pairs. In embodiments, the transgene or transposon has a size of about 150,000 base pairs. In embodiments, the transgene or transposon has a size of about 100,000 base pairs or less. In embodiments, the transgene or transposon has a size of at least 100,000 base pairs. In embodiments, the transgene or transposon has a size of about 100,000 base pairs. In embodiments, the transgene or transposon has a size of about 75,000 base pairs or less. In embodiments, the transgene or transposon has a size of at least 75,000 base pairs. In embodiments, the transgene or transposon has a size of about 75,000 base pairs. In embodiments, the transgene or transposon has a size of about 50,000 base pairs or less. In embodiments, the transgene or transposon has a size of at least 50,000 base pairs. In embodiments, the transgene or transposon has a size of about 50,000 base pairs.

In embodiments, the method of the present disclosure further comprises culturing the transfected cell in a medium that expands a population of the transfected cells to create a stably transfected packaging and producer cell line.

In embodiments, the stably transfected producer cell line is capable of producing replication-deficient viral particles in association with the transgene. In embodiments, the transfection comprises electroporation, nucleofection, lipofection, or calcium phosphate transfection. In embodiments, the method is helper virus-free.

In embodiments, the enzyme capable of performing targeted genomic integration is a recombinase. In embodiments, the recombinase is an integrase or a mobile element enzyme. In embodiments, the enzyme is a mobile element enzyme.

In embodiments, the enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*. In embodiments, the enzyme is an engineered version, including but not limited to hyperactive forms, of an enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

In embodiments, the mobile element enzyme is from one or more of the Sleeping beauty, Tn1, Tn2, Tn3, Tn5, Tn7, Tn9, Tn10, Tn552, Tn903, Tn1000/Gamma-delta, Tn/O, tnsA, tnsB, tnsC, tniQ, IS10, ISS, IS911, Minos, piggyBac, Tol2, Mos1, Himar1, Hermes, Tol2, Minos, Tel, P-element, MuA, Ty1, Chapaev, transib, Tc1/mariner, or Tc3 donor DNA system, or biologically active fragments variants thereof, inclusive of hyperactive variants. In embodiments, the mobile element enzyme has the amino acid sequence of SEQ ID NO: 1, or an amino acid sequence having at least about 80%, or an amino acid sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

SEQ ID NO: 1: MLT mobile element enzyme protein (amino acid sequence of a variant of the hyperactive mobile element enzyme with S at position 8 and C at position 13 (572 amino acids)

MAQHSDYSDDEFKADKLSNYSCDSLENASTSDEDSSDDEVMPRTLRRLRRISSSSSDSESIEGGREEWSHV
DNPPVLEDFLGHQGLNTDAVINNIEDAVKLFIGDDFFEFLEESNRYYNQNRNRFKLSKSLKWKDITPQEMKK
FLGLIVLMGQVRKDRRDDYWTTEPWTETPYFGKTMTRDRFRQIWKAWHFNNADIVNESDRLCKVRPVLDYFVP
KFINIYKPHQQLSLDEGIVPWRGRLFFRVYNAGKIVKYGILVRLLCESDTGYICNMEIYCCEGKRLLETIQTVV
SPYTDWYHIYMDNYNSVANCEALMKNKFRICGTIRKNRGIPKDFQTISLKKGETKFIKNDILLQVWQSKKP
VYLISIIHSAEMEEESQNIIDRTSKKKIVKPNALIDYNKHMKGVDRAQYLSYYSILRRTVKWTKRLAMYMINCAL
FNSYAVYKSVRQRKMGFKMFLKQTAIHWLTDDIPEDMDIVPDLQVPVSTSGMRAKPPTSDDPPCRLSMDMRKHTL
QAIVGSGKKKNILRRCRVCSVHKLRSSETRYMCKFCNIPHLKGACFEKYHTLKNY

In embodiments, the mobile element enzyme comprises an amino acid other than serine at the position corresponding to position 2 of SEQ ID NO: 1.

In embodiments, the amino acid is a non-polar aliphatic amino acid, optionally a non-polar aliphatic amino acid optionally selected from G, A, V, L, I and P, optionally A. In embodiments, the mobile element enzyme does not have additional residues at the C terminus relative to SEQ ID NO: 1.

In embodiments, the enzyme has one or more mutations which confer hyperactivity.

In embodiments, the enzyme has one or more amino acid substitutions selected from S8X₁ and/or C13X₂, or positions corresponding thereto relative to SEQ ID NO: 1. In embodiments, the enzyme has S8X₁ and/or C13X₂ substitutions, at positions corresponding thereto relative to SEQ ID NO: 1. In embodiments, the enzyme has S8X₁ and C13X₂ substitutions, at positions corresponding thereto relative to SEQ ID NO: 1. In embodiments, the enzyme has S8X₁ substitution, at position corresponding thereto relative to SEQ ID NO: 1. In embodiments, the enzyme has C13X₂ substitution, at positions corresponding thereto relative to SEQ ID NO: 1.

In embodiments, X₁ is selected from G, A, V, L, I, and P and X₂ is selected from K, R, and H. In embodiments, X₁ is P and X₂ is R.

In embodiments, the enzyme of the present disclosure comprises an amino acid sequence of SEQ ID NO: 11.

In embodiments, the enzyme comprises at least one, at least five, at least seven, at least nine, or ten substitutions at positions corresponding to: 5, 8, 9, 10, 11, 14, 22, 36, 37, 54, 130, 239, 281, 282, 283, 284, 285, 294, 300, 310, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 345, 375, 416, 427, 475, 481, 491, 520, and/or 561 of SEQ ID NO: 11.

In embodiments, the enzyme comprises one, two, three, four, five, six, seven, eight, nine, or ten substitutions at positions corresponding to: 5, 8, 9, 10, 11, 14, 22, 36, 37, 54, 130, 239, 281, 282, 283, 284, 285, 294, 300, 310, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 345, 375, 416, 427, 475, 481, 491, 520, and/or 561 of SEQ ID NO: 11.

In embodiments, the enzyme comprises at least one, at least five, at least seven, at least nine, or ten substitutions selected from S5P, S8F, D9G, D10G, E11G, A14V, T22C, S36G, T37C, S54N, K130T, G239R, Y281A, C282A, G283A, E284A, G285A, T294A, T300A, N310A, G330A, T331A, I332A, R333A, K334A, N335A, R336A, G337A, I338A, P339A, I345V, T375G, D416A, R427H, D475G, M481V, P491Q, A520T, and A561T, wherein the positions are corresponding to positions of SEQ ID NO: 11.

In embodiments, the enzyme comprises one, two, three, four, five, six, seven, eight, nine, or ten substitutions selected from S5P, S8F, D9G, D10G, E11G, A14V, T22C, S36G, T37C, S54N, K130T, G239R, Y281A, C282A, G283A, E284A, G285A, T294A, T300A, N310A, G330A, T331A, I332A, R333A, K334A, N335A, R336A, G337A, I338A, P339A, I345V, T375G, D416A, R427H, D475G, M481V, P491Q, A520T, and A561T, wherein the positions are corresponding to positions of SEQ ID NO: 11.

In embodiments, the mobile element enzyme is an engineered mammalian mobile element enzyme. In embodiments, the mobile element enzyme is a mammal-derived, helper RNA mobile element enzyme. In embodiments, the mobile element enzyme is a mammal-derived, helper DNA mobile element enzyme.

In embodiments, the enzyme is capable of inserting a donor DNA at a TA dinucleotide site. In embodiments, the enzyme is capable of inserting a donor DNA at a TTAA (SEQ ID NO: 440) tetranucleotide site.

In embodiments, the mobile element enzyme has gene cleavage activity (Exc+) and/or gene integration activity (Int+), and the mobile element enzyme having at least about 90% identity to the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 430, or a nucleotide sequence encoding the same.

SEQ ID NO: 3: PGBD4 Amino Acid Sequence (585 Amino Acids).

MSNPRKRSIP	MRDSNTGLEQ	LLAEDSFDES	DFSEIDSDSN	FSDSALEADK	50
IRPLSHLES	GKSSTSSDSG	RSMKWSARAM	IPRQRYDFTG	TPGRKVDVSD	100
ITDPLQYFEL	FFTEELVSKI	TRETNAQAAL	LASKPPGPKG	FSRMDKWKDT	150
DNDELKVFFA	VMLLQGIVQK	PELEMFWSTR	PLLDTPYLRLQ	IMTGERFLLL	200
FRCLHFFVNS	SISAGQSKAQ	ISLQKIKPVF	DPLVNKFSTV	YTPNENIAVD	250
ESLMLFKGPL	AMKQYLPTRK	VRFGKLKLYVL	CESQSGYVWN	ALVHTGPGMN	300
LKDSADGLKS	SRIVLTIVND	LLGQGYCVFL	DNFNISPMLE	RELHQNRDFA	350
VGTLRLNRKQ	IPNDLKKRIA	KGTTVARFCG	ELMALKWCDG	KEVTMLSTFH	400
NDTVIEVNNR	NGKTKRPRV	IVDYNENMGA	VDSADQMLTS	YPSEKRRHKV	450
WYKKFFHHLL	HITVLNSYIL	FKKDNPEHTM	SHINFRLALI	ERMLEKHHKP	500
GQQHLLRGRFC	SDDVTPLRLS	GRHFPKSI PA	TSGKQNPTGR	CKICCSQYDK	550
DGKKIRKETR	YFCAECDVPL	CVVPCFEIYH	TKKNY		585

SEQ ID NO: 4: PGBD4 Hyperactive Mutant (S8P, G17R, K134K) Amino Acid Sequence (585 Amino Acids).

MSNPRKRPIP	MRDSNTRLEQ	LLAEDSFDES	DFSEIDSDSN	FSDSALEADK	50
IRPLSHLES	GKSSTSSDSG	RSMKWSARAM	IPRQRYDFTG	TPGRKVDVSD	100
ITDPLQYFEL	FFTEELVSKI	TRETNAQAAL	LASKPPGPKG	FSRMDKWKDT	150
DNDELKVFFA	VMLLQGIVQK	PELEMFWSTR	PLLDTPYLRLQ	IMTGERFLLL	200
FRCLHFFVNS	SISAGQSKAQ	ISLQKIKPVF	DPLVNKFSTV	YTPNENIAVD	250
ESLMLFKGPL	AMKQYLPTRK	VRFGKLKLYVL	CESQSGYVWN	ALVHTGPGMN	300
LKDSADGLKS	SRIVLTIVND	LLGQGYCVFL	DNFNISPMLE	RELHQNRDFA	350
VGTLRLNRKQ	IPNDLKKRIA	KGTTVARFCG	ELMALKWCDG	KEVTMLSTFH	400
NDTVIEVNNR	NGKTKRPRV	IVDYNENMGA	VDSADQMLTS	YPSEKRRHKV	450
WYKKFFHHLL	HITVLNSYIL	FKKDNPEHTM	SHINFRLALI	ERMLEKHHKP	500
GQQHLLRGRFC	SDDVTPLRLS	GRHFPKSI PA	TSGKQNPTGR	CKICCSQYDK	550
DGKKIRKETR	YFCAECDVPL	CVVPCFEIYH	TKKNY		585

SEQ ID NO: 6: PGBD1 Amino Acid Sequence (809 Amino Acids).

MYEALPGPAP	ENEDGLVKVK	EEDPTWEQVC	NSQEGSSHTQ	EICRLRFRRHF	CYQEAHGPPQ	60
ALAQLRELCH	QWLRPEMHTK	EQIMELLVLE	QPLTILPKEL	QPCVKTYPLE	SGEEAVTVLE	120
NLETGSGDTG	QQASVYIQGQ	DMHPMVAEYQ	GVSLECSLQ	LLPGITTLKC	EPFQRPQGNP	180
QEVSGFVPHG	SAHLQEKNEP	DKAVVPVFNP	VRSQTLVKTE	EETAQAVAAE	KWSHLSLSTR	240
NLCGNSAQET	VMSLSPMTTE	IVTKDRLFKA	KQETSEEMEQ	SGEASGKPNR	ECAPQIPCST	300
PLATERTVAH	LNTLKDRHFG	DLWARMHISS	LEYAAGDITR	KGRKKDKARV	SELLQGLSFS	360
GDSDEKEDNE	PEIQPAQKKL	KVSCFPEKSW	TKRDIKPNFP	SWSALDSGLL	NLKSEKLNVP	420
ELFELFFDDE	TFNLIVNETN	NYASQKNVSL	EVTVQEMRCV	FGVLLLSGFM	RHPREMYWE	480
VSDTDQNLVR	DAIRDRDFEL	IFSNLHFADN	GHLDDQKDRFT	KLRPLIKQMN	KNFLLYAPLE	540
EYYCFDKSMC	ECFDSQDFLN	GKPIRIGYKI	WCGTTTQGYL	VWFEPYQEE	TMKVDEDPDL	600
GLGGNLVMNF	ADVLLERGQY	FYHLCFDSFF	TSVKLLSALK	KKGVRATGTT	RENRTKCPPL	660
MNVEHMKMKM	RGYPDFRIEE	NNEIILCRWY	GDGIISLCSN	AVGIEPVNEV	SCCDADNEEI	720

PQISQPSIVK VYDECKEGVA KMDQIISKYR VRIRSKKWYS ILVSYMIDVA MNNAWQLHRA 780
 CNPGASLDPL DFRRFVAHFY LEHNAHLSL 809

SEQ ID NO: 7: PGBD2 Amino Acid Sequence (592 Amino Acids).

MASTRDVIA GRGIHSKVK AKLEVLNAM EEEESNNRE EIFIAPPDNA AGEFTDEDSG 60
 DEDSQPGAH L PGSVLHASVL CEDSGTGEDN DDLELQPAKK RQKAVVKPQR IWTKRDIRPD 120
 PGSWTASDPH IEDLKSQELS PVGLFELEFFD EGTINFIVNE TNRYAWQKNV NLSLTAQELK 180
 CVLGILILSG YISYPRRRMF WETSPDSHHH LVADAIRDR FELIFSYLHF ADNNELDASD 240
 RFAKVRPLI RMNCNFQKHA PLEEFYSFGE SMCEYFGHRG SKQLHRGKPV RLGYKIWCGT 300
 TSGYLVWFE PSQGTLETKP DRSLDLGGSM VIKFVDALQE RGFLPYHIFV DKVFTSVKLM 360
 SILRKKGVKA TGTVREYRTE RCPLKDPKEL KMKRGSFDY KVDESEIIV CRWHDSSVVN 420
 ICSNAVGLTP VRLTSRHSQA AKTRTQVHQP SLVKLYQEKV GGVGRMDQNI AKYVKIRGM 480
 KWYSSFIGYV IDAALNNAWQ LHRICQDAQ VDLAFLRYI ACVYLESNAD TTSQGRRSRR 540
 LETESRFDMI GHWIIHQDKR TRCALCHSQT NTRCEKCQKG VHAKCFREYH IR 592

SEQ ID NO: 8: PGBD3 Amino Acid Sequence (593 Amino Acids).

MPRTLSEHEI TDLETTDDSI EASAIVIQPP ENATAPVSDE ESGDEEGGTI NNLPGLLHT 60
 AAYLIQDGS D AESDSDPSY APKDDSPDEV PSTFTVQPP PSRRRKMTKI LCKWKKADLT 120
 VQFVAGRVA PPNDFFTVMR TPTILELEFL DDEVIELIVK YSNLYACSKG VHLGLTSSEF 180
 KCFLGIIFLS GYVSVPRRRM FWEQRTDVHN VLVSAAMRD RFETIFSNLH VADNANLDPV 240
 DKFSKLRPLI SKLNERCMKF VPNETYFSFD EFMVEYFGRH GCKQFIRGKP IRFGYKFWCG 300
 ATCLGYICWF QPYQGNPNT KHEEYGVGAS LVLQFSEALT EAHPGQYHFV FNNFFTSIAL 360
 LDKLSSMGHQ ATGTVRKDH DRVPLESDVA LKKKERGTDP YRIDGKGNIV CRWNDNSVVT 420
 VASSGAGIHP LCLVSRYSQK LKKKIQQVQP NMIKVYNQFM GGVDRADENI DKYRASIRGK 480
 KWYSSPLLFC FELVLQNAWQ LHKTYDEKPV DFLEFRRRV CHYLETHGHP PEPGQKGRPQ 540
 KRNIDSRYDG INHVIKQKG QTRCAECHKN TTRCEKCDV ALHVKCSVEY HTE 593

SEQ ID NO: 9: PGBD5 Amino Acid Sequence (524 Amino Acids).

MAEGGGGAER RAPALLAAR ARYSLHISD DVFGESGPDG GGNPFYSTSA ASRSSSAASS 60
 DDEREPPGEP GAAPPPRAP DAQPEEDEEA GAGWSAALRD RPPPRFEDTG GPTRKMPESA 120
 SAVDFFQLPV PDNVLKMMVV QTNMYAKKFP ERPGSDGAWV EVTLTEMKAF LGYMISTSIS 180
 HCESVLSIWS GGFYSNRSLA LVMSQARFEK ILKYFHVAF RSSQTHGLY KVQPFLLSLQ 240
 NSFDSAPRPS QTQVLHEPLI DEDPVFIATC TERELRKRKK RKFSLWVRQC SSTGFIIQIY 300
 VHLKEGGGPD GLDALKNKPQ LHSMVARSLC RNAAGKNYII FTGPSITSLT LFEEFEKQGI 360
 YCCGLLRARK SDCTGLPLSM LTNPATPPAR GQYQIKMKGN MSLICWYNGK HFRFLTNAYS 420
 PVQQGVIIKR KSGEIPCPPLA VEAFAAHLSY ICYDDKYSK YFISHKPNKT WQVWFVAIS 480
 IAINNAYILY KMSDAYHVKR YSRAQFGERL VRELLGLEDA SPTH 524

SEQ ID NO: 10: *Myotis lucifugus* (Wild-type) Amino Acid Sequence with Hyperactive Mutations (S8P, C13R, N125K) (572 Amino Acids).

MSQHSDDYDD EFRADKLSNY SCDSLENAS TSDEDSSDDE VMVRPRTLRR RRISSSSSDS 60
 ESDIEGGREE WSHVDNPPVL EDPLGHQGLN TDAVINNIED AVKLFIGDDF FEFLVEESNR 120
 YYNQKRNNFK LSKKSLKWKD ITPQEMKKFL GLIVLMGQVR KDRRDDYWT EPWTETPYFG 180
 KTMTRDRFRQ IWKAWHFNNN ADIVNESDRL CKVRPVLDF VPKFINIYKP HQQLSLDEGI 240
 VPWRGRLFFR VYNAGKIVKY GILVRLLCES DTGYICNMEI YCGEGKRLLE TIQTVVSPYT 300
 DSWYHIYMDN YNSVANACEA LMKNKFRICG TIRKNGIPK DFQTISLKKG ETKFIRKNDI 360
 LLQVWQSKKP VYLISIIHSA EMEESQNIDR TSKKKIVKPN ALIDYNKHMK GVDRAQYLS 420
 YYSILRRITV WTKRLAMYMI NCALFNYSYV YKSVRQRKM FKMFLKQTAI HWLTDDIPED 480
 MDIVPDLQPV PSTSGMRAPK PTDSPFCRLS MDMRKHTLQA IVGSGKKKNI LRRCRVCSVH 540
 KLRSETRYMC KFCNIPLHKG ACEPKYHTLK NY 572

SEQ ID NO: 11: *Myotis lucifugus* Corrected Amino Acid Sequence with Hyperactive Mutations (S8P, C13R) (571 Amino Acids).

MAQHSDDYDD EFRADKLSNY SCDSLENASTSDEDSSDDEVVMVRPRTLRRRISSSSSDSESDIEGGREEWSHVDNPPVLEDPLGH
 QGLNTDAVINNIEDAVKLFIGDDFFEFLVEESNRYNNQNRNNFKLSKSLKWKDITPQEMKKFLGLIVLMGQVRKDRRDDYWTTEP

WTETPYFGKTMTRDRFRQIWKAWHFNNNADI VNESDR LCKVRPVL DYFVPKFINIYKPHQQLSLDEGIVPWRGR LFFRVYNAGKI V
 KYGILVRL LCESDTGYICNMEIYCGEGKR LLETIQTVVSPYTD SWYHIYMDNYNSVANCEALMKNKFRICGTIRKNRGIPKDFQT
 ISLKKGETKFI RKN DILLQVWQSKKFVYL ISSIHS AEMEE SQNIDRTSKKKIVKPNALIDY NKHMKGVDRADQYL SYYSILRRTVK
 WTKRLAMYMIN CALFNSYAVYKSVRQRKMGFKMFLKQTAIHWLTD DI PEDMDI VPDLPVPSTSGMRAPPTSDPPCRLSMDMRKH
 TLQAI VSGKKKNILRRCRVCSVHKL RSETRYMCKFCNI PLHGACFEKYHTLKNY

SEQ ID NO: 430: Extended *Pteropus vampyrus* Amino Acid Sequence 584 Amino Acids.

MSNPRKRSIP	TCDVNFVLEQ	LLAEDSFDES	DFSEIDDSDD	FSDSASEDYT	VRPPSDSESD	60
GNSPTSADSG	RALKWSTRVM	IPRQRYDFTG	TPGRKVDVSD	TTDFLQYFEL	FFTEELVSKI	120
TSEMNAQAAL	LASKPPGPKG	FSRMDKWKDT	DNDELKVFFA	VMLLQGIVQK	PELEMPWSTR	180
PLLDIPYLRQ	IMTGERFLLL	LRCLHFNNS	SISAGQSKAQ	ISLQKIKPVF	DFLVNKFSTV	240
YTPNRNIAVD	ESLMLFKGRL	AMKQYIPTKC	ARFGLKLYVL	CESQSGYVWN	ALVHTGPSMN	300
LKDSADGLKS	SCIVLTLVND	LLGGQYCVFL	NNFYTSPMLF	RELHQNR TDA	VG TARLNRKQ	360
MFNDLKKRIA	KGTTVARFCG	ELMALKWCDK	KEVTMLSTFH	NDTVIEVDNR	NGKKTKKPCV	420
IVDYNENMGA	VDSADQMLTS	YPTERRRHKF	WYKKFFRHLL	NITVLNSYIL	FKKDNPEHTI	480
SHVNFRLLTI	ERMLEKHHKF	GQQRLRGRPC	SDDVTPLRLS	GRHFPSKIPP	TSKGQNPTGR	540
CKVCCSHDKD	GKKIRRETLY	PCAECDVPLC	VVPCFEIYHT	KKNY		

In embodiments, the mobile element enzyme has at least about 95%, or at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 430.

In embodiments, the mobile element enzyme has one or more mutations which confer hyperactivity.

In embodiments, the mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 10, or SEQ ID NO: 11 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has the nucleotide sequence having at least about 90% identity to SEQ ID NO: 5 or a codon-optimized form thereof.

SEQ ID NO: 5: PGBD4 Hyperactive Mutant (S8P, G17R, K134K) Nucleotide Sequence (1758 bp).

ATGTCAAATC	CTAGAAAACG	TCCCATTCCCT	ATGCGTGATA	GTAATACCCG	TCTCGAACAG	60
TTGTTGGCTG	AAGATTCATT	TGATGAATCT	GATTTTTCGG	AAATAGATGA	TTCTGATAAT	120
TTTTCCGGATA	GTGCTTTAGA	AGCCGATAAG	ATCAGGCCCTC	TGTCCCATT	AGAATCTGAT	180
GGAAAGAGCT	CTACATCAAG	TGACTCAGGG	CGCTCCATGA	AATGGTCAGC	TCGTGCTATG	240
ATTCCACGTC	AAAGGTATGA	CTTTACCGGC	ACACCTGGCA	GAAAAGTCCA	TGTCAGTGAT	300
ATCACTGACC	CATTGCAGTA	TTTGAACCTG	TTCTTTACTG	AGGAATTAGT	TTCAAAAATT	360
ACTAGAGAAA	CAAATGCCCA	AGCTGCCTTG	TTGGCTTCAA	AGCCACCGGG	TCCGAAAGGA	420
TTTTCCGGAA	TGGATAAATG	GAAAGACACT	GACAATGACG	AGCTCAAAGT	CTTTTTTGCA	480
GTAATGTTAC	TGCAAGGTAT	TGTGCAGAAA	CCTGAGCTGG	AGATGTTTGG	GTCAACAAGG	540
CCTCTTTTGG	ATACACCTTA	TCTCAGGCAA	ATTATGACTG	GTGAAAGATT	TTTACTTTTG	600
TTTCGGTGCC	TGCATTTTGT	CAACAATTCT	TCTATATCTG	CTGGTCAATC	AAAGGCCAG	660
ATTTCAITGC	AGAAGATCAA	ACCTGTGTTC	GACTTTCTTG	TAAATAAATT	TTCCACTGTA	720
TATACTCCAA	ACAGAAACAT	TGCAGTTGAT	GAATCACTGA	TGCTGTTCAA	GGGGCCATTA	780
GCTATGAAGC	AGTACCTCCC	GACAAAACGA	GTACGATTTG	GTCTGAAGCT	ATATGTACTT	840
TGTGAAAGTC	AGTCTGGTTA	TGTGTGGAAT	GCGCTTGTTT	ACACAGGGCC	TGGCATGAAT	900
TTGAAAGATT	CAGCGGATGG	CCTGAAATCA	TCACGCATTG	TTCTTACCTT	GGTCAATGAC	960
CTTCTTGGCC	AAGGGTATTG	TGTCTTCCTC	GATAACTTTA	ATATATCTCC	CATGCTTTTC	1020
AGAGAATTAC	ATCAAAATAG	GACTGATGCA	GTTGGGACAG	CTCGTTTGAA	CAGAAAACAG	1080
ATTCCAAATG	ATCTGAAAAA	AAGGATTGCA	AAGGGGACGA	CTGTAGCCAG	ATTCTGTGGT	1140
GAACCTATTG	CAGTGAATATG	GTGTGACGGC	AAGGAGGTGA	CAATGTGTTC	AACATTCAC	1200
AATGATACTG	TGATTGAAGT	AAACAATAGA	AATGGAAGA	AAACTAAAAG	GCCACGTGTC	1260
ATTGTGGATT	ATAACGAGAA	TATGGGAGCA	GTCGACTCGG	CTGATCAAAT	GCTTACTTCT	1320
TATCCATCTG	AGCGCAAAAG	ACACAAGGTT	TGGTATAAGA	AATTCTTTCA	CCATCTTCTA	1380

CACATTACAG	TGCTGAACTC	CTACATCCTG	TTCAAGAAGG	ATAATCCTGA	GCACACGATG	1440
AGCCATATAA	ACTTCAGACT	GGCATTGATT	GAAAGAATGC	TGGAAAAGCA	TCACAAGCCA	1500
GGGCAGCAAC	ATCTTCGAGG	TCGTCCCTGC	TCCGATGATG	TCACACCTCT	TCGTCTGTCT	1560
GGAAGACATT	TCCCCAAGAG	CATACCAGCA	ACGTCCGGGA	AACAGAATCC	AACTGGTCGC	1620
TGCAAAATTT	GCTGCTCCCA	ATACGACAAG	GATGGCAAGA	AGATCCGGAA	AGAAACGCGC	1680
TATTTTTGTG	CCGAATGTGA	TGTTCCGCTT	TGTGTTGTTT	CGTGCTTTGA	AATTACCAC	1740
ACGAAAAAAA	ATTATTAA					1758

In embodiments, the mobile element enzyme has an amino acid sequence having S8P and G17R mutations relative to the amino acid sequence of SEQ ID NO: 3 or SEQ ID NO: 4, or a functional equivalent thereof. In embodiments, the mobile element enzyme has an amino acid sequence having I83P and/or V118R mutation relative to the amino acid sequence of SEQ ID NO: 6 or a functional equivalent thereof. In embodiments, the mobile element enzyme has an amino acid sequence having S20P and/or A29R mutation relative to the amino acid sequence of SEQ ID NO: 7 or a functional equivalent thereof. In embodiments, the mobile element enzyme has an amino acid sequence having A12P and/or I28R mutation and/or R152K mutation relative to the amino acid sequence of SEQ ID NO: 9 or a functional equivalent thereof. In embodiments, the mobile element enzyme has an amino acid sequence having T4P and/or L13R mutation relative to the amino acid sequence of SEQ ID NO: 8 or a functional equivalent thereof.

In embodiments, the donor DNA is included in a vector comprising left and right end sequences recognized by the mobile element enzyme.

In embodiments, the end sequences are selected from MER, MER75A, MER75B, and MER85. In embodiments, the end sequences are selected from nucleotide sequences of SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 441, and SEQ ID NO: 22, or a nucleotide sequence having at least about 90% identity thereto.

SEQ ID NO: 12: *Pteropus vampyrus* Left End Sequence

Sequence 381 bp.

TTAACCCTATT	TCTCTTTTGC	CCCGAGAATA	CTCACCAGCG	GCACTTGCGAG	CTGCAGCGTT	60
TACCCCGAGA	TAACCTGTCG	ATTACAGTCC	TAACCTTACC	CCCAAAGTTT	GCCATGAAAT	120
ATCTCGCTTT	TATTATTATT	TTGCGATCGC	TCTAGTATAT	CGATAGTCTT	TGGAAACAAA	180
TGACATCATT	CTATTTACAG	CATTCTGTTT	TTAGTAGTGG	TATTTCCATT	TACAAAATAT	240
AGTAATTTTC	TATCGCTGAA	AATGTCAAAT	CCTAGAAAAC	GTAGCATTCC	TACATGTGAT	300
GTTAACTTCG	TTCTCGAACA	GTTGTTAGCC	GAAGATTCAT	TTGATGAATC	CGATTTTTC	360
GAAATAGACG	ATTCTGATGA	T				381

SEQ ID NO: 13: PGBD4 Left End Nucleotide Sequence

Sequence 373 bp.

TTAACTCATT	TCTCCTTAGC	CCCGAGATTA	CGCGCTGCTG	TGCCTGCGAC	TGCAGCGTTT	60
ACGCCGAGAT	AACTCGTGGA	TTACAGTGCC	AACCTTACTC	CCAAAGTTTG	CCACGAAATA	120
TCTCGCTTCT	GTTATTTTCG	CATGGTTCTG	GTATATTGAC	TTTTGAAACA	AAAGACATCA	180
TTCTGTTTAT	AGCATCTCTG	TTTGTAGTAGT	GGGATTTCCA	TCTACAAAAT	ATAGTAATTC	240
TCGATCGCTG	AAATGTCAAA	TCCTAGAAAA	CGTAGCATT	CTATGCGTGA	TAGTAATACC	300
GGTCTCGAAC	AGTTGTTGGC	TGAAGATTCA	TTTGATGAAT	CTGATTTTTC	GGAAATAGAT	360
GATTCTGATA	ATT					373

SEQ ID NO: 14: MER75 Left End Nucleotide Sequence

Sequence 344 bp.

TTAACCCTTT	TCCCGTTTGC	CCCGAGAATA	CTCGCCGGCG	GCGCTTGCGG	CTGCAGCGTT	60
TACCCCGAGA	TAACTTTGCC	ACGAAATATC	TCGCTTTTAT	TATTATTTTC	GCATCGCTCT	120
AGTATATCGA	CTTTGGAAAC	AAAAGACATC	ATTCTATTTA	TAGCATTCTG	TTTTTAGTAG	180
TGGTATTTCC	ATTTACAAAA	TATAGTAATT	CTCGATCGCT	GAAAATGTCA	AATCCTAGAA	240
AACGTAGCAT	TCCTACGCGT	GATGTTAACA	TCGTTCTCGA	ACAGTTGTTG	GCCGAAGATT	300
CATTTGATGA	ATCCGATTTT	TCCGAAATAG	ACGATTCTGA	TGAT		344

SEQ ID NO: 15: MER75B Left End Nucleotide Sequence

Sequence 91 bp.

TTAACCCTTT	TCCCGTTTGC	CCCGAGAATA	CTCTTGTCTC	TAATCCTAAT	GTAACATCAT	60
ATACATTTCT	GTTACATTAG	GATTAGAGAC	A			91

SEQ ID NO: 16: MER75A Left End Nucleotide Sequence

Sequence 32 bp.

TTAACCCTTT	TCCCGTTTGC	CCCGAGAATA	CT	32
------------	------------	------------	----	----

SEQ ID NO: 17: *Pteropus vampyrus* Right End Sequence

Sequence 171 bp.

TAGGATTAGA	GACAAGTTCT	GTTTAGAAAT	AACTCCAAGA	ACAGTTTTTA	TATTTTATTT	60
TCACATTGAA	AACCAGTCAG	ATTGCTTCA	GCCTCAAAGA	GCATGTTTAT	GTAAAATTAA	120
ATTAACGCTG	GCAGCGAGCT	GCACTTTTTT	TCTAAACGGG	AAATGGGTTA	A	171

SEQ ID NO: 18: PGBD4 Right End Nucleotide Sequences

Sequence 176 bp.

CCTGGGATTA	TAGGCATGAG	CCACTGCGCC	TAGCACCAAG	AACAGTTTTT	ATATTTTATT	60
TTACATTGAA	AAATCAGTCA	GATTGCTTTC	AGCCTCAAAG	AGGGGTGTTA	TGTAAACTAA	120
AATGAGTGCA	GGCAGCGAGC	TACACTTTTT	TTTTTCCTAA	ATGGAAAAATG	GGTTAA	176

SEQ ID NO: 19: MER75 Right End Nucleotide Sequences

Sequence 178 bp.

TCAGACGATT	CTGATGTTAG	TTCTGTTTAG	AAATAACTCC	AAGAACAGTT	TTTATATTTT	60
ATTTTCACAT	TGAAAATCAG	TCAGATTTGC	TTGAGCCTCA	AAGAGCGTGT	TTATGTAAAA	120
TTAAATGAGC	GCTGGCAGCG	AGCTGCACCT	TTTTTTTCT	AAACGGGAAA	AGGGTTAA	178

SEQ ID NO: 20: MER75B Right End Nucleotide Sequences

Sequence 160 bp.

AGTTCTGTTT	AGAAATAACT	CCAAGAACAG	TTTTTATATT	TTATTTTCAC	ATTGAAAATC	60
AGTCAGATTT	GCTTCAGCCT	CAAAGAGCGT	GTTTATGTAA	AATTAAATGA	GCGCTGGCAG	120
CGAGCTGCAC	TTTTTTTTTT	CTAAACGGGA	AAAGGGTTAA			160

SEQ ID NO: 21: MER75A Right End Nucleotide Sequences

Sequence 46 bp.

CGCTGGCAGC	GAGCTGCACT	TTTTTTCTAA	ACGGGAAATG	GGTTAA	46
------------	------------	------------	------------	--------	----

In embodiments, one or more of the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 12, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 12 is positioned at the 5' end of the donor DNA.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 17, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 17 is positioned at the 3' end of the donor DNA.

In embodiments, the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 13, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 13 is positioned at the 5' end of the donor DNA.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 18, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 18 is positioned at the 3' end of the donor DNA.

In embodiments, the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 14, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 14 is positioned at the 5' end of the donor DNA.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 19, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 19 is positioned at the 3' end of the donor DNA.

In embodiments, the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 15, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 15 is positioned at the 5' end of the donor DNA.

In embodiments, end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 20, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 20 is positioned at the 3' end of the donor DNA.

In embodiments, the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 16, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 16 is positioned at the 5' end of the donor DNA.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441 is positioned at the 3' end of the donor DNA.

In embodiments, the mobile element enzyme is an engineered form of a mobile element enzyme reconstructed from *Homo sapiens* or a predecessor thereof.

In embodiments, the enzyme is in a monomeric or dimeric form. In embodiments, the enzyme is in a multimeric form.

In embodiments, the method of the present disclosure provides an enzyme comprising (a) a targeting element, and (b) an enzyme that is capable of inserting the donor DNA comprising a transgene, optionally at a TA dinucleotide site or a TTAA (SEQ ID NO: 440) tetranucleotide site in a genomic safe harbor site (GSHS).

In embodiments, the donor DNA comprises a transgene encoding a complete polypeptide. In embodiments, the donor DNA comprises a transgene which is defective or substantially absent in a disease state. In embodiments, the enzyme has one or more mutations which confer hyperactivity. In embodiments, the enzyme has gene cleavage (Exc) and/or gene integration activity (Int+). In embodiments, the enzyme has gene cleavage (Exc) and/or a lack of gene integration activity (Int-).

In embodiments, the mobile element enzyme is a chimeric mobile element enzyme.

In embodiments, the targeting element comprises one or more of a gRNA, optionally associated with a Cas enzyme, which is optionally catalytically inactive, transcription activator-like effector (TALE), catalytically inactive Zinc finger, catalytically inactive transcription factor, nickase, a transcriptional activator, a transcriptional repressor, a recombinase, a DNA methyltransferase, a histone methyltransferase, a paternally expressed gene 10 (PEG10), and a TnsD.

In embodiments, the targeting element comprises a transcription activator-like effector (TALE) DNA binding domain (DBD). In embodiments, the TALE DBD comprises one or more repeat sequences. In embodiments, the TALE DBD comprises about 14, or about 15, or about 16, or about 17, or about 18, or about 18.5 repeat sequences. In embodiments, the TALE DBD repeat sequences comprise 33 or 34 amino acids. In embodiments, the one or more of the TALE DBD repeat sequences comprise a repeat variable di-residue (RVD) at residue 12 or 13 of the 33 or 34 amino acids. In embodiments, the RVD recognizes one base pair in the nucleic acid molecule. In embodiments, the RVD recognizes a C residue in the nucleic acid molecule and is selected from HD, N(gap), HA, ND, and HI. In embodiments, the RVD recognizes a G residue in the nucleic acid molecule and is selected from NN, NH, NK, HN, and NA. In embodiments, the RVD recognizes an A residue in the nucleic acid molecule and is selected from NI and NS.

In embodiments, the RVD recognizes a T residue in the nucleic acid molecule and is selected from NG, HG, H(gap), and IG. In embodiments, the GSHS is in an open chromatin location in a chromosome.

In embodiments, the GSHS is selected from adeno-associated virus site 1 (AAVS1), chemokine (C-C motif) receptor 5 (CCR5) gene, HIV-1 coreceptor, and human Rosa26 locus. In embodiments, the GSHS is located on human chromosome 2, 4, 6, 10, 11, 17, 22, or X.

In embodiments, the GSHS is selected from TALC1, TALC2, TALC3, TALC4, TALC5, TALC7, TALC8, AVS1, AVS2, AVS3, ROSA1, ROSA2, TALER1, TALER2, TALER3, TALER4, TALER5, SHCHR2-1, SHCHR2-2, SHCHR2-3, SHCHR2-4, SHCHR4-1, SHCHR4-2, SHCHR4-3, SHCHR6-1, SHCHR6-2, SHCHR6-3, SHCHR6-4, SHCHR10-1, SHCHR10-2, SHCHR10-3, SHCHR10-4, SHCHR10-5, SHCHR11-1, SHCHR11-2, SHCHR11-3, SHCHR17-1, SHCHR17-2, SHCHR17-3, and SHCHR17-4.

In embodiments, the targeting element comprises a Cas9 enzyme guide RNA complex. In embodiments, the Cas9 enzyme guide RNA complex comprises a nuclease-deficient dCas9 guide RNA complex. In embodiments, the targeting element comprises a Cas12 enzyme guide RNA complex or wherein the targeting element comprises a nuclease-deficient dCas12 guide RNA complex, optionally dCas12j guide RNA complex or dCas12a guide RNA complex.

In embodiments, the targeting element comprises: a gRNA of or comprising a sequence of **TABLE 3A-3F**, or a variant thereof; or a TALE DBD of or comprising a sequence of **TABLE 4A-4F**, or a variant thereof; or a ZNF of or comprising a sequence of **TABLE 5A-5E**, or a variant thereof.

In embodiments, the targeting element is or comprises a nucleic acid binding component of the gene-editing system. In embodiments, the enzyme and the targeting element are connected. In embodiments, the enzyme and the targeting element are fused to one another or linked via a linker to one another. In embodiments, the linker is a flexible linker. In embodiments, the flexible linker is substantially comprised of glycine and serine residues, optionally wherein the flexible

linker comprises $(\text{Gly}_4\text{Ser})_n$, where n is from about 1 to about 12. In embodiments, the flexible linker is of about 20, or about 30, or about 40, or about 50, or about 60 amino acid residues.

In embodiments, the donor DNA comprises a gene encoding a complete polypeptide. In embodiments, the donor DNA comprises a gene which is defective or substantially absent in a disease state. In embodiments, the donor DNA is flanked by one or more inverted terminal ends.

In embodiments, at least one of the first nucleic acid and the second nucleic acid is in the form of a lipid nanoparticle (LNP). In embodiments, the first nucleic acid encoding the enzyme and the second nucleic acid encoding the donor DNA are in the form of the same LNP, optionally in a co-formulation. In embodiments, the LNP comprises one or more lipids selected from 1,2-dioleoyl-3-trimethylammonium propane (DOTAP), a cationic cholesterol derivative mixed with dimethylaminoethane-carbamoyl (DC-Chol), phosphatidylcholine (PC), triolein (glyceryl trioleate), and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[carboxy(polyethylene glycol)-2000] (DSPE-PEG), 1,2-dimyristoyl-rac-glycero-3-methoxypolyethyleneglycol – 2000 (DMG-PEG 2K), and 1,2 distearol -sn-glycerol-3phosphocholine (DSPC) and/or comprising of one or more molecules selected from polyethylenimine (PEI) and poly(lactic-co-glycolic acid) (PLGA), and N-Acetylgalactosamine (GalNAc).

In embodiments, the enzyme is encoded by a recombinant or synthetic nucleic acid. In embodiments, the nucleic acid is mRNA or a helper RNA. In embodiments, the nucleic acid is RNA that has a 5'-m7G cap (cap0, cap1, or cap2) with pseudouridinesubstitution, and a poly-A tail of about 30, or about 50, or about 100, of about 150 nucleotides in length.

In embodiments, the enzyme is incorporated into a vector or a vector-like particle. In embodiments, the vector is a non-viral vector. In embodiments, the enzyme and the donor DNA are included in the same vector. In embodiments, the enzyme and the donor DNA are included in different vectors.

In embodiments, the enzyme and the donor DNA are included in a single pharmaceutical composition. In embodiments, the enzyme and the donor DNA are included in different pharmaceutical compositions. In embodiments, the enzyme and the donor DNA are co-administered. In embodiments, the enzyme and the donor DNA are administered separately.

In embodiments, the present disclosure provides a method of producing an AAV bearing a gene of interest, comprising employing a method of the present disclosure to produce the AAV bearing the gene of interest. In embodiments, the present disclosure provides a cell for gene therapy, generated by a method of the present disclosure.

In embodiments, the present disclosure provides a method of delivering a cell therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of the present disclosure. In embodiments, the present disclosure provides a method of treating a disease or condition using a cell therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of the present disclosure.

In embodiments, the present disclosure provides a method of treating a disease or condition using a biologic, e.g., antibody, therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of the present disclosure.

Enzymes

In embodiments, the enzyme capable of targeted genomic integration is any type of an enzyme that cause a transgene to be inserted from one location (e.g., without limitation, donor DNA) to a specific site and/or locus in a subject's genome.

In embodiments, the enzyme capable of targeted genomic integration is a recombinase.

In embodiments, the recombinase is an integrase. In embodiments, the enzyme is a mobile element enzyme. In embodiments, the recombinase is an integrase or a mobile element enzyme.

In embodiments, the mobile element enzyme is an engineered mammalian mobile element enzyme. In embodiments, the mobile element enzyme is a mammal-derived, helper RNA mobile element enzyme. Messenger RNA (mRNA) is an effective alternative to DNA as a source of a mobile element enzyme for targeting somatic cells and tissues, given that RNA is a safer alternative to DNA as a source of a mobile element enzyme for somatic gene therapy applications. See, e.g., Wilber *et al.*, *Mol. Ther.* 13, 625–630 (2006). Successful use of *in vitro*-transcribed mRNA as a transient source of mobile element enzyme and subsequent transposition in cultured human cells and in live mice was previously reported for Sleeping Beauty mobile element enzyme. See Wilber *et al.*, *Mol. Ther.* 13, 625–630 (2006). It was demonstrated that *in vitro*-transcribed, UTR-stabilized mobile element enzyme-encoding mRNA can be used as a source of mobile element enzyme for Sleeping Beauty-mediated transposition in cultured somatic cells. Wilber *et al.*, *Mol. Ther.* 13, 625–630 (2006). Also, Hoerr *et al.* reported that a specific cytotoxic T cell response and circulating antigen-specific antibodies were detected after administration of *in vitro*-transcribed, UTR-stabilized, and protamine-condensed bacterial lacZ mRNA into the ear pinna of Balb/C mice. Hoerr *et al.*, *Eur. J. Immunol.* 2000; 30: 1-7; see also Wilber *et al.* (2006).

In embodiments, the mobile element enzyme is a mammal-derived, DNA mobile element enzyme. In embodiments, the mobile element enzyme is a chimeric mobile element enzyme.

In embodiments, the enzyme capable of targeted genomic integration is a mobile element enzyme, and the mobile element enzyme comprises (a) a targeting element that is or comprises a gene-editing system, and (b) a mobile element enzyme that is capable of inserting the donor DNA (e.g., mobile element) comprising a transgene at a TA dinucleotide site or a TTAA (SEQ ID NO: 440) tetranucleotide site in a GSHS, as described elsewhere herein.

In embodiments, the enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*. In embodiments, the enzyme is an

engineered version, including but not limited to hyperactive forms, of an enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

In embodiments, the enzyme is a mobile element enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*. In embodiments, the enzyme is an engineered version, including but not limited to hyperactive forms, of a mobile element enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

In embodiments, the mobile element enzyme is from one or more of the Tn1, Tn2, Tn3, Tn5, Tn7, Tn9, Tn10, Tn552, Tn903, Tn1000/Gamma-delta, Tn/O, tnsA, tnsB, tnsC, tniQ, IS10, ISS, IS911, Minos, Sleeping beauty, piggyBac, Tol2, Mos1, Himar1, Hermes, Tol2, Minos, Tel, P-element, MuA, Ty1, Chapaev, transib, Tc1/mariner, or Tc3 donor DNA system, or biologically active fragments variants thereof, inclusive of hyperactive mutants (e.g., without limitation selected from **TABLE 1**, or equivalents thereof).

In embodiments, the mobile element enzyme is from a MLT donor DNA system that is based on a cut-and-paste MLT element obtained from the little brown bat (*Myotis lucifugus*) or other bat mobile element enzymes, such as *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pipistrellus kuhlii*, and *Molossus molossus*. See Mitra *et al.*, *Proc Natl Acad Sci U S A*. 2013 Jan 2;110(1):234-9; Jebb *et al.*, *Nature*, volume 583, pages 578–584 (2020), which are incorporated by reference herein in their entireties. In embodiments, hyperactive forms of a bat mobile element enzyme are used. The MLT mobile element enzyme has been shown to be capable of transposition in bat, human, mammalian, and yeast cells. The hyperactive forms of the MLT mobile element enzyme enhance the transposition process. In addition, chimeric MLT mobile element enzymes are capable of site-specific excision without genomic integration.

In embodiments, the mobile element enzyme is a *Myotis lucifugus* mobile element enzyme (MLT), which is either the wild type, monomer, dimer, tetramer (or another multimer), hyperactive, an Int-mutant, or of any other form.

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, or a variant having at least about 80%, at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, and one or more mutations selected from L573X, E574X, and S2X, wherein X is any amino acid or no amino acid, optionally X is A, G, or a deletion, optionally the mutations are L573del E574del, and S2A). In embodiments, the MLT mobile element enzyme has the nucleotide sequence of SEQ ID NO: 2

(which is a codon-optimized form of MLT), or a nucleotide sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

SEQ ID NO: 1 is:

MAQHSDDYSDDEFCADKLSNYSCDSLENASTSDEDSSDDEVMVRPRTLRRRRISSSSSDSESDIEGGREWSHV
DNPPVLEDFLGHQGLNTDAVINNIEDAVKLFIGDDFFEFLVEESNRYYNQNRNNFKLSKKSLKWKDITPQEMKK
FLGLIVLMGQVRKDRDDYWTTEPWTETPYFGKTMTRDRFRQIWKAWHFNNNADIVNESDRLCKVRPVLDFYFVP
KFINIYKPHQQLSLDEGI VPWGRGLFFRVYNAGKIVKYGILVRLLCESDTGYICNMEIYCGEGRLLLETIQTVV
SPYTDSWYHIYMDNYNSVANCEALMKNKFRICGTIRKNRGIPKDFQTISLKKGETKFIRKNDILLQVWQSKKP
VYLISIIHSAEMEESQNIDRTSKKKIVKPNALIDYNKHMKGVDRAQYLSYYSILRRTVKWTKRLAMYMINCAL
FNSYAVYKSVRQPKMGFKMFLKQTAIHWLTDDIPEDMDIVPDLQVPVSTSGMRAKPPTSDPPCRLSMDMRKHTL
QAIVGSGKKKNILRRRCRVCSVHKLRSETRYMCKFCNIPLHKGACFEKYHTLKNY

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1 or a variant having at least about 80%, at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto and comprises an amino acid other than serine at the position corresponding to position 2 of SEQ ID NO: 1. In embodiments, the amino acid is a non-polar aliphatic amino acid, optionally a non-polar aliphatic amino acid optionally selected from G, A, V, L, I and P, optionally A. In embodiments, the mobile element enzyme does not have additional residues at the C terminus relative to SEQ ID NO: 1.

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1 or a variant having at least about 80%, at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto and comprises an alanine at the position corresponding to position 2 of SEQ ID NO: 1.

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1 or a variant having at least about 80%, at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto and comprises an alanine at the position corresponding to position 2 of SEQ ID NO: 1 and no additional amino acids at the C terminal end.

In embodiments, the MLT mobile element enzyme has a nucleotide sequence of SEQ ID NO: 2 (which is codon-optimized) and an amino acid sequence SEQ ID NO: 1, respectively. In embodiments, the MLT mobile element enzyme has a nucleotide sequence of SEQ ID NO: 2, or a nucleotide sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, or a codon-optimized form thereof. In embodiments, the MLT mobile element enzyme has an amino acid sequence SEQ ID NO: 1, or an amino acid sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

In embodiments, the mobile element enzyme can act on an MLT left terminal end, or a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, wherein the nucleotide sequence of the MLT left terminal end (5' to 3') is as follows:

ttaacactggattgogggaaaacgagttaagtcggctcgctgaattgcgcgtactccgogggagccgtcttaactcgggtcatatag
atttcgggtggagtgogggaaaacgtgtaactcgggccgattgtaactgcgtattacaaaatattgtt (SEQ ID NO: 21)

In embodiments, the mobile element enzyme can act on an MLT right terminal end, or a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, wherein the nucleotide sequence of the MLT right terminal end (5' to 3') is as follows:

aattattatgtactgaatagataaaaaaigtcgtgattgaataaattttacacaaagaaacgaaaatttcattcaatcgaa
occatacttcaaaagatataggcattttaactaactctgatttgcgogggaaaacctaaataattgcgcgcgcatttatatttggcg
ggaaaitcacccgacacgtAgtgttaa (SEQ ID NO: 22)

In embodiments, the donor DNA is flanked by one or more terminal ends. In embodiments, the donor DNA is or comprises a gene encoding a complete polypeptide. In embodiments, the donor DNA is or comprises a gene which is defective or substantially absent in a disease state.

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme, e.g., without limitation, MLT mobile element enzyme), inclusive of any described herein has one or more mutations which confer hyperactivity.

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme, e.g., without limitation, MLT mobile element enzyme) has gene cleavage activity (Exo+) and/or gene integration activity (Int+). In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme, e.g., without limitation, MLT mobile element enzyme) has gene cleavage activity (Exo+) and/or a lack of gene integration activity (Int-).

In embodiments, the mobile element enzyme, e.g., without limitation, MLT mobile element enzyme includes a hyperactive mutation, e.g., about 1, or about 2, or about 3, or about 4, or about 5 hyperactive mutations or combinations thereof. In embodiments, the mobile element enzyme can include any number of any of the hyperactive mutations, or equivalents thereof, described herein.

In embodiments, the MLT mobile element enzyme includes a hyperactive mutation, e.g., about 1, or about 2, or about 3, or about 4, or about 5 hyperactive mutations, or combinations thereof. In embodiments, the mobile element enzyme can include any number of any of the hyperactive mutations, or equivalents thereof, described herein.

In embodiments, the enzyme comprises one or more mutations corresponding to **TABLE 1**, which, without being bound by theory, provides hyperactive mutations. Numbering relative to the amino acid sequence of protein of SEQ ID NO: 1, and nucleic acid sequence of SEQ ID NO: 2.

TABLE 1:

Nucleotide Change	Amino Acid Change
T13C	S5P
T22C	S8P
T22C/T37C	S8P/C13R
A26G	D9G
A29G	D10G
A32G	E11G
T37C	C13R
C41T	A14V
A106G	S36G
G161A	S54N
T375G	N125K
A389C	K130T
G715A	G239S
A880G	T294A
A898G	T300A
A1033G	I345V
G1280A	R427H
A1424G	D475G
A1441G	M481V
C1472A	P491Q
G1558A	A520T
G1681A	A561T

In embodiments, the MLT mobile element enzyme has one or more amino acid substitutions selected from S8X1, C13X2 and/or N125X3, or positions corresponding thereto, relative to SEQ ID NO: 1, wherein X1 is selected from G, A, V, L, I and P, X2 is selected from K, R, and H, and X3 is selected from K, R, and H, or wherein: X1 is P, X2 is R, and/or X3 is K.

In embodiments, the MLT mobile element enzyme has S8X1, C13X2 and N125X3 substitutions, at positions corresponding to SEQ ID NO: 1, wherein X1 is selected from G, A, V, L, I and P, X2 is selected from K, R, and H, and X3 is selected from K, R, and H, or wherein: X1 is P, X2 is R, and/or X3 is K.

In embodiments, the MLT mobile element enzyme has S8X1 and C13X2 substitutions, at positions corresponding to SEQ ID NO: 1, wherein X1 is selected from G, A, V, L, I and P, X2 is selected from K, R, and H, and X3 is selected from K, R, and H, or wherein: X1 is P, X2 is R, and/or X3 is K.

In embodiments, the MLT mobile element enzyme has S8X₁ and N125X₃ substitutions, at positions corresponding to SEQ ID NO: 1, wherein X₁ is selected from G, A, V, L, I and P, X₂ is selected from K, R, and H, and X₃ is selected from K, R, and H, or wherein: X₁ is P, X₂ is R, and/or X₃ is K.

In embodiments, the MLT mobile element enzyme has C13X₂ and N125X₃ substitutions, at positions corresponding to SEQ ID NO: 1, wherein X₁ is selected from G, A, V, L, I and P, X₂ is selected from K, R, and H, and X₃ is selected from K, R, and H, or wherein: X₁ is P, X₂ is R, and/or X₃ is K.

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, or a variant thereof, and S8P and C13R mutations (SEQ ID NO: 11). In embodiments, the MLT mobile element enzyme has an amino acid sequence having mutations at positions which correspond to at least one of S8P and C13R mutations relative to the amino acid of SEQ ID NO: 1 or a functional equivalent thereof. In embodiments, the MLT mobile element enzyme has an amino acid sequence having mutations at positions which correspond to S8P and C13R mutations relative to the amino acid of SEQ ID NO: 1 or a functional equivalent thereof.

In embodiments, the MLT mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, or a variant thereof, and S8P, C13R, and N125K mutations (SEQ ID NO: 10).

In embodiments, a MLT mobile element enzyme comprising the amino acid sequence of SEQ ID NO: 1, or a variant thereof, and includes one or more hyperactive mutations selected from a substitution or deletion at one or more of positions S5, S8, D9, D10, E11, C13, A14, S36, S54, N125, K130, G239, T294, T300, I345, R427, D475, M481, P491, A520, and A561, or positions corresponding thereto.

In embodiments, a MLT mobile element enzyme comprising the amino acid sequence of SEQ ID NO: 1, or a variant thereof, and includes one or more hyperactive mutations selected from S5P, S8P, S8P/C13R, D9G, D10G, E11G, C13R, A14V, S36G, S54N, N125K, K130T, G239S, T294A, T300A, I345V, R427H, D475G, M481V, P491Q, A520T, and A561T, or positions corresponding thereto.

In embodiments, the MLT mobile element enzyme comprises one or more of hyperactive mutants selected from S8X₁, C13X₂ and/or N125X₃ (e.g., all of S8X₁, C13X₂ and N125X₃, S8X₁ and C13X₂, S8X₁ and N125X₃, and C13X₂ and N125X₃), where X₁, X₂, and X₃ is each independently any amino acid, or X₁ is a non-polar aliphatic amino acid, selected from G, A, V, L, I and P, X₂ is a positively charged amino acid selected from K, R, and H, and/or X₃ is a positively charged amino acid selected from K, R, and H. In embodiments, X₁ is P, X₂ is R, and/or X₃ is K.

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme, e.g., without limitation, MLT mobile element enzyme) has gene cleavage activity (Exc+) and/or gene integration activity (Int+). In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme) has gene cleavage activity (Exc+) and/or a lack of gene integration activity (Int-). In embodiments, the MLT mobile element enzyme has gene cleavage activity (Exc+) and/or gene

integration activity (Int+). In embodiments, the MLT mobile element enzyme has gene cleavage activity (Exc+) and/or a lack of gene integration activity (Int-).

In embodiments, the mobile element enzyme, e.g., without limitation, MLT mobile element enzyme includes an integration reduced or deficient mutation, e.g., about 1, or about 2, or about 3, or about 4, or about 5 integration reduced or deficient mutations or combinations thereof. In embodiments, the mobile element enzyme can include any number of any of the integration reduced or deficient mutations, or equivalents thereof, described herein.

In embodiments, the MLT mobile element enzyme includes a integration reduced or deficient mutations, e.g. about 1, or about 2, or about 3, or about 4, or about 5 integration reduced or deficient mutations, or combinations thereof. In embodiments, the mobile element enzyme can include any number of any of the integration reduced or deficient mutations, or equivalents thereof, described herein.

In embodiments, the enzyme comprises one or more mutations corresponding to **TABLE 2A**, or positions corresponding thereto, which, without being bound by theory, provides integration reduced or deficient mutations. Numbering relative to the amino acid sequence of protein of SEQ ID NO: 1.

TABLE 2A:

Amino Acid Change
Y281A
C282A
G283A
E284A
G285A
N310A
G330A
T331A
I332A
R333A
K334A
N335A
R336A
G337A
I338A
P339A
D416A
K286A
R287A
N310A

K286A/R287A
K286A/N310A
K286A/K369A
R287A/N310A
R287A/K369A
N310A/K369A
R287A/N310A

In embodiments, the enzyme comprises one or more mutations corresponding to **TABLE 2B**, or positions corresponding thereto, which, without being bound by theory, provides excision positive and integration deficient mutations. Numbering relative to the amino acid sequence of protein of SEQ ID NO: 1, and nucleic acid sequence of SEQ ID NO: 2.

TABLE 2B:

MLT Backbone	MLT Mutant 1	MLT Mutant 2	MLT Mutant 3
S8P/C13R	R164N	0	0
S8P/C13R	W168V	0	0
S8P/C13R	W168V	K369A	0
S8P/C13R	M278A	0	0
S8P/C13R	K286A	0	0
S8P/C13R	K286A	R287	
S8P/C13R	K286A	N310N	
S8P/C13R	K286A	K369A	
S8P/C13R	R287A	0	0
S8P/C13R	R287A	N310A	
S8P/C13R	R287A	K369A	
S8P/C13R	R287A	N310A	K369A
S8P/C13R	N310A	K369A	
S8P/C13R	R333A	0	0
S8P/C13R	R333A	E284A	0
S8P/C13R	R333A	E284A	R336A
S8P/C13R	K334A	0	0
S8P/C13R	N335A	0	0
S8P/C13R	K349A	0	0
S8P/C13R	K350A	0	0
S8P/C13R	K368A	0	0

S8P/C13R	K369A	0	0
S8P/C13R	D416N	0	0
S8P/C13R	D416N	K286A	0
S8P/C13R	D416N	R287A	0
S8P/C13R	D416N	R333A	0
S8P/C13R	D416N	K334A	0
S8P/C13R	D416N	R336A	0
S8P/C13R	D416N	K349A	0
S8P/C13R	D416N	K350A	0
S8P/C13R	D416N	K368A	0
S8P/C13R	D416N	K369A	0
S8P/C13R	D416N	N310A	0

In embodiments, a MLT mobile element enzyme comprising the amino acid sequence of SEQ ID NO: 1, or a variant thereof, and includes one or more mutations selected from S8P and/or C13R and one of R164N, W168V, M278A, K286A, R287A, R333A, K334A, N335A, K349A, K350A, K368A, K369A, and D416N, or positions corresponding thereto.

In embodiments, a MLT mobile element enzyme comprising the amino acid sequence of SEQ ID NO: 1, or a variant thereof, and includes one or more mutations selected from S8P and/or C13R and one of R164N, W168V, M278A, K286A, R287A, R333A, K334A, N335A, K349A, K350A, K368A, K369A, and D416N and/or one or more of E284A, K286A, R287A, N310A, R333A, K334A, R336A, K349A, K350A, K368A, and K369A, or positions corresponding thereto.

In embodiments, a MLT mobile element enzyme comprising the amino acid sequence of SEQ ID NO: 1, or a variant thereof, and includes one or more mutations selected from S8P and/or C13R and one of R164N, W168V, M278A, K286A, R287A, R333A, K334A, N335A, K349A, K350A, K368A, K369A, and D416N and/or one or more of E284A, K286A, R287A, N310A, R333A, K334A, R336A, K349A, K350A, K368A, and K369A and/or one R336A, or positions corresponding thereto.

In embodiments, the mobile element enzyme is or is derived from any of *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Myotis lucifugus*, *Pipistrellus kuhlii*, *Pteropus vampyrus*, and *Molossus molossus*. In embodiments, the mobile element enzyme is or is derived from any of *Trichoplusia ni* (SEQ ID NO: 433), *Myotis myotis* (SEQ ID NO: 435, SEQ ID NO: 436, SEQ ID NO: 438, or SEQ ID NO: 439), or *Pteropus vampyrus* (SEQ ID NO: 434). In embodiments, the mobile element enzymes have one or more hyperactive and/or integration deficient mutations selected from **TABLE 1**, **TABLE 2A**, and/or **TABLE 2B**, or equivalents thereof. One skilled in the art can correspond such mutants to mobile element enzymes

from any of *Trichoplusia ni* (SEQ ID NO: 433), *Myotis lucifugus* (SEQ ID NO: 437), *Myotis myotis* (SEQ ID NO: 435, SEQ ID NO: 436, SEQ ID NO: 438, or SEQ ID NO: 439), or *Pteropus vampyrus* (SEQ ID NO: 434), e.g.:

Trichoplusia ni

1 MGSSLDDEHI LSALLQSDDE LVGEDSDSEI SDHVSDDVQ SDTEEAFIDE VHEVQPTSSG
 61 SEILDEQNV I EQPGSSLASN KILTLPQRTI RGKNKHCWST SKSTRRSRVS ALNIVRSQRG
 121 PTRMCRNIYD PLLCPKLFET DEIISEIVKW TNAEISLKRR ESMTGATFRD TNEDEIYAFF
 181 GILVMTAVRK DNHMSTDDLF DRSLSMVYVS VMSRDRFDFL IRCLRMDDKS IRPTLRENDV
 241 FTFVRKIWDL FIHQCIQNYT PGAHLTIDEQ LLGFRGRCPF RMYIPNKPSK YGIKILMMCD
 301 SGTKYMINGM PYLGRGTQTN GVPLGEYYVK ELSKPVRGSC RNITCDNWFT SIPLAKNLLQ
 361 EPYKLTIVGT VRSNKRRIPE VLKNSRSRPV GTSMFCFDGP LTLVSYKPKP AKMVYLLSSC
 421 DEDASINEST GKPKQVMYYN QTKGGVDITD QMCSVMTCSE KTNRWPMALL YGMINIACIN
 481 SFIIYSHNVS SKGEKVQSRK KFMENLYMSL TSSFMKRLE APTLKRYLRD NISNILENEV
 541 PGTSDDSTEE PVTKKRTYCT YCPSKIRKA NASCKKCKKV ICREHNIDMC QSCF
 (SEQ ID NO: 433)

Pteropus vampyrus

1 MSNPRKRSIP TCDVNFVLEQ LLAEDSFEDES DFSEIDDSDD FSDSASEDYT VRPPSDSESD
 61 GNSPTSADSG RALKWSTRVM IPRQRYDFTG TPGRKVDVSD TTDPLQYFEL FFTEELVSKI
 121 TSEMNAQAAL LASKPPGPKG FSRMDKWKDT DNDELKVFFA VMLLQGIVQK PELEMFWSTR
 181 PLLDIPYLRQ IMTGERFLLL LRCLHFVNNS SISAGQSKAQ ISLQKIKPVF DFLVNKFSTV
 241 YTPNRNIAVD ESLMLFKGRL AMKQYIFTKM NLKDSADGLK (SEQ ID NO: 434)

Myotis myotis ("2a")

1 MDLRCQHTVL SIRESRGLLP NLKMKTSRMK KGDIIFSRKG DILLLAWKDK RVVRMISIH
 61 TSVSTTGKKK RKTGENIVKP ACIKEYNAHM KGVDRADQFL SCCSILRKMM KWTCKVVLVYL
 121 INCGLFNSFR VYNVLNPQAK MKYKQFLLSV ARDWIMDDNN EGSPEPETNL SSPSPGGARR
 181 APRKDPKRL SGDMKQHEPT CIPASGKKKF PTRACRVCAH GKRSESRYLK KFCVLPLHRG
 241 KCFTQYHTLK KY (SEQ ID NO: 435)

Myotis myotis ("1")

1 MKAFLGVILN MGVLNHPNLQ SYWSMDFESH IPFFRSVFKR ERFLQIFWML HLKNDQKSSK
 61 DLRTRETKVN CFLSYLEMKF RERFCPGREI AVDEAVVGFK GKIHFITYNP KKPTKWGIRL
 121 YVLSDSKCGY VHSFVPYYGG ITSETLVPRD LPFTSRIVLE LHERLKNSVP GSQGYHFFTD
 181 RYYTSVTLAK ELFKBKTHLT GTIMPNRKDN PPVIKHQKLE KGEIVAFRDE NVMLLAWKDK
 241 RIVTLSTWDS ETESVERRVG GGKEIVLKP VVTNYTKFMG GVDIADYTST YCFMRKTLKW
 301 WRTLFFWGLE VSVVNSYILY KECQKRENEK PITHVKFIRK LVHDLVGEFR DGTLTSGRL
 361 LSTNLEQRLD GKLHIITPHP NKKHKDCVVC SNRKIKGGRR ETIYICETCE CKPGLHVGE
 421 PKKYHTMKNY RD (SEQ ID NO: 436)

Myotis lucifugus ("2")

1 MPRLRKRKET NETDTLPEVF NDNLS DIPSE IEDADDCFDG SGDDSTDSTD SEIIRPVRRR
 61 KVAVLSSDS TDEATDNCWS EIDTPPRLQM FEGHAGVTTF PSQCDSVPSV TNLFFGDEL
 121 EMLCKELSNY HDQTAMKRKT PSRTLKWSVP TQKDIKKFLG LIILMGQTRK DSLKDYWSTD
 181 PLICTPIFPQ TMSRHRFEQI WTFWHFNDNA KMDSRSGRLF KIQPVLDYFL HKFRTIYKPK
 241 QQLSLDEGMI PWRGRFKFRT YNPAKITKYG LLVRMVCEST TGYICSMEIY TAEGRLQET
 301 VLSVLGPYLG IWHHIYQDNY YNATSTAEEL LQNKTRVCGT IRESRGLPPN LEMKTSRMKK
 361 GDIIFSRKG ILLLAWKDKR VVRMISTIH TSVSTTGKKK RKTGENIVKP TCIKEYNAHM
 421 KGVDRADQFL SCCSILRKTM KWTCKVVLVYL INCGLFNSFR VYNVLNPQAK MKYKQFLLSV

481 ARDWITDDNN EGSPEPETNL SSPSPGGARR APRKDPPKRL SGDMKQHEPT CIPASGKKKF
 541 PTRACRVCAA HGKRSESRYL CKFCLVPLHR GKCFTQYHTL KKYMDLRCQH TVLSTVGRGY
 601 SVLARFKPRT NERTGSSHCH VQVPAGGQGF PSTIIANGCG CKLEPMVRTR SPTCLVIEFG
 661 CM (SEQ ID NO: 437)

Myotis myotis ("2")

1 MPRLRRKKEE NETDTLPEVF NDNLSDIPSE IEDADDCFDD SGDDSTDSTE SEIIRFVRKR
 61 KVAVLSSDSN TDEATDNCWS EIDTPPRLQM FEGHAGVTTF PSQCDSVPSV TNLFFGDLEF
 121 EMLCKELSNY HDQTAMKRKT PSRTLKWSVP TQKDIKKFLG LIILMGQTRK DSWKDYWSTD
 181 PLICTPIFPQ TMSRHRFEQI WTFWHFNDNA KMDSCSGRLF KIQFVLDYFL HKFRTIYKPK
 241 QQLSLDEGMI PWRGRLKFTY NPAITKYGLL VRMVCESDTG YICNMEIYTA ERKKLQETVL
 301 SVLGPYLGIV HHIYQDNYYN ATSTAELLQ NKTRVCGTIR ESRGLPPNLK MKTSMKKKGD
 361 IIFSRKGDIL LLAWKDKRVV RMISTIHDT S VSTTGKKNRK TGENIVKPTC IKEYNAHMKG
 421 VDRADQFLSC CSILRKTTKW TKKVVLYLIN CGLFNSFRVY NILNPQAKMK YKQFLLSVAR
 481 DWITDDNNEG SPEPETNLSS PSSGGARRAP RKDQPKRLSG DMKQHEPTCI PASGKKKFPT
 541 ACRVCAAHGK RSESRYLKRF CFVPLRGKCF MYHTLKKYSE LFSLIVVSKI QNVIIYKTTK
 601 VYMRVYMRSH CFLSFLVFAP SVKDRSRVFS FFTRHLLWTL DVNTLSCPHR MKRSHWWKPC
 661 RSIYEKLYNC TNP (SEQ ID NO: 438)

Myotis myotis ("2b")

1 MDLRCQHTVL SIRESRGLPP NLKMKTSRMK KGDIIFSRKG DILLAWKDK RVVRMISTIH
 61 DTSVSTTGKK NRKTGENIVK PACIKEYNAH MKGVDRADQF LSCCSILRKT MKWTKKVVLY
 121 LINCGLFNSF RVYNVLNPQA KMKYQFLLS VARDWITDDN NEGSPEPETN LSSPSGGAR
 181 RAPRKDPPKR LSGDMKQHEP TCIPASGKKK FPTRACRVCA AHGKRSESRV LCKFCLVPLH
 241 RGKCFQYHT LKKY (SEQ ID NO: 439)

In embodiments, the mobile element enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, or *Trichoplusia ni*. In embodiments, the mobile element enzyme is an engineered version of a mobile element enzyme, including but not limited to monomers, dimers, tetramers, hyperactive, or Int-forms, derived from *Bombyx mori*, *Xenopus tropicalis*, or *Trichoplusia ni*.

In embodiments, the mobile element enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, or *Myotis lucifugus*. In embodiments, the mobile element enzyme is an engineered version, including but not limited to a mobile element enzyme that is a monomer, dimer, tetramer (or another multimer), hyperactive, or has a reduced interaction with non-TTAA (SEQ ID NO: 440) recognitions sites (Int-), derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni* or *Myotis lucifugus*. In embodiments, the mobile element enzymes have one or more hyperactive and/or integration deficient mutations selected from **TABLE 1**, **TABLE 2A**, and **TABLE 2B**, or equivalents thereof.

In embodiments, one skilled in the art can correspond such mutants to mobile element enzymes from any of *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Myotis lucifugus*, *Pipistrellus kuhlii*, *Pteropus vampyrus*, *Pan troglodytes*, and *Molossus molossus*.

In embodiments, the mobile element enzyme has a nucleotide sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to a nucleotide sequence of any of *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Myotis lucifugus*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Pan troglodytes*, and *Molossus molossus*. In embodiments, the mobile element enzyme has an amino acid sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to an amino acid sequence of any of *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Myotis lucifugus*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, and *Molossus molossus*. See Jebb, et al. (2020).

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme) is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*. In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme) is an engineered version, including but not limited to hyperactive forms, of a mobile element enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*. The enzyme is either the wild type, monomer, dimer, tetramer, hyperactive, or an Int-mutant. In embodiments, the mobile element enzymes have one or more hyperactive and/or integration deficient mutations selected from **TABLE 1**, **TABLE 2A**, and/or **TABLE 2B**, or equivalents thereof.

In embodiments, the mobile element enzyme has a nucleotide sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to a nucleotide sequence of any of *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, and *Pan troglodytes*. In embodiments, the mobile element enzyme has an amino acid sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to an amino acid sequence of any of *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, and *Homo sapiens*.

In embodiments, the mobile element enzyme is an engineered version, including but not limited to a mobile element enzyme that is a monomer, dimer, tetramer, hyperactive, or has a reduced interaction with non-TTAA (SEQ ID NO: 440) recognitions sites (Int-), derived from any of *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Myotis lucifugus*, *Pipistrellus kuhlii*, *Pteropus vampyrus*, and *Molossus molossus* *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Pan troglodytes*, *Myotis*

lucifugus, and *Homo sapiens*. The mobile element enzyme is either the wild type, monomer, dimer, tetramer or another multimer, hyperactive, or a an Int-mutant.

In embodiments, the mobile element enzyme is from a *Tc1*/mariner donor DNA system. See, e.g., Plasterk *et al.* *Trends in Genetics*. 1999; 15(8):326–32.

In embodiments, the mobile element enzyme is from a Sleeping Beauty donor DNA system (see, e.g., *Cell*. 1997;91:501–510), e.g., a hyperactive form of Sleeping Beauty (hypSB), e.g., SB100X (see *Gene Therapy* volume 18, pages 849–856(2011), or a piggyBac (PB) donor DNA system (see, e.g., *Trends Biotechnol.* 2015 Sep;33(9):525-33, which is incorporated herein by reference in its entirety), e.g., a hyperactive form of PB mobile element enzyme (hypPB), e.g., with seven amino acid substitutions (e.g., I30V, S103P, G165S, M282V, S509G, N570S, N538K on mPB, or functional equivalents in non-mPB, see *Mol Ther Nucleic Acids*. 2012 Oct; 1(10): e50, which is incorporated herein by reference in its entirety); see also Yusa *et al.*, *PNAS* January 25, 2011 108 (4) 1531-1536; Voigt *et al.*, *Nature Communications* volume 7, Article number: 11126 (2016).

The piggyBac mobile element enzymes belong to the IS4 mobile element enzyme family. De Palmenaer *et al.*, *BMC Evolutionary Biology*. 2008;8:18. doi: 10.1186/1471-2148-8-18. The piggyBac family includes a large diversity of donor DNAs, and any of these donor DNAs can be used in embodiments of the present disclosure. See, e.g., Bouallègue *et al.*, *Genome Biol Evol.* 2017;9(2):323-339. The founding member of the piggyBac (super)family, insect piggyBac, was originally identified in the cabbage looper moth (*Trichoplusia ni*) and studied both *in vivo* and *in vitro*. Insect piggyBac is known to transpose by a canonical cut-and-paste mechanism promoted by an element-encoded mobile element enzyme with a catalytic site resembling the RNase H fold shared by many recombinases. The insect piggyBac donor DNA system has been shown to be highly active in a wide range of animals, including *Drosophila* and mice, where it has been developed as a powerful tool for gene tagging and genome engineering. Other donor DNAs affiliated to the piggyBac superfamily are common in arthropods and vertebrates including *Xenopus* and *Bombyx*. Mammalian piggyBac donor DNAs and mobile element enzymes, including hyperactive mammalian piggyBac variants, which can be used in embodiments of the present disclosure, are described, e.g., in International Application WO2010085699, which is incorporated herein by reference in its entirety.

In embodiments, the mobile element enzyme is from a LEAP-IN 1 type or LEAP-IN donor DNA system (*Biotechnol J*. 2018 Oct;13(10):e1700748. doi: 10.1002/biot.201700748. Epub 2018 Jun 11). The LEAPIN mobile element enzyme system includes a mobile element enzyme (e.g., without limitation, a mobile element enzyme mRNA) and a vector containing one or more genes of interest (donor DNAs), selection markers, regulatory elements, insulators, etc., flanked by the donor DNA cognate inverted terminal ends and the transposition recognition motif (TTAT). Upon co-transfection of vector DNA and mobile element enzyme mRNA, the transiently expressed enzyme catalyzes high-efficiency and precise integration of a single copy of the donor DNA cassette (all sequences between the terminal ends) at one or

more sites across the genome of the host cell. Hottentot *et al.* In *Genotyping: Methods and Protocols*. White SJ, Cantsilieris S, eds: 185–196. (New York, NY: Springer): 2017. pp. 185–196. The LEAPIN mobile element enzyme generates stable transgene integrants with various advantageous characteristics, including single copy integrations at multiple genomic loci, primarily in open chromatin segments; no payload limit, so multiple independent transcriptional units may be expressed from a single construct; the integrated transgenes maintain their structural and functional integrity; and maintenance of transgene integrity ensures the desired chain ratio in every recombinant cell.

In embodiments, the mobile element enzyme is an engineered form of a mobile element enzyme reconstructed from *Homo sapiens* or a predecessor thereof.

Donor DNAs in Humans have 5 inactive elements, designated PiggyBac domain (PGBD)1, PGBD2, PGBD3, PGBD4, and PGBD5. PGBD1, PGBD2, and PGBD3 have multiple coding exons, but in each case the mobile element enzyme-related sequence is encoded by a single uninterrupted 3' terminal exon. Thus, PGBD1 and PGBD2 may resemble the PGBD3 donor DNA in which the mobile element enzyme ORF is flanked upstream by a 3' splice site and downstream by a polyadenylation site. See Newman *et al.*, *PLoS Genet* 2008;4:e1000031. *PLoS Genet* 4(3): e1000031. <https://doi.org/10.1371/journal.pgen.1000031>; Gray *et al.*, *PLoS Genet* 8(9): e1002972. <https://doi.org/10.1371/journal.pgen.1002972>.

The PGBD5 inactive mobile element enzyme sequence belongs to the RNase H clan of Pfam structures, while PGBD3 has sustained only a single D to N mutation in the essential catalytic triad DDD(D) and retains the ability to bind the upstream piggyBac terminal inverted repeat. Bailey *et al.*, *DNA Repair* (Amst) 2012;11:488-501. The PGBD5 mobile element enzyme does not retain the catalytic DDD (D) motif found in active elements, and the mobile element enzyme is not only inactive but fails to associate with either DNA or chromatin *in vivo*. Pavelitz *et al.*, *Mob DNA* 2013;4:23. However, *in vitro* studies showed that it is transpositionally active in HEK293 cells. See Henssen *et al.*, *Elife* 2015;4. PGBD1 and PGBD2 are thought to be present in the common ancestor of mammals, while PGBD3 and PGBD4 are restricted to primates. See Sarkar *et al.*, *Mol Genet Genomics* 2003;270:173-80. The *Pteropus vampyrus* mobile element enzyme is closely related to PGBD4 and shares DDD catalytic domain and the C-terminal region that are involved in excision mechanisms. See Mitra *et al.*, *EMBO J* 2008;27:1097-109.

A mammalian mobile element enzyme, which has gene cleavage and/or gene integration activity, can be constructed based on alignment of the amino acid sequence of *Pteropus vampyrus* mobile element enzyme to PGBD1, PGBD2, PGBD3, PGBD4, and PGBD5 sequences. Also, in embodiments, the mammalian mobile element enzyme has mutations that confers hyperactivity to a recombinant mammalian mobile element enzyme. Accordingly, in embodiments, the mobile element enzyme has gene cleavage activity (Exc+) and/or gene integration activity (Int+). In embodiments, the mobile element enzyme has gene cleavage activity (Exc+) and/or lacks gene integration activity (Int-).

In some aspects, an enzyme capable of targeted genomic integration is a recombinant mammalian mobile element enzyme that was derived by, in part, aligning several inactive mobile element enzyme sequences from a human genome to *Pteropus vampyrus* mobile element enzyme sequence. In embodiments, the *Pteropus vampyrus* mobile element enzyme has an amino acid sequence having at least 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to SEQ ID NO: 430 (or a functional equivalent thereof). In embodiments, the *Pteropus vampyrus* mobile element enzyme has an amino acid sequence of SEQ ID NO: 430, or a functional equivalent thereof. In embodiments, the *Pteropus vampyrus* mobile element enzyme has a nucleotide sequence having at least 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to SEQ ID NO: 429 or a codon-optimized variant thereof.

In embodiments, the mobile element enzyme is a mammalian mobile element enzyme, such as a mobile element enzyme from a bat, e.g., without limitation, *Pteropus vampyrus*.

In embodiments, the mobile element enzyme is an engineered form that is based on a mobile element enzyme reconstructed from *Homo sapiens* or a predecessor thereof. In embodiments, the mobile element enzyme includes but is not limited to an engineered version that is a monomer, dimer, tetramer (or another multimer), hyperactive, or has a reduced interaction with non-TTAA (SEQ ID NO: 440) recognitions sites (Int-), of an engineered version of a mobile element enzyme reconstructed from *Homo sapiens* or a predecessor thereof.

In embodiments, the mobile element enzyme is an engineered form that is based on a mobile element enzyme reconstructed from mammalian species. In embodiments, the mobile element enzyme includes but is not limited to an engineered that is a monomer, dimer, tetramer (or another multimer), hyperactive, or has a reduced interaction with non-TTAA (SEQ ID NO: 440) recognitions sites (Int-), of a mobile element enzyme reconstructed from mammalian species.

In embodiments, the donor DNA is included in a vector comprising left and right end sequences recognized by the mobile element enzyme.

In embodiments, the end sequences are selected from MER, MER75A, MER75B, and MER85.

In embodiments, the end sequences are selected from nucleotide sequences of SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 441, and SEQ ID NO: 22, or a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) thereto. In embodiments, one or more of the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) identity to the nucleotide sequence of SEQ ID NO: 12, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 12 is positioned at the 5' end of the donor DNA. The end sequences can further include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 17, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 17 is positioned at the 3' end of the donor DNA. The end sequences, which can be from, e.g., *Pteropus vampyrus*, are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 13, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 13 is positioned at the 5' end of the donor DNA. The end sequences can further include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 18, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 18 is positioned at the 3' end of the donor DNA. The end sequences, which can be, e.g., PGBD4, are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 14, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 14 is positioned at the 5' end of the donor DNA. The end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence

having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 18, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 19 is positioned at the 3' end of the donor DNA. The end sequences, which can be, e.g., MER75, are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 15, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 15 is positioned at the 5' end of the donor DNA. The end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 20, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 20 is positioned at the 3' end of the donor DNA. The end sequences, which can be, e.g., MER75B, are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, the end sequences include at least one repeat from a nucleotide sequence having at least about 90% (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) identity to the nucleotide sequence of SEQ ID NO: 16, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 16 is positioned at the 5' end of the donor DNA. The end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity (e.g., a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441 is positioned at the 3' end of the donor DNA. The end sequences, which can be, e.g., MER75A, are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

In embodiments, a donor DNA is or comprises a vector comprising a donor DNA comprising one or more end sequences recognized by an enzyme such as, for example a mobile element enzyme. In embodiments, the end sequences are selected from *Pteropus vampyrus*, MER75, MER75A, and MER75B. MERs contain end sequences with similarity to piggyBac-like mobile elements and exhibit duplications of their presumed TTAA (SEQ ID NO: 440) target sites. In embodiments, the end sequences are selected from nucleotide sequences of SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 441, and SEQ ID NO: 22, or a nucleotide sequence having at least about 90% identity (e.g. a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity) thereto.

In embodiments, the mobile element enzyme has an amino acid sequence of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, or SEQ ID NO: 9, or a variant sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the mobile element enzyme has an amino acid sequence having S8P, G17R, and/or K134K mutation relative to the amino acid sequence of SEQ ID NO: 4 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has an amino acid sequence having S8P, G17R, and/or K134K mutation relative to the amino acid sequence of SEQ ID NO: 5 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has an amino acid sequence having I83P and/or V118R mutation relative to the amino acid sequence of SEQ ID NO: 6 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has an amino acid sequence having S20P and/or A29R mutation relative to the amino acid sequence of SEQ ID NO: 7 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has an amino acid sequence having T4P and/or L13R mutation relative to the amino acid sequence of SEQ ID NO: 8 or a functional equivalent thereof.

In embodiments, the mobile element enzyme has an amino acid sequence having A12P and/or I28R mutation and/or R152K mutation relative to the amino acid sequence of SEQ ID NO: 9 or a functional equivalent thereof.

In embodiments, the enzyme capable of targeted genomic integration (e.g., without limitations, a mobile element enzyme) is in a monomeric or dimeric form. In embodiments, the enzyme capable of targeted genomic integration (e.g., without limitations, a mobile element enzyme) is in a multimeric form.

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme) is an engineered version, including but not limited to a mobile element enzyme that is a monomer, dimer, tetramer, hyperactive, or has a reduced interaction with non-TTAA (SEQ ID NO: 440) recognition sites (Int-), and is derived from any of *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Pan troglodytes*, *Molossus molossus*, or *Homo sapiens*. In embodiments, the mobile element enzyme is either the wild type, monomer, dimer, tetramer or another multimer, hyperactive, or an Int-mutant.

Targeting Chimeric Constructs

In aspects, the present disclosure provides for targeted chimeras, e.g., in embodiments, the enzyme, without limitation, a mobile element enzyme, comprises a targeting element.

In embodiments, the targeting element is or comprises a gene-editing system, e.g., that is catalytically inactive (or "dead").

In embodiments, the enzyme, without limitation, a mobile element enzyme, associated with the targeting element, is capable of inserting the donor DNA (e.g., mobile element) comprising a transgene, optionally at a TA dinucleotide site or a TTAA (SEQ ID NO: 440) tetranucleotide site in a GSHS.

In embodiments, the enzyme, without limitation, a mobile element enzyme, associated with the targeting element has one or more mutations which confer hyperactivity.

In embodiments, the enzyme, without limitation, a mobile element enzyme, associated with the targeting element has gene cleavage activity (Exc+) and/or gene integration activity (Int+).

In embodiments, the enzyme, without limitation, a mobile element enzyme, associated with the targeting element has gene cleavage activity (Exc+) and/or a lack of gene integration activity (Int-).

In embodiments, the targeting element comprises one or more proteins or nucleic acids that are capable of binding to a nucleic acid.

In embodiments, the targeting element comprises one or more of a of a gRNA, optionally associated with a Cas enzyme, which is optionally catalytically inactive, transcription activator-like effector (TALE), catalytically inactive Zinc finger, catalytically inactive transcription factor, nickase, a transcriptional activator, a transcriptional repressor, a recombinase, a DNA methyltransferase, a histone methyltransferase, paternally expressed gene 10 (PEG10), and TnsD.

In embodiments, the targeting element comprises a transcription activator-like effector (TALE) DNA binding domain (DBD).

In embodiments, the TALE DBD comprises one or more repeat sequences. In embodiments, the TALE DBD comprises about 14, or about 15, or about 16, or about 17, or about 18, or about 18.5 repeat sequences. In embodiments, the TALE DBD repeat sequences comprise 33 or 34 amino acids. In embodiments, the TALE DBD repeat sequences comprise a repeat variable di-residue (RVD) at residue 12 or 13 of the 33 or 34 amino acids. In embodiments, the RVD recognizes one base pair in the nucleic acid molecule. In embodiments, the RVD recognizes a C residue in the nucleic acid molecule and is selected from HD, N(gap), HA, ND, and HI. In embodiments, the RVD recognizes a G residue in the nucleic acid molecule and is selected from NN, NH, NK, HN, and NA. In embodiments, the RVD recognizes an A residue in the nucleic acid molecule and is selected from NI and NS. In embodiments, the RVD recognizes a T residue in the nucleic acid molecule and is selected from NG, HG, H(gap), and IG. In embodiments, the GSHS is in an open chromatin location in a chromosome. In embodiments, the GSHS is selected from adeno-associated virus site 1 (AAVS1), chemokine (C-C motif) receptor 5 (CCR5) gene, HIV-1 coreceptor, and human Rosa26 locus. In embodiments, the GSHS is located on human chromosome 2, 4, 6, 10, 11, 17, 22, or X. In embodiments, the GSHS is selected from TALC1, TALC2, TALC3, TALC4, TALC5, TALC7, TALC8, AVS1, AVS2, AVS3, ROSA1, ROSA2, TALER1, TALER2, TALER3, TALER4, TALER5, SHCHR2-1, SHCHR2-2, SHCHR2-3, SHCHR2-4, SHCHR4-1, SHCHR4-2, SHCHR4-3, SHCHR6-1, SHCHR6-2, SHCHR6-3, SHCHR6-4, SHCHR10-1, SHCHR10-2, SHCHR10-3, SHCHR10-4, SHCHR10-5, SHCHR11-1, SHCHR11-2, SHCHR11-3, SHCHR17-1, SHCHR17-2, SHCHR17-3, and SHCHR17-4.

In embodiments, the targeting element comprises a Cas9 enzyme guide RNA complex. In embodiments, the Cas9 enzyme guide RNA complex comprises a nuclease-deficient dCas9 guide RNA complex. In embodiments, the targeting element comprises a Cas12 enzyme guide RNA complex. In embodiments, the targeting element comprises a nuclease-deficient dCas12 guide RNA complex, optionally dCas12j guide RNA complex or dCas12a guide RNA complex. In embodiments, the targeting element comprises a Cas12k enzyme guide RNA complex. In embodiments, the targeting element comprises a nuclease-deficient dCas12 guide RNA complex, optionally dCas12k guide RNA complex.

In embodiments, a targeting chimeric system or construct, having a DBD fused to a mobile element enzyme, directs binding of an enzyme capable of targeted genomic integration (e.g., without limitation, a mobile element enzyme) to a specific sequence (e.g., transcription activator-like effector proteins (TALE) repeat variable di-residues (RVD) or gRNA) near an enzyme recognition site. The enzyme is thus prevented from binding to random recognition sites. In embodiments, the targeting chimeric construct binds to human GSHS. In embodiments, dCas9 (*i.e.*, deficient for nuclease activity) is programmed with gRNAs directed to bind at a desired sequence of DNA in GSHS.

In embodiments, TALEs described herein can physically sequester the enzyme such as, e.g., a mobile element enzyme, to GSHS and promote transposition to nearby TTAA (SEQ ID NO: 440) sequences in close proximity to the RVD TALE

nucleotide sequences. GSHs in open chromatin sites are specifically targeted based on the predilection for mobile element enzymes to insert into open chromatin.

In embodiments, an enzyme capable of targeted genomic integration (e.g., without limitation, a recombinase, integrase, or a mobile element enzyme such as, without limitation, a mammalian mobile element enzyme) is linked to or fused with a TALE DNA binding domain (DBD) or a Cas-based gene-editing system, such as, e.g., Cas9 or a variant thereof.

In embodiments, the targeting element targets the enzyme to a locus of interest. In embodiments, the targeting element comprises CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) associated protein 9 (Cas9), or a variant thereof. A CRISPR/Cas9 tool only requires Cas9 nuclease for DNA cleavage and a single-guide RNA (sgRNA) for target specificity. See Jinek *et al.* (2012) *Science* 337, 816–821; Chylinski *et al.* (2014) *Nucleic Acids Res* 42, 6091–6105. The inactivated form of Cas9, which is a nuclease-deficient (or inactive, or "catalytically dead" Cas9, is typically denoted as "dCas9," has no substantial nuclease activity. Qi, L. S. *et al.* (2013). *Cell* 152, 1173–1183. CRISPR/dCas9 binds precisely to specific genomic sequences through targeting of guide RNA (gRNA) sequences. See Dominguez *et al.*, *Nat Rev Mol Cell Biol.* 2016;17:5–15; Wang *et al.*, *Annu Rev Biochem.* 2016;85:227–64. dCas9 is utilized to edit gene expression when applied to the transcription binding site of a desired site and/or locus in a genome. When the dCas9 protein is coupled to guide RNA (gRNA) to create dCas9 guide RNA complex, dCas9 prevents the proliferation of repeating codons and DNA sequences that might be harmful to an organism's genome. Essentially, when multiple repeat codons are produced, it elicits a response, or recruits an abundance of dCas9 to combat the overproduction of those codons and results in the shut-down of transcription. Thus, dCas9 works synergistically with gRNA and directly affects the DNA polymerase II from continuing transcription.

In embodiments, the targeting element comprises a nuclease-deficient Cas enzyme guide RNA complex. In embodiments, the targeting element comprises a nuclease-deficient (or inactive, or "catalytically dead" Cas, e.g., Cas9, typically denoted as "dCas" or "dCas9") guide RNA complex.

In embodiments, the dCas9/gRNA complex comprises a guide RNA selected from: GTTAGCTCACCCGTGAGCC (SEQ ID NO: 91), CCCAATATTATTGTTCTCTG (SEQ ID NO: 92), GGGGTGGGATAGGGGATACG (SEQ ID NO: 93), GGATCCCCCTCTACATTTAA (SEQ ID NO: 94), GTGATCTTGTACAAATCATT (SEQ ID NO: 95), CTACACAGAATCTGTTAGAA (SEQ ID NO: 96), TAAGCTAGAGAATAGATCTC (SEQ ID NO: 97), and TCAATACACTTAATGATTTA (SEQ ID NO: 98), wherein the guide RNA directs the enzyme to a chemokine (C-C motif) receptor 5 (CCR5) gene.

In embodiments, the dCas9/gRNA complex comprises a guide RNA selected from:

CACCGGGAGCCACGAAAACAGATCC (SEQ ID NO: 99); CACCGCGAAAACAGATCCAGGGACA (SEQ ID NO: 100); CACCGAGATCCAGGGACACGGTGCT (SEQ ID NO: 101); CACCGGACACGGTGCTAGGACAGTG (SEQ ID NO: 102);

102); CACCGGAAAATGACCCAAACAGCCTC (SEQ ID NO: 103); CACCGGCCTGGCCGGCCTGACCACT (SEQ ID NO: 104); CACCGCTGAGCACTGAAGGCCTGGC (SEQ ID NO: 105); CACCGTGGTTTCCACTGAGCACTGA (SEQ ID NO: 106); CACCGGATAGCCAGGAGTCCTTTCG (SEQ ID NO: 107); CACCGGCGCTTCCAGTGCTCAGACT (SEQ ID NO: 108); CACCGCAGTGCTCAGACTAGGGAAG (SEQ ID NO: 109); CACCGGCCCCCTCCTCCTTCAGAGCC (SEQ ID NO: 110); CACCGTCCTTCAGAGCCAGGAGTCC (SEQ ID NO: 111); CACCGTGGTTTCCGAGCTTGACCCT (SEQ ID NO: 112); CACCGCTGCAGAGTATCTGCTGGGG (SEQ ID NO: 113); CACCGCGTTCCTGCAGAGTATCTGC (SEQ ID NO: 114); TCCCCTCCCAGAAAGACCTG (SEQ ID NO: 131); TGGGCTCCAAGCAATCCTGG (SEQ ID NO: 132); GTGGCTCAGGAGGTACCTGG (SEQ ID NO: 133); GAGCCACGAAAACAGATCCA (SEQ ID NO: 134); AAGTGAACGGGGAAGGGAGG (SEQ ID NO: 135); GACAAAAGCCGAAGTCCAGG (SEQ ID NO: 136); GTGGTTGATAAACCACGTG (SEQ ID NO: 137); TGGAACAGCCACAGCAGGG (SEQ ID NO: 138); GCAGGGGAACGGGGATGCAG (SEQ ID NO: 139); GAGATGGTGGACGAGGAAGG (SEQ ID NO: 140); GAGATGGCTCCAGGAAATGG (SEQ ID NO: 141); TAAGGAATCTGCCTAACAGG (SEQ ID NO: 142); TCAGGAGACTAGGAAGGAGG (SEQ ID NO: 143); TATAAGGTGGTCCCAGCTCG (SEQ ID NO: 144); CTGGAAGATGCCATGACAGG (SEQ ID NO: 145); GCACAGACTAGAGAGGTAAG (SEQ ID NO: 146); ACAGACTAGAGAGGTAAGGG (SEQ ID NO: 147); GAGAGGTGACCCGAATCCAC (SEQ ID NO: 148); GCACAGGCCCCAGAAAGGAGA (SEQ ID NO: 149); CCGGAGAGGACCCAGACACG (SEQ ID NO: 150); GAGAGGACCCAGACACGGGG (SEQ ID NO: 151); GCAACACAGCAGAGAGCAAG (SEQ ID NO: 152); GAAGAGGGAGTGGAGGAAGA (SEQ ID NO: 153); AAGACGGAACCTGAAGGAGG (SEQ ID NO: 154); AGAAAGCGGCACAGGCCAG (SEQ ID NO: 155); GGGAAACAGTGGGCCAGAGG (SEQ ID NO: 156); GTCCGGACTCAGGAGAGAGA (SEQ ID NO: 157); GGCACAGCAAGGGCACTCGG (SEQ ID NO: 158); GAAGAGGGGAAGTCGAGGGA (SEQ ID NO: 159); GGGAAATGGTAAGGAGGCCTG (SEQ ID NO: 160); GCAGAGTGGTCAGCACAGAG (SEQ ID NO: 161); GCACAGAGTGGCTAAGCCCA (SEQ ID NO: 162); GACGGGGTGTCAGCATAGGG (SEQ ID NO: 163); GCCCAGGGCCAGGAACGACG (SEQ ID NO: 164); GGTGGAGTCCAGCACGGCGC (SEQ ID NO: 165); ACAGGCCGCCAGGAACCTCGG (SEQ ID NO: 166); ACTAGGAAGTGTGTAGCACC (SEQ ID NO: 167); ATGAATAGCAGACTGCCCGG (SEQ ID NO: 168); ACACCCCTAAAAGCACAGTG (SEQ ID NO: 169); CAAGGAGTTCCAGCAGGTGG (SEQ ID NO: 170); AAGGAGTTCCAGCAGGTGGG (SEQ ID NO: 171); TGGAAGAGGAGGGAAGAGG (SEQ ID NO: 172); TCGAATTCCTAACTGCCCCG (SEQ ID NO: 173); GACCTGCCCAGCACACCCTG (SEQ ID NO: 174); GGAGCAGCTGCGGCAGTGGG (SEQ ID NO: 175); GGGAGGGAGAGCTTGGCAGG (SEQ ID NO: 176); GTTACGTGGCCAAGAAGCAG (SEQ ID NO: 177); GCTGAACAGAGAAGAGCTGG (SEQ ID NO: 178); TCTGAGGGTGGAGGGACTGG (SEQ ID NO: 179); GGAGAGGTGAGGGACTTGGG (SEQ ID NO: 180); GTGAACCAGGCAGACAACGA (SEQ ID NO: 181); CAGGTACCTCCTGAGCCACG (SEQ ID NO: 182); GGGGGAGTAGGGGCATGCAG (SEQ ID NO: 183);

GCAAATGGCCAGCAAGGGTG (SEQ ID NO: 184); CAAATGGCCAGCAAGGGTGG (SEQ ID NO: 309);
GCAGAACCTGAGGATATGGA (SEQ ID NO: 310); AATACACAGAATGAAAATAG (SEQ ID NO: 311);
CTGGTGA CTAGAATAGCCAG (SEQ ID NO: 312); TGGTGA CTAGAATAGCCAGT (SEQ ID NO: 313);
TAAAAGAATGTGAAAAGATG (SEQ ID NO: 314); TCAGGAGTTCAAGACCACCC (SEQ ID NO: 315);
TG TAGTCCCAGTTATGCAGG (SEQ ID NO: 316); GGGTTCACACCACAAATGCA (SEQ ID NO: 317);
GGCAAATGGCCAGCAAGGGT (SEQ ID NO: 318); AGAAACCAATCCCAAAGCAA (SEQ ID NO: 319);
GCCAAGGACACCAAAACCCA (SEQ ID NO: 320); AGTGGTGATAAGGCCAACAGT (SEQ ID NO: 321);
CCTGAGACAGAAGTATTAAG (SEQ ID NO: 322); AAGGTCACACAATGAATAGG (SEQ ID NO: 323);
CACCATACTAGGGAAGAAGA (SEQ ID NO: 324); CAATACCCTGCCCTTAGTGG (SEQ ID NO: 327);
AATACCCTGCCCTTAGTGGG (SEQ ID NO: 325); TTAGTGGGGGGTGGAGTGGG (SEQ ID NO: 326);
GTGGGGGGTGGAGTGGGGGG (SEQ ID NO: 328); GGGGGGTGGAGTGGGGGGTG (SEQ ID NO: 329);
GGGGTGGAGTGGGGGGTGGG (SEQ ID NO: 330); GGGTGGAGTGGGGGGTGGG (SEQ ID NO: 331);
GGGGGTGGGGAAAGACATCG (SEQ ID NO: 332); GCAGCTGTGAATTCTGATAG (SEQ ID NO: 333);
GAGATCAGAGAAACCAGATG (SEQ ID NO: 334); TCTATACTGATTGCAGCCAG (SEQ ID NO: 335);
CACCGAATCGAGAAGCGACTCGACA (SEQ ID NO: 185); CACCGGTCCCTGGGCGTTGCCCTGC (SEQ ID NO:
186); CACCGCCCTGGGCGTTGCCCTGCAG (SEQ ID NO: 187); CACCGCCGTGGGAAGATAAACTAAT (SEQ ID
NO: 188); CACCGTCCCTGCAGGGCAACGCC (SEQ ID NO: 189); CACCGGTGAGTCGCTTCTCGATTA (SEQ
ID NO: 190); CACCGCTGCTGCCTCCCGTCTTGTA (SEQ ID NO: 191); CACCGGAGTCCGCAATACCTTTAT (SEQ
ID NO: 192); CACCGACACTTTGGTGGTGCAGCAA (SEQ ID NO: 193); CACCGTCTCAAATGGTATAAACTC (SEQ
ID NO: 194); CACCGAATCCCGCCCATATCGAGA (SEQ ID NO: 195); CACCGTCCCGCCCATATCGAGAAG
(SEQ ID NO: 196); CACCGCCCATATCGAGAAGCGACT (SEQ ID NO: 197);
CACCGGAGAAGCGACTCGACATGGA (SEQ ID NO: 198); CACCGGAAGCGACTCGACATGGAGG (SEQ ID NO:
199); CACCGGCGACTCGACATGGAGGCGA (SEQ ID NO: 200); AAAGTGTGAGTCGCTTCTCGATTG (SEQ ID
NO: 201); AAACGCAGGGCAACGCCAGGGACC (SEQ ID NO: 202); AAACCTGCAGGGCAACGCCAGGGC (SEQ
ID NO: 203); AAACATTAGTTTATCTTCCACGGC (SEQ ID NO: 204); AAACGGCGTTGCCCTGCAGGGGAC (SEQ
ID NO: 205); AAATAATCGAGAAGCGACTCGACC (SEQ ID NO: 206); AAATAACAAGACGGGAGGCAGCAGC
(SEQ ID NO: 207); AAACATAAAGGTATTGCGGCACTCC (SEQ ID NO: 208); AAATTGCTGCACCACCAAAGTGTC
(SEQ ID NO: 209); AAACGAGTTTATACCATTTGAGAC (SEQ ID NO: 210); AAATCTCGATTATGGGCGGGATTG
(SEQ ID NO: 211); AAACCTTCTCGATTATGGGCGGGAC (SEQ ID NO: 212); AAACAGTCGCTTCTCGATTATGGGC
(SEQ ID NO: 213); AAATCCATGTGAGTCGCTTCTCC (SEQ ID NO: 214); AAACCTCCATGTGAGTCGCTTCC
(SEQ ID NO: 215); AAATCGCCTCCATGTGAGTCGCC (SEQ ID NO: 216); CACCGACAGGGTTAATGTGAAGTCC
(SEQ ID NO: 217); CACCGTCCCCCTCTACATTTAAAGT (SEQ ID NO: 218); CACCGCATTTAAAGTTGGTTAAAGT
(SEQ ID NO: 219); CACCGTTAGAAAATATAAAGAATAA (SEQ ID NO: 220); CACCGTAAATGCTTACTGGTTTGAA

(SEQ ID NO: 221); CACCGTCCTGGGTCCAGAAAAAGAT (SEQ ID NO: 222); CACCGTTGGGTGGTGAGCATCTGTG (SEQ ID NO: 223); CACCGCGGGGAGAGTGGAGAAAAAG (SEQ ID NO: 224); CACCGGTTAAACTCTTTAGACAAC (SEQ ID NO: 225); CACCGGAAAATCCCCACTAAGATCC (SEQ ID NO: 226); AAACGGACTTCACATTAACCCTGTC (SEQ ID NO: 227); AAACACTTTAAATGTAGAGGGGGAC (SEQ ID NO: 228); AAACACTTAAACCAACTTTAAATGC (SEQ ID NO: 229); AAACCTTATTCTTTATATTTTCTAAC (SEQ ID NO: 230); AAACCTCAAACCAGTAAGCATTTAC (SEQ ID NO: 231); AAACATCTTTTTCTGGACCCAGGAC (SEQ ID NO: 232); AAACCACAGATGCTCACCACCCAAC (SEQ ID NO: 233); AAACCTTTTTCTCCACTCTCCCCGC (SEQ ID NO: 234); AAACGTTGTCTAAAGAGTTTAAACC (SEQ ID NO: 235); AAACGGATCTTAGTGGGGATTTTCC (SEQ ID NO: 236); AGTAGCAGTAATGAAGCTGG (SEQ ID NO: 237); ATACCCAGACGAGAAAGCTG (SEQ ID NO: 238); TACCCAGACGAGAAAGCTGA (SEQ ID NO: 239); GGTGGTGAGCATCTGTGTGG (SEQ ID NO: 240); AAATGAGAAGAAGAGGCACA (SEQ ID NO: 241); CTTGTGGCCTGGGAGAGCTG (SEQ ID NO: 242); GCTGTAGAAGGAGACAGAGC (SEQ ID NO: 243); GAGCTGGTTGGGAAGACATG (SEQ ID NO: 244); CTGGTTGGGAAGACATGGGG (SEQ ID NO: 245); CGTGAGGATGGGAAGGAGGG (SEQ ID NO: 246); ATGCAGAGTCAGCAGAACTG (SEQ ID NO: 247); AAGACATCAAGCACAGAAGG (SEQ ID NO: 248); TCAAGCACAGAAGGAGGAGG (SEQ ID NO: 249); AACCGTCAATAGGCAAAGGG (SEQ ID NO: 250); CCGTATTTCACTGAATGG (SEQ ID NO: 251); GAGAGGACAGGTGCTACAGG (SEQ ID NO: 252); AACCAAGGAAGGGCAGGAGG (SEQ ID NO: 253); GACCTCTGGGTGGAGACAGA (SEQ ID NO: 254); CAGATGACCATGACAAGCAG (SEQ ID NO: 255); AACACCAGTGAGTAGAGCGG (SEQ ID NO: 256); AGGACCTTGAAGCACAGAGA (SEQ ID NO: 257); TACAGAGGCAGACTAACCCA (SEQ ID NO: 258); ACAGAGGCAGACTAACCCAG (SEQ ID NO: 259); TAAATGACGTGCTAGACCTG (SEQ ID NO: 260); AGTAACCACTCAGGACAGGG (SEQ ID NO: 261); ACCACAAAACAGAAACACCA (SEQ ID NO: 262); GTTTGAAGACAAGCCTGAGG (SEQ ID NO: 263); GCTGAACCCCAAAAGACAGG (SEQ ID NO: 264); GCAGCTGAGACACACACCAG (SEQ ID NO: 265); AGGACACCCCAAAGAAGCTG (SEQ ID NO: 266); GGACACCCCAAAGAAGCTGA (SEQ ID NO: 267); CCAGTGCAATGGACAGAAGA (SEQ ID NO: 268); AGAAGAGGGAGCCTGCAAGT (SEQ ID NO: 269); GTGTTTGGGCCCTAGAGCGA (SEQ ID NO: 270); CATGTGCCTGGTGCAATGCA (SEQ ID NO: 271); TACAAAGAGGAAGATAAGTG (SEQ ID NO: 272); GTCACAGAATACACCACTAG (SEQ ID NO: 273); GGGTTACCCTGGACATGGAA (SEQ ID NO: 274); CATGGAAGGGTATTTACTCG (SEQ ID NO: 275); AGAGTGGCCTAGACAGGCTG (SEQ ID NO: 276); CATGCTGGACAGCTCGGCAG (SEQ ID NO: 277); AGTGAAAGAAGAGAAAATTC (SEQ ID NO: 278); TGGTAAGTCTAAGAAACCTA (SEQ ID NO: 279); CCCACAGCCTAACCACCTA (SEQ ID NO: 280); AATATTTCAAAGCCCTAGGG (SEQ ID NO: 281); GCACTCGGAACAGGGTCTGG (SEQ ID NO: 282); AGATAGGAGCTCCAACAGTG (SEQ ID NO: 283); AAGTTAGAGCAGCCAGGAAA (SEQ ID NO: 284); TAGAGCAGCCAGGAAAAGGGA (SEQ ID NO: 285); TGAATACCCCTCCATGTCCA (SEQ ID NO: 286);

CCTGCATTGCACCAGGCACA (SEQ ID NO: 287); TCTAGGGCCCCAACACACCT (SEQ ID NO: 288); TCCCTCCATCTATCAAAAGG (SEQ ID NO: 289); AGCCCTGAGACAGAAGCAGG (SEQ ID NO: 290); GCCCTGAGACAGAAGCAGGT (SEQ ID NO: 291); AGGAGATGCAGTGATACGCA (SEQ ID NO: 292); ACAATACCAAGGGTATCCGG (SEQ ID NO: 293); TGATAAAGAAAACAAAGTGA (SEQ ID NO: 294); AAAGAAAACAAAGTGAGGGA (SEQ ID NO: 295); GTGGCAAGTGGAGAAATTGA (SEQ ID NO: 296); CAAGTGGAGAAATTGAGGGA (SEQ ID NO: 297); GTGGTGATGATTGCAGCTGG (SEQ ID NO: 298); CTATGTGCCTGACACACAGG (SEQ ID NO: 299); GGGTTGGACCAGGAAAGAGG (SEQ ID NO: 300); GATGCCTGGAAAAGGAAAGA (SEQ ID NO: 301); TAGTATGCACCTGCAAGAGG (SEQ ID NO: 302); TATGCACCTGCAAGAGGCCGG (SEQ ID NO: 303); AGGGGAAGAAGAGAAGCAGA (SEQ ID NO: 304); GCTGAATCAAGAGACAAGCG (SEQ ID NO: 305); AAGCAAATAAATCTCCTGGG (SEQ ID NO: 306); AGATGAGTGCTAGAGACTGG (SEQ ID NO: 307); and CTGATGGTTGAGCACAGCAG (SEQ ID NO: 308).

In embodiments, the guide RNAs are: AATCGAGAAGCGACTCGACA (SEQ ID NO: 425), and tgcctgcaggggagtgago (SEQ ID NO: 426). In embodiments, the guide RNAs are gaagcgactcgacatggagg (SEQ ID NO: 427) and octgcaggggagtgagcage (SEQ ID NO: 428).

In embodiments, guide RNAs (gRNAs) for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, *e.g.*, without limitation dCas, in areas of open chromatin are as shown in **TABLE 3A-3F**.

In embodiments, guide RNAs (gRNAs) for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, *e.g.*, without limitation dCas, in areas of open chromatin are as shown in **TABLE 3A**:

GSHS	Identifier	Sequence
AAVS1	14F	ggagccacgaaaacagatcc (SEQ ID NO: 800)
AAVS1	15F	cgaaaacagatccagggaca (SEQ ID NO: 801)
AAVS1	16F	agatccagggacacggtgct (SEQ ID NO: 802)
AAVS1	17F	gacacggtgctaggacagt (SEQ ID NO: 803)
AAVS1	18F	gaaaatgacccaacagcctc (SEQ ID NO: 804)
AAVS1	19F	gcctggcggcctgaccact (SEQ ID NO: 805)
AAVS1	20F	ctgagcactgaaggcctggc (SEQ ID NO: 806)
AAVS1	21F	tggttccactgagcactga (SEQ ID NO: 807)
AAVS1	22F	gatagccaggagtccttccg (SEQ ID NO: 808)
AAVS1	23F	gcgttccagtgctcagact (SEQ ID NO: 809)
AAVS1	24F	cagtgcacagactaggaag (SEQ ID NO: 810)
AAVS1	25F	gcccctcctcctcagagcc (SEQ ID NO: 811)
AAVS1	26F	tccttcagagccaggagtc (SEQ ID NO: 812)
AAVS1	27F	tggttccagcttgacct (SEQ ID NO: 813)
AAVS1	28F	ctgcagagtatctgctggg (SEQ ID NO: 814)

AAVS1	29F	cgttcctgcagagtatctgc (SEQ ID NO: 815)
AAVS1	AAVS1	TCCCCTCCCAGAAAGACCTG (SEQ ID NO: 131)
AAVS1	gAAVS2	TGGGCTCCAAGCAATCCTGG (SEQ ID NO: 132)
AAVS1	gAAVS3	GTGGCTCAGGAGGTACCTGG (SEQ ID NO: 133)
AAVS1	gAAVS4	GAGCCACGAAAACAGATCCA (SEQ ID NO: 134)
AAVS1	gAAVS5	AAGTGAACGGGGAAGGGAGG (SEQ ID NO: 135)
AAVS1	gAAVS6	GACAAAAGCCGAAGTCCAGG (SEQ ID NO: 136)
AAVS1	gAAVS7	GTGGTTGATAAACCCACGTG (SEQ ID NO: 137)
AAVS1	gAAVS8	TGGGAACAGCCACAGCAGGG (SEQ ID NO: 138)
AAVS1	gAAVS9	GCAGGGGAACGGGGATGCAG (SEQ ID NO: 139)
AAVS1	gAAVS10	GAGATGGTGGACGAGGAAGG (SEQ ID NO: 140)
AAVS1	gAAVS11	GAGATGGCTCCAGGAAATGG (SEQ ID NO: 141)
AAVS1	gAAVS12	TAAGGAATCTGCCTAACAGG (SEQ ID NO: 142)
AAVS1	gAAVS13	TCAGGAGACTAGGAAGGAGG (SEQ ID NO: 143)
AAVS1	gAAVS14	TATAAGGTGGTCCCAGCTCG (SEQ ID NO: 144)
AAVS1	gAAVS15	CTGGAAGATGCCATGACAGG (SEQ ID NO: 145)
AAVS1	gAAVS16	GCACAGACTAGAGAGGTAAG (SEQ ID NO: 146)
AAVS1	gAAVS17	ACAGACTAGAGAGGTAAGGG (SEQ ID NO: 147)
AAVS1	gAAVS18	GAGAGGTGACCCGAATCCAC (SEQ ID NO: 148)
AAVS1	gAAVS19	GCACAGGCCCCCAGAAGGAGA (SEQ ID NO: 149)
AAVS1	gAAVS20	CCGGAGAGGACCCAGACACG (SEQ ID NO: 150)
AAVS1	gAAVS21	GAGAGGACCCAGACACGGGG (SEQ ID NO: 151)
AAVS1	gAAVS22	GCAACACAGCAGAGAGCAAG (SEQ ID NO: 152)
AAVS1	gAAVS23	GAAGAGGGAGTGGAGGAAGA (SEQ ID NO: 153)
AAVS1	gAAVS24	AAGACGGAACCTGAAGGAGG (SEQ ID NO: 154)
AAVS1	gAAVS25	AGAAAGCGGCACAGGCCCCAG (SEQ ID NO: 155)
AAVS1	gAAVS26	GGGAAACAGTGGGCCAGAGG (SEQ ID NO: 156)
AAVS1	gAAVS27	GTCCGGACTCAGGAGAGAGA (SEQ ID NO: 157)
AAVS1	gAAVS28	GGCACAGCAAGGGCACTCGG (SEQ ID NO: 158)
AAVS1	gAAVS29	GAAGAGGGGAAGTCGAGGGA (SEQ ID NO: 159)
AAVS1	gAAVS30	GGGAATGGTAAGGAGGCCTG (SEQ ID NO: 160)
AAVS1	gAAVS31	GCAGAGTGGTCAGCACAGAG (SEQ ID NO: 161)
AAVS1	gAAVS32	GCACAGAGTGGCTAAGCCCA (SEQ ID NO: 162)
AAVS1	gAAVS33	GACGGGGTGTCAGCATAGGG (SEQ ID NO: 163)
AAVS1	gAAVS34	GCCCAGGGCCAGGAACGACG (SEQ ID NO: 164)
AAVS1	gAAVS35	GGTGGAGTCCAGCACGGCGC (SEQ ID NO: 165)
AAVS1	gAAVS36	ACAGGCCGCCAGGAACTCGG (SEQ ID NO: 166)
AAVS1	gAAVS37	ACTAGGAAGTGTGTAGCACC (SEQ ID NO: 167)
AAVS1	gAAVS38	ATGAATAGCAGACTGCCCCG (SEQ ID NO: 168)

AAVS1	gAAVS39	ACACCCCTAAAAGCACAGTG (SEQ ID NO: 169)
AAVS1	gAAVS40	CAAGGAGTTCAGCAGGTGG (SEQ ID NO: 170)
AAVS1	gAAVS41	AAGGAGTTCAGCAGGTGGG (SEQ ID NO: 171)
AAVS1	gAAVS42	TGGAAGAGGAGGGAAGAGG (SEQ ID NO: 172)
AAVS1	gAAVS43	TCGAATTCCTAACTGCCCCG (SEQ ID NO: 173)
AAVS1	gAAVS44	GACCTGCCCAGCACACCCTG (SEQ ID NO: 174)
AAVS1	gAAVS45	GGAGCAGCTGCGCCAGTGGG (SEQ ID NO: 175)
AAVS1	gAAVS46	GGGAGGGAGAGCTTGGCAGG (SEQ ID NO: 176)
AAVS1	gAAVS47	GTTACGTGGCCAAGAAGCAG (SEQ ID NO: 177)
AAVS1	gAAVS48	GCTGAACAGAGAAGAGCTGG (SEQ ID NO: 178)
AAVS1	gAAVS49	TCTGAGGGTGGAGGGACTGG (SEQ ID NO: 179)
AAVS1	gAAVS50	GGAGAGGTGAGGGACTTGGG (SEQ ID NO: 180)
AAVS1	gAAVS51	GTGAACCAGGCAGACAAACGA (SEQ ID NO: 181)
AAVS1	gAAVS52	CAGGTACCTCCTGAGCCACG (SEQ ID NO: 182)
AAVS1	gAAVS53	GGGGGAGTAGGGGCATGCAG (SEQ ID NO: 183)
hROSA26	gHROSA26-1	GCAAATGGCCAGCAAGGGTG (SEQ ID NO: 184)
hROSA26	gHROSA26-2	CAATGGCCAGCAAGGGTGG (SEQ ID NO: 309)
hROSA26	gHROSA26-3	GCAGAACCTGAGGATATGGA (SEQ ID NO: 310)
hROSA26	gHROSA26-3	AATACACAGAATGAAAATAG (SEQ ID NO: 311)
hROSA26	gHROSA26-4	CTGGTGACTAGAATAGGCAG (SEQ ID NO: 312)
hROSA26	gHROSA26-5	TGGTGACTAGAATAGGCAGT (SEQ ID NO: 313)
hROSA26	gHROSA26-6	TAAAAGAATGTGAAAAGATG (SEQ ID NO: 314)
hROSA26	gHROSA26-7	TCAGGAGTTCAAGACCACCC (SEQ ID NO: 315)
hROSA26	gHROSA26-8	TGTAGTCCCAGTTATGCAGG (SEQ ID NO: 316)
hROSA26	gHROSA26-9	GGGTTCACACCACAAATGCA (SEQ ID NO: 317)
hROSA26	gHROSA26-10	GGCAAATGGCCAGCAAGGGT (SEQ ID NO: 318)
hROSA26	gHROSA26-11	AGAAACCAATCCCAAAGCAA (SEQ ID NO: 319)
hROSA26	gHROSA26-12	GCCAAGGACACCAAAAACCCA (SEQ ID NO: 320)
hROSA26	gHROSA26-13	AGTGGTGATAAGGCAACAGT (SEQ ID NO: 321)
hROSA26	gHROSA26-14	CCTGAGACAGAAGTATTAAG (SEQ ID NO: 322)
hROSA26	gHROSA26-15	AAGGTCACACAATGAATAGG (SEQ ID NO: 323)
hROSA26	gHROSA26-16	CACCATACTAGGGAAGAAGA (SEQ ID NO: 324)
hROSA26	gHROSA26-17	CAATACCCTGCCCTTAGTGG (SEQ ID NO: 327)
hROSA26	gHROSA26-18	AATACCCTGCCCTTAGTGGG (SEQ ID NO: 325)
hROSA26	gHROSA26-19	TTAGTGGGGGGTGGAGTGGG (SEQ ID NO: 326)
hROSA26	gHROSA26-20	GTGGGGGGTGGAGTGGGGGG (SEQ ID NO: 328)
hROSA26	gHROSA26-21	GGGGGGTGGAGTGGGGGGTG (SEQ ID NO: 329)
hROSA26	gHROSA26-22	GGGGTGGAGTGGGGGGTGGG (SEQ ID NO: 330)
hROSA26	gHROSA26-23	GGGTGGAGTGGGGGGTGGGG (SEQ ID NO: 331)

hROSA26	gHROSA26-24	GGGGGTGGGGAAAAGACATCG (SEQ ID NO: 332)
hROSA26	gHROSA26-25	GCAAATGGCCAGCAAGGGTG (SEQ ID NO: 184)
hROSA26	gHROSA26-26	CAAAATGGCCAGCAAGGGTGG (SEQ ID NO: 309)
hROSA26	gHROSA26-27	GCAGAACCTGAGGATATGGA (SEQ ID NO: 310)
hROSA26	gHROSA26-28	AATACACAGAATGAAAATAG (SEQ ID NO: 311)
hROSA26	gHROSA26-29	CTGGTGACTIONAGATAGGCAG (SEQ ID NO: 312)
hROSA26	gHROSA26-30	TGGTGACTIONAGATAGGCAGT (SEQ ID NO: 313)
hROSA26	gHROSA26-31	TAAAGAATGTGAAAAGATG (SEQ ID NO: 314)
hROSA26	gHROSA26-32	TCAGGAGTTCAAGACCACCC (SEQ ID NO: 315)
hROSA26	gHROSA26-33	TGTAGTCCCAGTTATGCAGG (SEQ ID NO: 316)
hROSA26	gHROSA26-34	GGGTTCACACCACAAATGCA (SEQ ID NO: 317)
hROSA26	gHROSA26-35	GGCAAATGGCCAGCAAGGGT (SEQ ID NO: 318)
hROSA26	gHROSA26-36	AGAAACCAATCCCAAAGCAA (SEQ ID NO: 319)
hROSA26	gHROSA26-37	GCCAAGGACACCAAAACCCA (SEQ ID NO: 320)
hROSA26	gHROSA26-38	AGTGGTGATAAGGCAACAGT (SEQ ID NO: 321)
hROSA26	gHROSA26-39	CCTGAGACAGAAAGTATTAAG (SEQ ID NO: 322)
hROSA26	gHROSA26-40	AAGGTCACACAATGAATAGG (SEQ ID NO: 323)
hROSA26	gHROSA26-41	CACCATACTAGGGAAGAAGA (SEQ ID NO: 324)
hROSA26	gHROSA26-42	CAATACCCTGCCCTTAGTGG (SEQ ID NO: 327)
hROSA26	gHROSA26-43	AATACCCTGCCCTTAGTGGG (SEQ ID NO: 325)
hROSA26	gHROSA26-44	TTAGTGGGGGTGGAGTGGG (SEQ ID NO: 326)
hROSA26	gHROSA26-45	GTGGGGGTGGAGTGGGGG (SEQ ID NO: 328)
hROSA26	gHROSA26-46	GGGGGTGGAGTGGGGGTG (SEQ ID NO: 329)
hROSA26	gHROSA26-47	GGGTGGAGTGGGGGTGGG (SEQ ID NO: 330)
hROSA26	gHROSA26-48	GGGTGGAGTGGGGGTGGG (SEQ ID NO: 331)
hROSA26	gHROSA26-49	GGGGGTGGGGAAAAGACATCG (SEQ ID NO: 332)
hROSA26	gHROSA26-50	GCAGCTGTGAATTCTGATAG (SEQ ID NO: 333)
hROSA26	gHROSA26-51	GAGATCAGAGAAACCAGATG (SEQ ID NO: 334)
hROSA26	gHROSA26-52	TCTATACTGATTGCAGCCAG (SEQ ID NO: 335)
hROSA26	gHROSA26-1	GCAAATGGCCAGCAAGGGTG (SEQ ID NO: 184)
hROSA26	44F	CACCGAATCGAGAAGCGACTCGACA (SEQ ID NO: 185)
hROSA26	45F	CACCGGTCCCTGGGCGTTGCCCTGC (SEQ ID NO: 186)
hROSA26	46F	CACCGCCCTGGGCGTTGCCCTGCAG (SEQ ID NO: 187)
hROSA26	1nF	CACCGCCGTGGGAAGATAAACTAAT (SEQ ID NO: 188)
hROSA26	2nF	CACCGTCCCTGCAGGGCAACGCC (SEQ ID NO: 189)
hROSA26	3nF	CACCGTCCAGTCCGCTTCTCGATTA (SEQ ID NO: 190)
hROSA26	4nF	CACCGTCCGCTCCCTCCGCTTTGTA (SEQ ID NO: 191)
hROSA26	5nF	CACCGAGTGCCGCAATACCTTTAT (SEQ ID NO: 192)
hROSA26	6nF	CACCGACACTTTGGTGGTGCAGCAA (SEQ ID NO: 193)

hROSA26	7nF	CACCGTCTCAAATGGTATAAACTC (SEQ ID NO: 194)
hROSA26	8nF	CACCGCCGTGGGAAGATAAACTAAT (SEQ ID NO: 188)
hROSA26	9F	CACCGAATCCCGCCCATAATCGAGA (SEQ ID NO: 195)
hROSA26	10F	CACCGTCCCGCCCATAATCGAGAAG (SEQ ID NO: 196)
hROSA26	11F	CACCGCCCATAATCGAGAAGCGACT (SEQ ID NO: 197)
hROSA26	12F	CACCGGAGAAGCGACTCGACATGGA (SEQ ID NO: 198)
hROSA26	13F	CACCGGAAGCGACTCGACATGGAGG (SEQ ID NO: 199)
hROSA26	14F	CACCGGCGACTCGACATGGAGGCGA (SEQ ID NO: 200)
hROSA26	44F	AAACTGTCGAGTCGCTTCTCGATTG (SEQ ID NO: 201)
hROSA26	45F	AAACGCAGGGCAACGCCAGGGACC (SEQ ID NO: 202)
hROSA26	46F	AAACCTGCAGGGCAACGCCAGGGC (SEQ ID NO: 203)
hROSA26	1nR	AAACATTAGTTTATCTTCCACGGC (SEQ ID NO: 204)
hROSA26	2nR	AAACGGGCGTTGCCCTGCAGGGGAC (SEQ ID NO: 205)
hROSA26	3nR	AAACTAATCGAGAAGCGACTCGACC (SEQ ID NO: 206)
hROSA26	4nR	AAACTACAAGACGGGAGGCAGCAGC (SEQ ID NO: 207)
hROSA26	5nR	AAACATAAAGGTATTGCGGCACTCC (SEQ ID NO: 208)
hROSA26	6nR	AAACTTGCTGCACCACCAAAGTGTC (SEQ ID NO: 209)
hROSA26	7nR	AAACGAGTTTATACCATTGAGAC (SEQ ID NO: 210)
hROSA26	8nR	AAACATTAGTTTATCTTCCACGGC (SEQ ID NO: 204)
hROSA26	9R	AAACTCTCGATTATGGGCGGGATTG (SEQ ID NO: 211)
hROSA26	10R	AAACCTTCTCGATTATGGGCGGGAC (SEQ ID NO: 212)
hROSA26	11R	AAACAGTCGCTTCTCGATTATGGGC (SEQ ID NO: 213)
hROSA26	12R	AAACTCCATGTCGAGTCGCTTCTCC (SEQ ID NO: 214)
hROSA26	13R	AAACCTCCATGTCGAGTCGCTTCC (SEQ ID NO: 215)
hROSA26	14R	AAACTCGCCTCCATGTCGAGTCGCC (SEQ ID NO: 216)
CCR5	1F	CACCGACAGGGTTAATGTGAAGTCC (SEQ ID NO: 217)
CCR5	2F	CACCGTCCCCCTCTACATTTAAAGT (SEQ ID NO: 218)
CCR5	3F	CACCGCATTAAAGTTGGTTTAAGT (SEQ ID NO: 219)
CCR5	4F	CACCGTTAGAAAATATAAAGAATAA (SEQ ID NO: 220)
CCR5	5	CACCGTAAATGCTTACTGGTTTGAA (SEQ ID NO: 221)
CCR5	6F	CACCGTCCTGGGTCCAGAAAAAGAT (SEQ ID NO: 222)
CCR5	7F	CACCGTTGGGTGGTGAGCATCTGTG (SEQ ID NO: 223)
CCR5	8F	CACCGCGGGGAGAGTGGAGAAAAAG (SEQ ID NO: 224)
CCR5	9F	CACCGGTTAAAACTCTTTAGACAAC (SEQ ID NO: 225)
CCR5	10F	CACCGGAAAATCCCCACTAAGATCC (SEQ ID NO: 226)
CCR5	1R	AAACGGACTTCACATTAACCCTGTC (SEQ ID NO: 227)
CCR5	2R	AAACACTTTAAATGTAGAGGGGGAC (SEQ ID NO: 228)
CCR5	3R	AAACACTTAAACCACTTTAAATGC (SEQ ID NO: 229)
CCR5	4R	AAACTTATTCTTTATATTTTCTAAC (SEQ ID NO: 230)

CCR5	5R	AAACTTCAAACCAGTAAGCATTTAC (SEQ ID NO: 231)
CCR5	6R	AAACATCTTTTTCTGGACCCAGGAC (SEQ ID NO: 232)
CCR5	7R	AAACCACAGATGCTCACCACCCAAC (SEQ ID NO: 233)
CCR5	8R	AAACCTTTTTCTCCACTCTCCCCGC (SEQ ID NO: 234)
CCR5	9R	AAACGTTGTCTAAAGAGTTTAAACC (SEQ ID NO: 235)
CCR5	10R	AAACGGATCTTAGTGGGGATTTTCC (SEQ ID NO: 236)
CCR5	gCCR5-1	AGTAGCAGTAATGAAGCTGG (SEQ ID NO: 237)
CCR5	gCCR5-2	ATACCCAGACGAGAAAGCTG (SEQ ID NO: 238)
CCR5	gCCR5-3	TACCCAGACGAGAAAGCTGA (SEQ ID NO: 239)
CCR5	gCCR5-4	GGTGGTGAGCATCTGTGTGG (SEQ ID NO: 240)
CCR5	gCCR5-5	AAATGAGAAGAAGAGGCACA (SEQ ID NO: 241)
CCR5	gCCR5-6	CTTGTGGCCTGGGAGAGCTG (SEQ ID NO: 242)
CCR5	gCCR5-7	GCTGTAGAAGGAGACAGAGC (SEQ ID NO: 243)
CCR5	gCCR5-8	GAGCTGGTTGGGAAGACATG (SEQ ID NO: 244)
CCR5	gCCR5-9	CTGGTTGGGAAGACATGGGG (SEQ ID NO: 245)
CCR5	gCCR5-10	CGTGAGGATGGGAAGGAGGG (SEQ ID NO: 246)
CCR5	gCCR5-11	ATGCAGAGTCAGCAGAACTG (SEQ ID NO: 247)
CCR5	gCCR5-12	AAGACATCAAGCACAGAAGG (SEQ ID NO: 248)
CCR5	gCCR5-13	TCAAGCACAGAAGGAGGAGG (SEQ ID NO: 249)
CCR5	gCCR5-14	AACCGTCAATAGGCAAAGGG (SEQ ID NO: 250)
CCR5	gCCR5-15	CCGTATTTCAAGTGAATGG (SEQ ID NO: 251)
CCR5	gCCR5-16	GAGAGGACAGGTGCTACAGG (SEQ ID NO: 252)
CCR5	gCCR5-17	AACCAAGGAAGGGCAGGAGG (SEQ ID NO: 253)
CCR5	gCCR5-18	GACCTCTGGGTGGAGACAGA (SEQ ID NO: 254)
CCR5	gCCR5-19	CAGATGACCATGACAAGCAG (SEQ ID NO: 255)
CCR5	gCCR5-20	AACACCAGTGAGTAGAGCGG (SEQ ID NO: 256)
CCR5	gCCR5-21	AGGACCTTGAAGCACAGAGA (SEQ ID NO: 257)
CCR5	gCCR5-22	TACAGAGGCAGACTAACCCA (SEQ ID NO: 258)
CCR5	gCCR5-23	ACAGAGGCAGACTAACCCAG (SEQ ID NO: 259)
CCR5	gCCR5-24	TAAATGACGTGCTAGACCTG (SEQ ID NO: 260)
CCR5	gCCR5-25	AGTAACCACTCAGGACAGGG (SEQ ID NO: 261)
chr2	gchr2-1	ACCACAAAACAGAAACACCA (SEQ ID NO: 262)
chr2	gchr2-2	GTTTGAAGACAAGCCTGAGG (SEQ ID NO: 263)
chr4	gchr4-1	GCTGAACCCCAAAGACAGG (SEQ ID NO: 264)
chr4	gchr4-2	GCAGCTGAGACACACACCAG (SEQ ID NO: 265)
chr4	gchr4-3	AGGACACCCCAAAGAAGCTG (SEQ ID NO: 266)
chr4	gchr4-4	GGACACCCCAAAGAAGCTGA (SEQ ID NO: 267)
chr6	gchr6-1	CCAGTGCAATGGACAGAAGA (SEQ ID NO: 268)
chr6	gchr6-2	AGAAGAGGGAGCCTGCAAGT (SEQ ID NO: 269)

chr6	gchr6-3	GTGTTTGGGCCCTAGAGCGA (SEQ ID NO: 270)
chr6	gchr6-4	CATGTGCCTGGTGCAATGCA (SEQ ID NO: 271)
chr6	gchr6-5	TACAAAGAGGAAGATAAGTG (SEQ ID NO: 272)
chr6	gchr6-6	GTCACAGAATACACCACTAG (SEQ ID NO: 273)
chr6	gchr6-7	GGGTTACCCTGGACATGGAA (SEQ ID NO: 274)
chr6	gchr6-8	CATGGAAGGGTATTCACTCG (SEQ ID NO: 275)
chr6	gchr6-9	AGAGTGGCCTAGACAGGCTG (SEQ ID NO: 276)
chr6	gchr6-10	CATGCTGGACAGCTCGGCAG (SEQ ID NO: 277)
chr6	gchr6-11	AGTGAAAGAAGAGAAAATTC (SEQ ID NO: 278)
chr6	gchr6-12	TGGTAAGTCTAAGAAACCTA (SEQ ID NO: 279)
chr6	gchr6-13	CCCACAGCCTAACCACCCTA (SEQ ID NO: 280)
chr6	gchr6-14	AATATTTCAAAGCCCTAGGG (SEQ ID NO: 281)
chr6	gchr6-15	GCACTCGGAACAGGGTCTGG (SEQ ID NO: 282)
chr6	gchr6-16	AGATAGGAGCTCCAACAGTG (SEQ ID NO: 283)
chr6	gchr6-17	AAGTTAGAGCAGCCAGGAAA (SEQ ID NO: 284)
chr6	gchr6-18	TAGAGCAGCCAGGAAAGGGA (SEQ ID NO: 285)
chr6	gchr6-19	TGAATACCCTTCCATGTCCA (SEQ ID NO: 286)
chr6	gchr6-20	CCTGCATTGCACCAGGCACA (SEQ ID NO: 287)
chr6	gchr6-21	TCTAGGGCCCCAACACACCT (SEQ ID NO: 288)
chr6	gchr6-22	TCCCTCCATCTATCAAAAGG (SEQ ID NO: 289)
chr10	gchr10-1	AGCCCTGAGACAGAAGCAGG (SEQ ID NO: 290)
chr10	gchr10-2	GCCCTGAGACAGAAGCAGGT (SEQ ID NO: 291)
chr10	gchr10-3	AGGAGATGCAGTGATACGCA (SEQ ID NO: 292)
chr10	gchr10-4	ACAATACCAAGGGTATCCGG (SEQ ID NO: 293)
chr10	gchr10-5	TGATAAAGAAAACAAAGTGA (SEQ ID NO: 294)
chr10	gchr10-6	AAAGAAAACAAAGTGAGGGA (SEQ ID NO: 295)
chr10	gchr10-7	GTGGCAAGTGGAGAAATTGA (SEQ ID NO: 296)
chr10	gchr10-8	CAAGTGGAGAAATTGAGGGA (SEQ ID NO: 297)
chr10	gchr10-9	GTGGTGATGATTGCAGCTGG (SEQ ID NO: 298)
chr11	gchr11-1	CTATGTGCCTGACACACAGG (SEQ ID NO: 299)
chr11	gchr11-2	GGGTTGGACCAGGAAAGAGG (SEQ ID NO: 300)
chr17	gchr17-1	GATGCCTGGAAAAGGAAAGA (SEQ ID NO: 301)
chr17	gchr17-2	TAGTATGCACCTGCAAGAGG (SEQ ID NO: 302)
chr17	gchr17-3	TATGCACCTGCAAGAGGCGG (SEQ ID NO: 303)
chr17	gchr17-4	AGGGGAAGAAGAGAAGCAGA (SEQ ID NO: 304)
chr17	gchr17-5	GCTGAATCAAGAGACAAGCG (SEQ ID NO: 305)
chr17	gchr17-6	AAGCAAATAAATCTCCTGGG (SEQ ID NO: 306)
chr17	gchr17-7	AGATGAGTGCTAGAGACTGG (SEQ ID NO: 307)
chr17	gchr17-8	CTGATGGTTGAGCACAGCAG (SEQ ID NO: 308)

In embodiments, gRNAs for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, e.g., without limitation dCas, to the TTAA site in hROSA26 (e.g., hg38 chr3:9,396,133-9,396,305) are shown in TABLE 3B:

HROSA26 GUIDE NO.	DNA SEQUENCE	SEQ ID NO:
GUIDE 44	AATCGAGAAGCGACTCGACA	425
GUIDE 45-C	GTCCCTGGGCGTTGCCCTGC	442
GUIDE 46-C	CCCTGGGCGTTGCCCTGCAG	443
SPG GUIDE1-C	GAGTGAGCAGCTGTAAGATT	444
SPG GUIDE2-C	CAGGGGAGTGAGCAGCTGTA	445
SPG GUIDE3-C	CCTGCAGGGGAGTGAGCAGC	428
SPG GUIDE4-C	TGCCCTGCAGGGGAGTGAGC	426
SPG GUIDE5-C	CGTTGCCCTGCAGGGGAGTG	446
SPG GUIDE6-C	TGGGCGTTGCCCTGCAGGGG	447
SPG GUIDE7-C	TTGGTCCCTGGGCGTTGCCC	448
SPG GUIDE8	AAGAATCCCGCCCATATCG	449
SPG GUIDE9	AATCCCGCCCATATCGAGA	450
SPG GUIDE10	TCCCGCCCATATCGAGAAG	451
SPG GUIDE11	CCCATATCGAGAAGCGACT	452
SPG GUIDE12	GAGAAGCGACTCGACATGGA	453
SPG GUIDE13	GAAGCGACTCGACATGGAGG	427
SPG GUIDE14	GCGACTCGACATGGAGGCGA	454
GUIDE N1	CCGTGGGAAGATAAACTAAT	455
GUIDE N2	TCCCCTGCAGGGCAACGCCC	456
GUIDE N3-C	GTCGAGTCGCTTCTCGATTA	457
GUIDE O12	CGACACCAACTCTAGTCCGT	458
GUIDE O13	CAGCTGCTCACTCCCCTGCA	459

GUIDE 014-C

AGTCGCTTCTCGATTATGGG

460

In embodiments, gRNAs for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, e.g., without limitation dCas, to the AAVS1 (e.g., hg38 chr19:55,112,851-55,113,324) are shown in TABLE 3C:

AAVS1 GUIDE NO.	DNA SEQUENCE	SEQ ID NO:
AAV GUIDE 12	ACCCTTGAAGGACCTGGCTGGG	461
AAV GUIDE 13c	TCCGAGCTTGACCCTTGGAA	462
AAV GUIDE 14	GGAGCCACGAAAACAGATCCAGG	463
AAV GUIDE 14c	TGGTTTCCGAGCTTGACCCT	112
AAV GUIDE 15	GGAGCCACGAAAACAGATCCAGG	463
AAV GUIDE 16	AGATCCAGGGACACGGTGCTAGG	464
AAV GUIDE 17	GACACGGTGCTAGGACAGTGGGG	465
AAV GUIDE 18	GAAAATGACCCAACAGCCTCTGG	466
AAV GUIDE 19	GCCTGGCCGGCCTGACCACTGGG	467
AAV GUIDE 20	CTGAGCACTGAAGGCCTGGCCGG	468
AAV GUIDE 21	TGGTTTCCACTGAGCACTGAAGG	469
AAV GUIDE 22	GGTGCTTTCCTGAGGACCGATAG	470
AAV GUIDE 23	GCGCTTCCAGTGCTCAGACTAGG	471
AAV GUIDE 24	CAGTGCTCAGACTAGGGAAGAGG	472
AAV GUIDE 25	GCCCCCTCCTCCTTCAGAGCCAGG	473
AAV GUIDE 26	TCCTTCAGAGCCAGGAGTCCTGG	474
AAV GUIDE 27	CCAAGGGTCAAGCTCGGAAACCA	475
AAV GUIDE 28	CTGCAGAGTATCTGCTGGGGTGG	476
AAV GUIDE 29	CGTTCCTGCAGAGTATCTGCTGG	477
AAV GUIDE 30c	GTGGGGAAAATGACCCAACA	478
AAV GUIDE 31	GAAGGCCTGGCCGGCCTGAC	479

AAV GUIDE 32c	ACTCCTGGCTCTGAAGGAGG	480
AAV GUIDE 33c	GGGCTGGGGGCCAGGACTCC	481
AAV GUIDE 34	GTCCTTCCAAGGGTCAAGCT	482
AAV GUIDE 35	TCAAGCTCGGAAACCACCCC	483

In embodiments, gRNAs for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, e.g., without limitation dCas, to Chromosome 4 (e.g., hg38 chr4:30,793,534-30,875,476 or hg38 chr4:30,793,533-30,793,537 (9677); chr4:30,875,472-30,875,476 (8948)) are shown in **TABLE 3D**:

CHR4 GUIDE NO.	DNA SEQUENCE	SEQ ID NO:
Guide C4-1	ATTGTCTTCACTAAACCCGTTGG	484
Guide C4-2	TAAACCCGTTGGGAATACAATGG	485
Guide C4-3	TTGTCTTCACTAAACCCGTTGGG	486
Guide C4-4	TGATTCATAGGAGTCTATTAAGG	487
Guide C4-5	TTACATATGCTTCGAGTTTGTGG	488
Example 1Guide C4-6	ACTCTTAAGGTAGGACTAATTGG	489
Guide C4-7	TATGTGTGCAATAGCGTTAAAGG	490
Guide C4-8	CGTTGGGAATACAATGGCTTAGG	491
Guide C4-9	TCACAATGGAAGCTCTGCCTTTGG	492
Guide C4-10	GACCACAAATCAATGCCCAAAGG	493
Guide C4-11	CTAAGCCATTGTATTCCCAACGG	494
Guide C4-12	AGCATTCTGGAGTGTCAACAATGG	495
Guide C4-13	CAATAGCCCACTTTAATACTAGG	496
Guide C4-14	CTTTATCCAAGTGAATCCTTTGG	497
Guide C4-15	GGCATTGATTTGTGGTCATTTGG	498
Guide C4-16	TAAGCCATTGTATTCCCAACGGG	499
Guide C4-17	AATACAATCACTCTTAAGGTAGG	500
Guide C4-18	GAAGTACCTTTCACTATTTTGGG	501

Guide C4-19	CAAGCAACAAATGACTTCTAAGG	502
Guide C4-20	TTTGAATACAATCACTCTTAAGG	503
Guide C4A1	ACAAACGGACTACGTAACTTGG	504
Guide C4A2	ACAAGATGTGAACACGACGATGG	505
Guide C4A3	GTTGCACCGTTGATTCCTTCAGG	506
Guide C4A4	AGTAATATTGAATTAGGGCGTGG	507
Guide C4A5	CCTGATGTTGGCTCGACATTAGG	508
Guide C4A6	CTTTGTTGGGTCTTAGCTTAAGG	509
Guide C4A7	TCGGAACAGCTCCTTCCTGAAGG	510
Guide C4A8	AGTAGTTTCTGAGGTCATGTTGG	511
Guide C4A9	CTTGAAAATACGATGATGTGAGG	512
Guide C4A10	GCATTAATCTAGAGAGAGGGAGG	513
Guide C4A11	GGGTCATGTTAGAATTCATGTGG	514
Guide C4A12	TGATGCATTAATCTAGAGAGAGG	515
Guide C4A13	ACATCATCGTATTTTCAAGTTGG	516
Guide C4A14	CTAGCTGACAAACATGTGAGTGG	517
Guide C4A15	AACATGACCCAAGTGAGTCCAGG	518
Guide C4A16	GATTCCGTATTTGCTTTGTTGGG	519
Guide C4A17	TACGATGATGTGAGGAAATAAGG	520
Guide C4A18	GTAATATGTCTAAGTACTGATGG	521
Guide C4A19	GTAAAGTGAGCTGGTTCATTAGG	522
Guide C4A20	ACTAGAGTCCTTAAGAAGGGGGG	523

In embodiments, gRNAs for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, e.g., without limitation dCas, to Chromosome 22 (e.g., hg38 chr22:35,370,000-35,380,000 or hg38 chr22:35,373,912-35,373,916 (861); chr22:35,377,843-35,377,847 (1153)) are shown in **TABLE 3E**:

CHR22 GUIDE NO.	DNA SEQUENCE	SEQ ID NO:
------------------------	---------------------	-------------------

Guide C22-1	ATAACACGTGAGCCGTCCTAAGG	524
Guide C22-2	GGAAGACTTTTCTCTATACGAGG	525
Guide C22-3	GCATTCTTTTCATCCATGGCAGG	526
Guide C22-4	GACATATGGTTATAAAAATCAGG	527
Guide C22-5	GGAGTGCAAGTCCCTGACATATGG	528
Guide C22-6	GTGGGTTAGGGTGGTTAACTGGG	529
Guide C22-7	AGGTGCAAAAAGGTTGCTGTGGG	530
Guide C22-8	CGTGACAAGGCAAAGTGGCGTGG	531
Guide C22-9	GAAGGACTGCCCCTGACGTCAGG	532
Guide C22-10	CTGCCCCTGACGTCAGGAGTTGG	533
Guide C22-11	TGTGGGTTAGGGTGGTTAACTGG	534
Guide C22-12	ACCCTTTTAGAGTTTTCTGCTGG	535
Guide C22-13	AACTTCCTGCCATGGATGAAAGG	536
Guide C22-14	GCAAAAAGGTTGCTGTGGGTGG	537
Guide C22-15	AATTTGGGGGTAGATAGGCATGG	538
Guide C22-16	AGAAACTCTAAAAGGGTATAGG	539
Guide C22-17	ATTAGCATTCTTTTCATCCATGG	540
Guide C22-18	CCCAGCAGAAACTCTAAAAGGG	541
Guide C22-19	CAGGTGCAAAAAGGTTGCTGTGG	542
Guide C22-20	GCAAGAGATGAAATTCCATATGG	543
Guide C22A1	GGGCTGTTCTAACGAAGTCTGGG	544
Guide C22A2	TGTCCATTTCAGCGACCCTAGAGG	545
Guide C22A3	GGCTGTTCTAACGAAGTCTGGGG	546
Guide C22A4	GTCCATTTCAGCGACCCTAGAGGG	547
Guide C22A5	GGGGCTGTTCTAACGAAGTCTGG	548
Guide C22A6	GGCTGAATCAGCATGCGAAAGGG	549
Guide C22A7	TTCCAATGGGGGGCATAGCCTGG	550

Guide C22A8	TACCCCTCTAGGGTCGCTGAATGG	551
Guide C22A9	ATCCTCTTGGGCCTTATAAGAGG	552
Guide C22A10	GGCCAGGCTATGCCCCCATTTGG	553
Guide C22A11	CTAGAGGACCAGAACAACCTCTGG	554
Guide C22A12	TCCCTCTTATAAGGCCCAAGAGG	555
Guide C22A13	AGGCTGAATCAGCATGCGAAAGG	556
Guide C22A14	GGACCAGAACAACCTCTGGCCTGG	557
Guide C22A15	GGGCTTTTATTTGGCCCAGCAGG	558
Guide C22A16	GTCGCTGAATGGACAGACTCTGG	559
Guide C22A17	CTCATGAGTTTTTACCCTCTAGGG	560
Guide C22A18	TCCTCTTGGGCCTTATAAGAGGG	561
Guide C22A19	TCTTGGGCCTTATAAGAGGGAGG	562
Guide C22A20	TAGAACAGCCCCCACACAGTGG	563

In embodiments, gRNAs for targeting human genomic safe harbor sites using any of the gRNA-based targeting elements, e.g., without limitation dCas, to Chromosome X (e.g., hg38 chrX:134,419,661-134,541,172 or hg38 chrX:134,476,304-134,476,307 (85); chrX:134,476,337-134,476,340 (51)) are shown in **TABLE 3F**:

CHRX GUIDE NO.	DNA SEQUENCE	SEQ ID NO:
Guide CX-1	GTTACGTTATGACTAATCTTTGG	564
Guide CX-2	TACGTTATGACTAATCTTTGGGG	565
Guide CX-3	GGAAGTAGTGTTATGATGTATGG	566
Guide CX-4	GTTATGATGTATGGGCATAAAGG	567
Guide CX-5	GAAGTAGTGTTATGATGTATGGG	568
Guide CX-6	ATAGCTGCTGGCAGTATAACTGG	569
Guide CX-7	GCATCACAACATTGACACTGTGG	570
Guide CX-8	AAGGCGAGTTTCTACAAAGATGG	571
Guide CX-9	TTACGTTATGACTAATCTTTGGG	572

Guide CX-10	CAAGACTGATTAAGACTGATGGG	573
Guide CX-11	AGCAGCAATGTATTAAAGGCTGG	574
Guide CX-12	CTACAGGATTGATGTAAACATGG	575
Guide CX-13	TGGGCATAAAGGGTTTAAATGGG	576
Guide CX-14	ACATCAATCCTGTAGGTGATTGG	577
Guide CX-15	ATTCTAGTCATTATAGCTGCTGG	578
Guide CX-16	CATCAATCCTGTAGGTGATTGGG	579
Guide CX-17	GTTATAAGATCAATTCTGAGTGG	580
Guide CX-18	GGCAGACTGTGGATCAAAAGTGG	581
Guide CX-19	ATGGCTGCCCAATCACCTACAGG	582
Guide CX-20	TCAAAGCATGTACTTAGAGTTGG	583

In embodiments, the gRNA comprises one or more of the sequences outlined herein or a variant sequence having at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, a Cas-based targeting element comprises Cas12 or a variant thereof, e.g., without limitation, Cas12a (e.g., dCas12a), or Cas12j (e.g., dCas12j), or Cas12k (e.g., dCas12k). In embodiments, the targeting element comprises a Cas12 enzyme guide RNA complex. In embodiments, comprises a nuclease-deficient dCas12 guide RNA complex, optionally dCas12j guide RNA complex or dCas12a guide RNA complex.

In embodiments, the targeting element is selected from a zinc finger (ZF), catalytically inactive Zinc finger, transcription activator-like effector (TALE), meganuclease, and clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein, any of which are, in embodiments, catalytically inactive. In embodiments, the CRISPR-associated protein is selected from Cas9, CasX, CasY, Cas12a (Cpf1), and gRNA complexes thereof. In embodiments, the CRISPR-associated protein is selected from Cas9, xCas9, Cas 6, Cas7, Cas8, Cas12a (Cpf1), Cas13a, Cas14, CasX, CasY, a Class 1 Cas protein, a Class 2 Cas protein, MAD7, MG1 nuclease, MG2 nuclease, MG3 nuclease, or catalytically inactive forms thereof, and gRNA complexes thereof.

In embodiments, the mobile element enzyme is capable of inserting a donor DNA at a TA dinucleotide site or a TTAA tetranucleotide site in a GSHS of a nucleic acid molecule. The mobile element enzyme is suitable for causing insertion of the donor DNA in a GSHS when contacted with a biological cell.

In embodiments, the targeting element is suitable for directing the mobile element enzyme to the GSHS sequence.

In embodiments, the targeting element comprises transcription activator-like effector (TALE) DNA binding domain (DBD). The TALE DBD comprises one or more repeat sequences. For example, in embodiments, the TALE DBD comprises about 14, or about 15, or about 16, or about 17, or about 18, or about 18.5 repeat sequences. In embodiments, the TALE DBD repeat sequences comprise 33 or 34 amino acids.

In embodiments, the one or more of the TALE DBD repeat sequences comprise a repeat variable di-residue (RVD) at residue 12 or 13 of the 33 or 34 amino acids.

In embodiments, the targeting element (e.g., TALE or Cas (e.g., Cas9 or Cas12, or variants thereof) DBDs cause the mammalian mobile element enzyme to bind specifically to human GSHS. In embodiments, the TALEs or Cas DBDs sequester the mobile element enzyme to GSHS and promote transposition to nearby TA dinucleotide or a TTAA tetranucleotide sites which can be located in proximity to the repeat variable di-residues (RVD) TALE or gRNA nucleotide sequences. The GSHS regions are located in open chromatin sites that are susceptible to mobile element enzyme activity. Accordingly, the mammalian mobile element enzyme does not only operate based on its ability to recognize TA or TTAA sites, but it also directs a donor DNA (having a transgene) to specific locations in proximity to a TALE or Cas DBD. The chimeric mobile element enzyme in accordance with embodiments of the present disclosure has negligible risk of genotoxicity and exhibits superior features as compared to existing gene therapies.

In embodiments, a chimeric mobile element enzyme is mutated to be characterized by reduced or inhibited binding of off-target sequences and consequently reliant on a DBD fused thereto, such as a TALE or Cas DBD, for transposition.

The described cells, compositions, and methods allow reducing vector and transgene insertions that increase a mutagenic risk. The described cells and methods make use of a gene transfer system that reduces genotoxicity compared to viral- and nuclease-mediated gene therapies. The dual system is designed to avoid the persistence of an active mobile element enzyme and efficiently transfect human cell lines without significant cytotoxicity.

In embodiments, TALE or Cas DBDs are customizable, such as a TALE or Cas DBDs is selected for targeting a specific genomic location. In embodiments, the genomic location is in proximity to a TA dinucleotide site or a TTAA (SEQ ID NO: 440) tetranucleotide site.

Embodiments of the present disclosure make use of the ability of TALE or Cas or dCas9/gRNA DBDs to target specific sites in a host genome. The DNA targeting ability of a TALE or Cas DBD or dCas9/gRNA DBD is provided by TALE repeat sequences (e.g., modular arrays) or gRNA which are linked together to recognize flanking DNA sequences. Each TALE or gRNA can recognize certain base pair(s) or residue(s).

TALE nucleases (TALENs) are a known tool for genome editing and introducing targeted double-stranded breaks. TALENs comprise endonucleases, such as FokI nuclease domain, fused to a customizable DBD. This DBD is

composed of highly conserved repeats from TALEs, which are proteins secreted by *Xanthomonas* bacteria to alter transcription of genes in host plant cells. The DBD includes a repeated highly conserved 33–34 amino acid sequence with divergent 12th and 13th amino acids. These two positions, referred to as the RVD, are highly variable and show a strong correlation with specific base pair or nucleotide recognition. This straightforward relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DBDs by selecting a combination of repeat segments containing the appropriate RVDs. Boch *et al. Nature Biotechnology*. 2011; 29 (2): 135–6.

Accordingly, TALENs can be readily designed using a “protein-DNA code” that relates modular DNA-binding TALE repeat domains to individual bases in a target-binding site. See Joung *et al. Nat Rev Mol Cell Biol*. 2013;14(1):49-55. doi:10.1038/nrm3486. The following table, for example, shows such code:

RVD	Nucleotide	RVD	Nucleotide
HD	C	NI	A
NH	G	NN	G, A
NK	G	NS	G, C, A
NG	T, mC		

It has been demonstrated that TALENs can be used to target essentially any DNA sequence of interest in human cell. Miller *et al. Nat Biotechnol*. 2011;29:143–148. Guidelines for selection of potential target sites and for use of particular TALE repeat domains (harboring NH residues at the hypervariable positions) for recognition of G bases have been proposed. See Streubel *et al. Nat Biotechnol*. 2012;30:593–595.

Accordingly, in embodiments, the TALE DBD comprises one or more repeat sequences. In embodiments, the TALE DBD comprises about 15, or about 16, or about 17, or about 18, or about 18.5 repeat sequences. In embodiments, the TALE DBD repeat sequences comprise 33 or 34 amino acids.

In embodiments, the one or more of the TALE DBD repeat sequences comprise an RVD at residue 12 or 13 of the 33 or 34 amino acids. The RVD can recognize certain base pair(s) or residue(s). In embodiments, the RVD recognizes one base pair in the nucleic acid molecule. In embodiments, the RVD recognizes a C residue in the nucleic acid molecule and is selected from HD, N(gap), HA, ND, and HI. In embodiments, the RVD recognizes a G residue in the nucleic acid molecule and is selected from NN, NH, NK, HN, and NA. In embodiments, the RVD recognizes an A residue in the nucleic acid molecule and is selected from NI and NS. In embodiments, the RVD recognizes a T residue in the nucleic acid molecule and is selected from NG, HG, H(gap), and IG.

In embodiments, the GSHS is in an open chromatin location in a chromosome. In embodiments, the GSHS is selected from adeno-associated virus site 1 (AAVS1), chemokine (C-C motif) receptor 5 (CCR5) gene, HIV-1 coreceptor, and human Rosa26 locus. In embodiments, the GSHS is located on human chromosome 2, 4, 6, 10, 11, 17, 22, or X.

In embodiments, the GSHS is selected from TALC1, TALC2, TALC3, TALC4, TALC5, TALC7, TALC8, AVS1, AVS2, AVS3, ROSA1, ROSA2, TALER1, TALER2, TALER3, TALER4, TALER5, SHCHR2-1, SHCHR2-2, SHCHR2-3, SHCHR2-4, SHCHR4-1, SHCHR4-2, SHCHR4-3, SHCHR6-1, SHCHR6-2, SHCHR6-3, SHCHR6-4, SHCHR10-1, SHCHR10-2, SHCHR10-3, SHCHR10-4, SHCHR10-5, SHCHR11-1, SHCHR11-2, SHCHR11-3, SHCHR17-1, SHCHR17-2, SHCHR17-3, and SHCHR17-4.

In embodiments, the GSHS comprises one or more of TGGCCGGCCTGACCACTGG (SEQ ID NO: 23), TGAAGGCCTGGCCGGCCTG (SEQ ID NO: 24), TGAGCACTGAAGGCCTGGC (SEQ ID NO: 25), TCCACTGAGCACTGAAGGC (SEQ ID NO: 26), TGGTTTCCACTGAGCACTG (SEQ ID NO: 27), TGGGGAAAATGACCCAACA (SEQ ID NO: 28), TAGGACAGTGGGGAAAATG (SEQ ID NO: 29), TCCAGGGACACGGTGCTAG (SEQ ID NO: 30), TCAGAGCCAGGAGTCCTGG (SEQ ID NO: 31), TCCTTCAGAGCCAGGAGTC (SEQ ID NO: 32), TCCTCCTTCAGAGCCAGGA (SEQ ID NO: 33), TCCAGCCCCTCCTCCTTCA (SEQ ID NO: 34), TCCGAGCTTGACCCTTGGA (SEQ ID NO: 35), TGGTTTCCGAGCTTGACCC (SEQ ID NO: 36), TGGGGTGGTTTCCGAGCTT (SEQ ID NO: 37), TCTGCTGGGGTGGTTTCCG (SEQ ID NO: 38), TGCAGAGTATCTGCTGGGG (SEQ ID NO: 39), CCAATCCCCTCAGT (SEQ ID NO: 40), CAGTGCTCAGTGGAA (SEQ ID NO: 41), GAAACATCCGGCGACTCA (SEQ ID NO: 42), TCGCCCCCTCAAATCTTACA (SEQ ID NO: 43), TCAAATCTTACAGCTGCTC (SEQ ID NO: 44), TCTTACAGCTGCTCACTCC (SEQ ID NO: 45), TACAGCTGCTCACTCCCCT (SEQ ID NO: 46), TGCTCACTCCCCTGCAGGG (SEQ ID NO: 47), TCCCCTGCAGGGCAACGCC (SEQ ID NO: 48), TGCAGGGCAACGCCAGGG (SEQ ID NO: 49), TCTCGATTATGGGCGGGAT (SEQ ID NO: 50), TCGCTTCTCGATTATGGGC (SEQ ID NO: 51), TGTCGAGTCGCTTCTCGAT (SEQ ID NO: 52), TCCATGTGCGAGTCGCTTCT (SEQ ID NO: 53), TCGCCTCCATGTGCGAGTCG (SEQ ID NO: 54), TCGTCATCGCCTCCATGTC (SEQ ID NO: 55), TGATCTCGTCATCGCCTCC (SEQ ID NO: 56), GCTTCAGCTTCCTA (SEQ ID NO: 57), CTGTGATCATGCCA (SEQ ID NO: 58), ACAGTGGTACACACCT (SEQ ID NO: 59), CCACCCCCCACTAAG (SEQ ID NO: 60), CATTGGCCGGGCAC (SEQ ID NO: 61), GCTTGAACCCAGGAGA (SEQ ID NO: 62), ACACCCGATCCACTGGG (SEQ ID NO: 63), GCTGCATCAACCCC (SEQ ID NO: 64), GCCACAAACAGAAATA (SEQ ID NO: 65), GGTGGCTCATGCCTG (SEQ ID NO: 66), GATTTGCACAGCTCAT (SEQ ID NO: 67), AAGCTCTGAGGAGCA (SEQ ID NO: 68), CCCTAGCTGTCCC (SEQ ID NO: 69), GCCTAGCATGCTAG (SEQ ID NO: 70), ATGGGCTTCACGGAT (SEQ ID NO: 71), GAACTATGCCTGC (SEQ ID NO: 72), GCACCATGCTCCC (SEQ ID NO: 73), GACATGCAACTCAG (SEQ ID NO: 74), ACACCACTAGGGGT (SEQ ID NO: 75), GTCTGCTAGACAGG (SEQ ID NO: 76), GGCCTAGACAGGCTG (SEQ ID NO: 77), GAGGCATTCTTATCG (SEQ ID NO: 78), GCCTGGAACGTTCC (SEQ ID NO: 79), GTGCTCTGACAATA (SEQ ID NO: 80), GTTTTGCAGCCTCC (SEQ ID NO: 81), ACAGCTGTGGAACGT (SEQ ID NO: 82), GGCTCTCTCCTCCT (SEQ ID NO: 83), CTATCCCAAACTCT (SEQ ID NO: 84), GAAAACTATGTAT (SEQ ID NO: 85), AGGCAGGCTGGTTGA (SEQ ID

NO: 86), CAATACAACCACGC (SEQ ID NO: 87), ATGACGGACTCAACT (SEQ ID NO: 88), CACAACATTTGTAA (SEQ ID NO: 89), and ATTTCCAGTGCACA (SEQ ID NO: 90).

In embodiments, the TALE DBD binds to one of TGGCCGGCCTGACCACTGG (SEQ ID NO: 23), TGAAGGCCTGGCCGGCCTG (SEQ ID NO: 24), TGAGCACTGAAGGCCTGGC (SEQ ID NO: 25), TCCACTGAGCACTGAAGGC (SEQ ID NO: 26), TGGTTTCCACTGAGCACTG (SEQ ID NO: 27), TGGGGAAAATGACCCAACA (SEQ ID NO: 28), TAGGACAGTGGGGAAAATG (SEQ ID NO: 29), TCCAGGGACACGGTGCTAG (SEQ ID NO: 30), TCAGAGCCAGGAGTCCTGG (SEQ ID NO: 31), TCCTTCAGAGCCAGGAGTC (SEQ ID NO: 32), TCCTCCTTCAGAGCCAGGA (SEQ ID NO: 33), TCCAGCCCCTCCTCCTTCA (SEQ ID NO: 34), TCCGAGCTTGACCCCTTGA (SEQ ID NO: 35), TGGTTTCCGAGCTTGACCC (SEQ ID NO: 36), TGGGGTGGTTTCCGAGCTT (SEQ ID NO: 37), TCTGCTGGGGTGGTTTCCG (SEQ ID NO: 38), TGCAGAGTATCTGCTGGGG (SEQ ID NO: 39), CCAATCCCCTCAGT (SEQ ID NO: 40), CAGTGCTCAGTGGAA (SEQ ID NO: 41), GAAACATCCGGCGACTCA (SEQ ID NO: 42), TCGCCCCTCAAATCTTACA (SEQ ID NO: 43), TCAAATCTTACAGCTGCTC (SEQ ID NO: 44), TCTTACAGCTGCTCACTCC (SEQ ID NO: 45), TACAGCTGCTCACTCCCCT (SEQ ID NO: 46), TGCTCACTCCCCTGCAGGG (SEQ ID NO: 47), TCCCCTGCAGGGCAACGCC (SEQ ID NO: 48), TGCAGGGCAACGCCAGGG (SEQ ID NO: 49), TCTCGATTATGGGCGGGAT (SEQ ID NO: 50), TCGCTTCTCGATTATGGGC (SEQ ID NO: 51), TGTCGAGTCGCTTCTCGAT (SEQ ID NO: 52), TCCATGTGCGAGTCGCTTCT (SEQ ID NO: 53), TCGCCTCCATGTGCGAGTCG (SEQ ID NO: 54), TCGTCATCGCCTCCATGTC (SEQ ID NO: 55), TGATCTCGTCATCGCCTCC (SEQ ID NO: 56), GCTTCAGCTTCCTA (SEQ ID NO: 57), CTGTGATCATGCCA (SEQ ID NO: 58), ACAGTGGTACACACCT (SEQ ID NO: 59), CCACCCCCCACTAAG (SEQ ID NO: 60), CATTGGCCGGGCAC (SEQ ID NO: 61), GCTTGAACCCAGGAGA (SEQ ID NO: 62), ACACCCGATCCACTGGG (SEQ ID NO: 63), GCTGCATCAACCCC (SEQ ID NO: 64), GCCACAAACAGAAATA (SEQ ID NO: 65), GGTGGCTCATGCCTG (SEQ ID NO: 66), GATTTGCACAGCTCAT (SEQ ID NO: 67), AAGCTCTGAGGAGCA (SEQ ID NO: 68), CCCTAGCTGTCCC (SEQ ID NO: 69), GCCTAGCATGCTAG (SEQ ID NO: 70), ATGGGCTTCAACGGAT (SEQ ID NO: 71), GAAACTATGCCTGC (SEQ ID NO: 72), GCACCATGCTCCC (SEQ ID NO: 73), GACATGCAACTCAG (SEQ ID NO: 74), ACACCACTAGGGGT (SEQ ID NO: 75), GTCTGCTAGACAGG (SEQ ID NO: 76), GGCCTAGACAGGCTG (SEQ ID NO: 77), GAGGCATTCTTATCG (SEQ ID NO: 78), GCCTGGAAACGTTCC (SEQ ID NO: 79), GTGCTCTGACAATA (SEQ ID NO: 80), GTTTTGCAGCCTCC (SEQ ID NO: 81), ACAGCTGTGGAACGT (SEQ ID NO: 82), GGCTCTCTTCCTCCT (SEQ ID NO: 83), CTATCCCAAACTCT (SEQ ID NO: 84), GAAAACTATGTAT (SEQ ID NO: 85), AGGCAGGCTGGTTGA (SEQ ID NO: 86), CAATACAACCACGC (SEQ ID NO: 87), ATGACGGACTCAACT (SEQ ID NO: 88), CACAACATTTGTAA (SEQ ID NO: 89), and ATTTCCAGTGCACA (SEQ ID NO: 90).

In embodiments, the TALE DBD comprises one or more of:

NH NH HD HD NH NH HD HD NG NH NI HD HD NI HD NG NH NH,
NH NI NI NH NH HD HD NG NH NH HD HD NH NH HD HD NG NH,
NH NI NH HD NI HD NG NH NI NI NH NH HD HD NG NH NH HD,
HD HD NI HD NG NH NI NH HD NI HD NG NH NI NI NH NH HD,
NH NH NG NG NG HD HD NI HD NG NH NI NH HD NI HD NG NH,
NH NH NH NH NI NI NI NI NG NH NI HD HD HD NI NI HD NI,
NI NH NH NI HD NI NH NG NH NH NH NH NI NI NI NI NG NH,
HD HD NI NH NH NH NI HD NI HD NH NH NG NH HD NG NI NH,
HD NI NH NI NH HD HD NI NH NH NI NH NG HD HD NG NH NH,
HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI NH NG HD,
HD HD NG HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI,
HD HD NI NH HD HD HD HD NG HD HD NG HD HD NG NG HD NI,
HD HD NH NI NH HD NG NG NH NI HD HD HD NG NG NH NH NI,
NH NH NG NG NG HD HD NH NI NH HD NG NG NH NI HD HD HD,
NH NH NH NH NG NH NH NG NG NG HD HD NH NI NH HD NG NG,
HD NG NH HD NG NH NH NH NH NG NH NH NG NG NG HD HD NH,
NH HD NI NH NI NH NG NI NG HD NG NH HD NG NH NH NH NH,
HD HD NI NI NG HD HD HD HD NG HD NI NH NG,
HD NI NH NG NH HD NG HD NI NH NG NH NH NI NI,
NH NI NI NI HD NI NG HD HD NH NH HD NH NI HD NG HD NI,
HD NH HD HD HD HD NG HD NI NI NI NG HD NG NG NI HD NI,
HD NI NI NI NG HD NG NG NI HD NI NH HD NG NH HD NG HD,
HD NG NG NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD,
NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD HD HD NG,
NH HD NG HD NI HD NG HD HD HD HD NG NH HD NI NH NH NH,

HD HD HD HD NG NH HD NI NH NH NH HD NI NI HD NH HD HD,
NH HD NI NH NH NH HD NI NI HD NH HD HD HD NI NH NH NH,
HD NG HD NH NI NG NG NI NG NH NH NH HD NH NH NH NI NG,
HD NH HD NG NG HD NG HD NH NI NG NG NI NG NH NH NH HD,
NH NG HD NH NI NH NG HD NH HD NG NG HD NG HD NH NI NG,
HD HD NI NG NH NG HD NH NI NH NG HD NH HD NG NG HD NG,
HD NH HD HD NG HD HD NI NG NH NG HD NH NI NH NG HD NH,
HD NH NG HD NI NG HD NH HD HD NG HD HD NI NG NH NG HD,
NH NI NG HD NG HD NH NG HD NI NG HD NH HD HD NG HD HD,
NH HD NG NG HD NI NH HD NG NG HD HD NG NI,
HD NG NK NG NH NI NG HD NI NG NH HD HD NI,
NI HD NI NN NG NN NN NG NI HD NI HD NI HD HD NG,
HD HD NI HD HD HD HD HD HD NI HD NG NI NI NN,
HD NI NG NG NN NN HD HD NN NN NN HD NI HD,
NN HD NG NG NN NI NI HD HD HD NI NN NN NI NN NI,
NI HD NI HD HD HD NN NI NG HD HD NI HD NG NN NN NN,
NN HD NG NN HD NI NG HD NI NI HD HD HD HD,
NN NN HD NI HD NN NI NI NI HD NI HD HD HD NG HD HD,
NN NN NG NN NN HD NG HD NI NG NN HD HD NG NN,
NN NI NG NG NG NN HD NI HD NI NN HD NG HD NI NG,
NI NI NH HD NG HD NG NH NI NH NH NI NH HD,
HD HD HD NG NI NK HD NG NH NG HD HD HD HD,
NH HD HD NG NI NH HD NI NG NH HD NG NI NH,
NI NG NH NH NH HD NG NG HD NI HD NH NH NI NG,
NH NI NI NI HD NG NI NG NH HD HD NG NH HD,

NH HD NI HD HD NI NG NG NH HD NG HD HD HD,
 NH NI HD NI NG NH HD NI NI HD NG HD NI NH,
 NI HD NI HD HD NI HD NG NI NH NH NH NH NG,
 NH NG HD NG NH HD NG NI NH NI HD NI NH NH,
 NH NH HD HD NG NI NH NI HD NI NH NH HD NG NH,
 NH NI NH NH HD NI NG NG HD NG NG NI NG HD NH,
 NN HD HD NG NN NN NI NI NI HD NN NG NG HD HD,
 NN NG NN HD NG HD NG NN NI HD NI NI NG NI,
 NN NG NG NG NG NN HD NI NN HD HD NG HD HD,
 NI HD NI NN HD NG NN NG NN NN NI NI HD NN NG,
 HD NI NI NN NI HD HD NN NI NN HD NI HD NG NN HD NG NN,
 HD NG NI NG HD HD HD NI NI NI NI HD NG HD NG,
 NH NI NI NI NI NI HD NG NI NG NH NG NI NG,
 NI NH NH HD NI NH NH HD NG NH NH NG NG NH NI,
 HD NI NI NG NI HD NI NI HD HD NI HD NN HD,
 NI NG NN NI HD NN NN NI HD NG HD NI NI HD NG,
 HD NI HD NI NI HD NI NG NG NG NN NG NI NI, and
 NI NG NG NG HD HD NI NN NG NN HD NI HD NI.

In embodiments, the TALE DBD comprises one or more of the sequences outlined herein or a variant sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

In embodiments, the GSHS and the TALE DBD sequences are selected from:

TGGCCGGCCTGACCACTGG (SEQ ID NO: 23) and NH NH HD HD NH NH HD HD NG NH NI HD HD NI HD NG NH
 NH;

TGAAGGCCTGGCCGGCCTG (SEQ ID NO: 24) and NH NI NI NH NH HD HD NG NH NH HD HD NH NH HD HD NG NH;

TGAGCACTGAAGGCCTGGC (SEQ ID NO: 25) and NH NI NH HD NI HD NG NH NI NI NH NH HD HD NG NH NH HD;

TCCACTGAGCACTGAAGGC (SEQ ID NO: 26) and HD HD NI HD NG NH NI NH HD NI HD NG NH NI NI NH NH HD;

TGGTTTCCACTGAGCACTG (SEQ ID NO: 27) and NH NH NG NG NG HD HD NI HD NG NH NI NH HD NI HD NG NH;

TGGGGAAAATGACCCAACA (SEQ ID NO: 28) and NH NH NH NH NI NI NI NI NG NH NI HD HD HD NI NI HD NI;

TAGGACAGTGGGGAAAATG (SEQ ID NO: 29) and NI NH NH NI HD NI NH NG NH NH NH NI NI NI NI NG NH;

TCCAGGGACACGGTGCTAG (SEQ ID NO: 30) and HD HD NI NH NH NH NI HD NI HD NH NH NG NH HD NG NI NH;

TCAGAGCCAGGAGTCCTGG (SEQ ID NO: 31) and HD NI NH NI NH HD HD NI NH NH NI NH NG HD HD NG NH NH;

TCCTTCAGAGCCAGGAGTC (SEQ ID NO: 32) and HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI NH NG HD;

TCCTCCTTCAGAGCCAGGA (SEQ ID NO: 33) and HD HD NG HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI;

TCCAGCCCCTCCTCCTTCA (SEQ ID NO: 34) and HD HD NI NH HD HD HD HD NG HD HD NG HD HD NG NG HD NI;

TCCGAGCTTGACCCCTTGA (SEQ ID NO: 35) and HD HD NH NI NH HD NG NG NH NI HD HD HD NG NG NH NH NI;

TGGTTTCCGAGCTTGACCC (SEQ ID NO: 36) and NH NH NG NG NG HD HD NH NI NH HD NG NG NH NI HD HD HD;

TGGGGTGGTTTCCGAGCTT (SEQ ID NO: 37) and NH NH NH NH NG NH NH NG NG NG HD HD NH NI NH HD NG NG;

TCTGCTGGGGTGGTTTCCG (SEQ ID NO: 38) and HD NG NH HD NG NH NH NH NH NG NH NH NG NG NG HD HD NH;

TGCAGAGTATCTGCTGGGG (SEQ ID NO: 39) and NH HD NI NH NI NH NG NI NG HD NG NH HD NG NH NH NH NH;

CCAATCCCCTCAGT (SEQ ID NO: 40) and HD HD NI NI NG HD HD HD HD NG HD NI NH NG;
CAGTGCTCAGTGGA (SEQ ID NO: 41) and HD NI NH NG NH HD NG HD NI NH NG NH NH NI NI;
GAAACATCCGGCGACTCA (SEQ ID NO: 42) and NH NI NI NI HD NI NG HD HD NH NH HD NH NI HD NG HD NI;
TCGCCCCCTCAAATCTTACA (SEQ ID NO: 43) and HD NH HD HD HD HD NG HD NI NI NI NG HD NG NG NI HD NI;
TCAAATCTTACAGCTGCTC (SEQ ID NO: 44) and HD NI NI NI NG HD NG NG NI HD NI NH HD NG NH HD NG HD;
TCTTACAGCTGCTCACTCC (SEQ ID NO: 45) and HD NG NG NI HD NI NH HD NG NH HD NG HD NI HD NG HD
HD;
TACAGCTGCTCACTCCOCT (SEQ ID NO: 46) and NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD HD HD
NG;
TGCTCACTCCOCTGCAGGG (SEQ ID NO: 47) and NH HD NG HD NI HD NG HD HD HD HD NG NH HD NI NH NH
NH;
TCCCCTGCAGGGCAACGCC (SEQ ID NO: 48) and HD HD HD HD NG NH HD NI NH NH NH HD NI NI HD NH HD
HD;
TGCAGGGCAACGCCAGGG (SEQ ID NO: 49) and NH HD NI NH NH NH HD NI NI HD NH HD HD HD NI NH NH
NH;
TCTCGATTATGGGCGGGAT (SEQ ID NO: 50) and HD NG HD NH NI NG NG NI NG NH NH NH HD NH NH NH NI
NG;
TCGCTTCTCGATTATGGGC (SEQ ID NO: 51) and HD NH HD NG NG HD NG HD NH NI NG NG NI NG NH NH NH
HD;
TGTCGAGTCGCTTCTCGAT (SEQ ID NO: 52) and NH NG HD NH NI NH NG HD NH HD NG NG HD NG HD NH NI
NG;
TCCATGTCGAGTCGCTTCT (SEQ ID NO: 53) and HD HD NI NG NH NG HD NH NI NH NG HD NH HD NG NG HD
NG;
TCGCCTCCATGTCGAGTCG (SEQ ID NO: 54) and HD NH HD HD NG HD HD NI NG NH NG HD NH NI NH NG HD
NH;
TCGTCATCGCCTCCATGTC (SEQ ID NO: 55) and HD NH NG HD NI NG HD NH HD HD NG HD HD NI NG NH NG
HD;

TGATCTCGTCATCGCCTCC (SEQ ID NO: 56) and NH NI NG HD NG HD NH NG HD NI NG HD NH HD HD NG HD HD;

GCTTCAGCTTCCTA (SEQ ID NO: 57) and NH HD NG NG HD NI NH HD NG NG HD HD NG NI;

CTGTGATCATGCCA (SEQ ID NO: 58) and HD NG NK NG NH NI NG HD NI NG NH HD HD NI;

ACAGTGGTACACACCT (SEQ ID NO: 59) and NI HD NI NN NG NN NN NG NI HD NI HD NI HD HD NG;

CCACCCCCCACTAAG (SEQ ID NO: 60) and HD HD NI HD HD HD HD HD HD NI HD NG NI NI NN;

CATTGGCCGGGCAC (SEQ ID NO: 61) and HD NI NG NG NN NN HD HD NN NN NN HD NI HD;

GCTTGAACCCAGGAGA (SEQ ID NO: 62) and NN HD NG NG NN NI NI HD HD HD NI NN NN NI NN NI;

ACACCCGATCCACTGGG (SEQ ID NO: 63) and NI HD NI HD HD HD NN NI NG HD HD NI HD NG NN NN NN;

GCTGCATCAACCCC (SEQ ID NO: 64) and NN HD NG NN HD NI NG HD NI NI HD HD HD HD;

GCCACAAACAGAAATA (SEQ ID NO: 65) and NN NN HD NI HD NN NI NI NI HD NI HD HD HD NG HD HD;

GGTGGCTCATGCCTG (SEQ ID NO: 66) and NN NN NG NN NN HD NG HD NI NG NN HD HD NG NN;

GATTTGCACAGCTCAT (SEQ ID NO: 67) and NN NI NG NG NG NN HD NI HD NI NN HD NG HD NI NG;

AAGCTCTGAGGAGCA (SEQ ID NO: 68) and NI NI NH HD NG HD NG NH NI NH NH NI NH HD;

CCCTAGCTGTCCC (SEQ ID NO: 69) and HD HD HD NG NI NK HD NG NH NG HD HD HD HD;

GCCTAGCATGCTAG (SEQ ID NO: 70) and NH HD HD NG NI NH HD NI NG NH HD NG NI NH;

ATGGGCTTCACGGAT (SEQ ID NO: 71) and NI NG NH NH NH HD NG NG HD NI HD NH NH NI NG;

GAACTATGCCTGC (SEQ ID NO: 72) and NH NI NI NI HD NG NI NG NH HD HD NG NH HD;

GCACCATTTGCTCCC (SEQ ID NO: 73) and NH HD NI HD HD NI NG NG NH HD NG HD HD HD;

GACATGCAACTCAG (SEQ ID NO: 74) and NH NI HD NI NG NH HD NI NI HD NG HD NI NH;

ACACCACTAGGGGT (SEQ ID NO: 75) and NI HD NI HD HD NI HD NG NI NH NH NH NH NG;

GTCTGCTAGACAGG (SEQ ID NO: 76) and NH NG HD NG NH HD NG NI NH NI HD NI NH NH;

GGCCTAGACAGGCTG (SEQ ID NO: 77) and NH NH HD HD NG NI NH NI HD NI NH NH HD NG NH;

GAGGCATTCTTATCG (SEQ ID NO: 78) and NH NI NH NH HD NI NG NG HD NG NG NI NG HD NH;

GCCTGGAAACGTTCC (SEQ ID NO: 79) and NN HD HD NG NN NN NI NI NI HD NN NG NG HD HD;

GTGCTCTGACAATA (SEQ ID NO: 80) and NN NG NN HD NG HD NG NN NI HD NI NI NG NI;

GTTTTGCAGCCTCC (SEQ ID NO: 81) and NN NG NG NG NG NN HD NI NN HD HD NG HD HD;
 ACAGCTGTGGAACGT (SEQ ID NO: 82) and NI HD NI NN HD NG NN NG NN NN NI NI HD NN NG;
 GGCTCTCTTCCTCCT (SEQ ID NO: 83) and HD NI NI NN NI HD HD NN NI NN HD NI HD NG NN HD NG NN;
 CTATCCCAAACTCT (SEQ ID NO: 84) and HD NG NI NG HD HD HD NI NI NI NI HD NG HD NG;
 GAAAACTATGTAT (SEQ ID NO: 85) and NH NI NI NI NI NI HD NG NI NG NH NG NI NG;
 AGGCAGCCTGGTTGA (SEQ ID NO: 86) and NI NH NH HD NI NH NH HD NG NH NH NG NG NH NI;
 CAATACAACCACGC (SEQ ID NO: 87) and HD NI NI NG NI HD NI NI HD HD NI HD NN HD;
 ATGACGGACTCAACT (SEQ ID NO: 88) and NI NG NN NI HD NN NN NI HD NG HD NI NI HD NG; and
 CACAACATTTGTAA (SEQ ID NO: 89) and HD NI HD NI NI HD NI NG NG NG NN NG NI NI.

In embodiments, the GSHS is within about 25, or about 50, or about 100, or about 150, or about 200, or about 300, or about 500 nucleotides of the TA dinucleotide site or TTAA (SEQ ID NO: 440) tetranucleotide site.

Illustrative DNA binding codes for human genomic safe harbor in areas of open chromatin via TALEs, encompassed by various embodiments are provided in **TABLE 4A-4F**. In embodiments, there is provided a variant of the TALEs, encompassed by various embodiments are provided in **TABLE 4A-4F**, e.g., having a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to any of the sequences in **TABLE 4A-4F**.

Illustrative DNA binding codes for human genomic safe harbor in areas of open chromatin via TALEs, encompassed by various embodiments are provided in **TABLE 4A**:

GSHS	ID	Sequence	TALE (DNA binding code)
AAVS1	1	TGGCCGGCCTGACCACTGG (SEQ ID NO: 23)	NH NH HD HD NH NH HD HD NG NH NI HD HD NI HD NG NH NH
AAVS1	2	TGAAGGCCTGGCCGGCCTG (SEQ ID NO: 24)	NH NI NI NH NH HD HD NG NH NH HD HD NH NH HD HD NG NH
AAVS1	3	TGAGCACTGAAGGCCTGGC (SEQ ID NO: 25)	NH NI NH HD NI HD NG NH NI NI NH NH HD HD NG NH NH HD
AAVS1	4	TCCACTGAGCACTGAAGGC (SEQ ID NO: 26)	HD HD NI HD NG NH NI NH HD NI HD NG NH NI NI NH NH HD
AAVS1	5	TGGTTTCCACTGAGCACTG (SEQ ID NO: 27)	NH NH NG NG NG HD HD NI HD NG NH NI NH HD NI HD NG NH
AAVS1	6	TGGGGAAAATGACCCAACA (SEQ ID NO: 28)	NH NH NH NH NI NI NI NI NG NH NI HD HD HD NI NI HD NI

AAVS1	7	TAGGACAGTGGGGAAAATG (SEQ ID NO: 29)	NI NH NH NI HD NI NH NG NH NH NH NH NI NI NI NI NG NH
AAVS1	8	TCCAGGGACACGGTGCTAG (SEQ ID NO: 30)	HD HD NI NH NH NH NI HD NI HD NH NH NG NH HD NG NI NH
AAVS1	9	TCAGAGCCAGGAGTCCTGG (SEQ ID NO: 31)	HD NI NH NI NH HD HD NI NH NH NI NH NG HD HD NG NH NH
AAVS1	10	TCCTTCAGAGCCAGGAGTC (SEQ ID NO: 32)	HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI NH NG HD
AAVS1	11	TCCTCCTTCAGAGCCAGGA (SEQ ID NO: 33)	HD HD NG HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI
AAVS1	12	TCCAGCCCCTCCTCCTTCA (SEQ ID NO: 34)	HD HD NI NH HD HD HD HD NG HD HD NG HD HD NG NG HD NI
AAVS1	13	TCCGAGCTTGACCCTTGGA (SEQ ID NO: 35)	HD HD NH NI NH HD NG NG NH NI HD HD HD NG NG NH NH NI
AAVS1	14	TGGTTTCCGAGCTTGACCC (SEQ ID NO: 36)	NH NH NG NG NG HD HD NH NI NH HD NG NG NH NI HD HD HD
AAVS1	15	TGGGGTGGTTTCCGAGCTT (SEQ ID NO: 37)	NH NH NH NH NG NH NH NG NG NG HD HD NH NI NH HD NG NG
AAVS1	16	TCTGCTGGGGTGGTTTCCG (SEQ ID NO: 38)	HD NG NH HD NG NH NH NH NH NG NH NH NG NG NG HD HD NH
AAVS1	17	TGCAGAGTATCTGCTGGGG (SEQ ID NO: 39)	NH HD NI NH NI NH NG NI NG HD NG NH HD NG NH NH NH NH
AAVS1	AVS1	CCAATCCCCTCAGT (SEQ ID NO: 40)	HD HD NI NI NG HD HD HD HD NG HD NI NH NG
AAVS1	AVS2	CAGTGCTCAGTGGA (SEQ ID NO: 41)	HD NI NH NG NH HD NG HD NI NH NG NH NH NI NI
AAVS1	AVS3	GAAACATCCGGCGACTCA (SEQ ID NO: 42)	NH NI NI NI HD NI NG HD HD NH NH HD NH NI HD NG HD NI
hROSA26	1F	TCGCCCTCAAATCTTACA (SEQ ID NO: 43)	HD NH HD HD HD HD NG HD NI NI NI NG HD NG NG NI HD NI
hROSA26	2F	TCAAATCTTACAGCTGCTC (SEQ ID NO: 44)	HD NI NI NI NG HD NG NG NI HD NI NH HD NG NH HD NG HD
hROSA26	3F	TCTTACAGCTGCTCACTCC (SEQ ID NO: 45)	HD NG NG NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD
hROSA26	4F	TACAGCTGCTCACTCCCCT (SEQ ID NO: 46)	NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD HD HD NG
hROSA26	5F	TGCTCACTCCCCTGCAGGG (SEQ ID NO: 47)	NH HD NG HD NI HD NG HD HD HD HD NG NH HD NI NH NH NH
hROSA26	6F	TCCCCTGCAGGGCAACGCC (SEQ ID NO: 48)	HD HD HD HD NG NH HD NI NH NH NH HD NI NI HD NH HD HD

hROSA26	7F	TGCAGGGCAACGCCAGGG (SEQ ID NO: 49)	NH HD NI NH NH NH HD NI NI HD NH HD HD HD NI NH NH NH
hROSA26	8R	TCTCGATTATGGGCGGGAT (SEQ ID NO: 50)	HD NG HD NH NI NG NG NI NG NH NH NH HD NH NH NH NI NG
hROSA26	9R	TCGCTTCTCGATTATGGGC (SEQ ID NO: 51)	HD NH HD NG NG HD NG HD NH NI NG NG NI NG NH NH NH HD
hROSA26	10R	TGTCGAGTCGCTTCTCGAT (SEQ ID NO: 52)	NH NG HD NH NI NH NG HD NH HD NG NG HD NG HD NH NI NG
hROSA26	11R	TCCATGTCGAGTCGCTTCT (SEQ ID NO: 53)	HD HD NI NG NH NG HD NH NI NH NG HD NH HD NG NG HD NG
hROSA26	12R	TCGCCTCCATGTCGAGTCG (SEQ ID NO: 54)	HD NH HD HD NG HD HD NI NG NH NG HD NH NI NH NG HD NH
hROSA26	13R	TCGTCATCGCCTCCATGTC (SEQ ID NO: 55)	HD NH NG HD NI NG HD NH HD HD NG HD HD NI NG NH NG HD
hROSA26	14R	TGATCTCGTCATCGCCTCC (SEQ ID NO: 56)	NH NI NG HD NG HD NH NG HD NI NG HD NH HD HD NG HD HD
hROSA26	ROSA1	GCTTCAGCTTCCTA (SEQ ID NO: 57)	NH HD NG NG HD NI NH HD NG NG HD HD NG NI
hROSA26	ROSA2	CTGTGATCATGCCA (SEQ ID NO: 58)	HD NG NK NG NH NI NG HD NI NG NH HD HD NI
hROSA26	TALER2	ACAGTGGTACACACCT (SEQ ID NO: 59)	NI HD NI NN NG NN NN NG NI HD NI HD NI HD HD NG
hROSA26	TALER3	CCACCCCCCACTAAG (SEQ ID NO: 60)	HD HD NI HD HD HD HD HD NI HD NG NI NI NN
hROSA26	TALER4	CATTGGCCGGGCAC (SEQ ID NO: 61)	HD NI NG NG NN NN HD HD NN NN NN HD NI HD
hROSA26	TALER5	GCTTGAACCCAGGAGA (SEQ ID NO: 62)	NN HD NG NG NN NI NI HD HD HD NI NN NN NI NN NI
CCR5	TALC3	ACACCCGATCCACTGGG (SEQ ID NO: 63)	NI HD NI HD HD HD NN NI NG HD HD NI HD NG NN NN NN
CCR5	TALC4	GCTGCATCAACCCC (SEQ ID NO: 64)	NN HD NG NN HD NI NG HD NI NI HD HD HD HD
CCR5	TALC5	GCCACAAACAGAAATA (SEQ ID NO: 65)	NN NN HD NI HD NN NI NI NI HD NI HD HD HD NG HD HD
CCR5	TALC7	GGTGGCTCATGCCTG (SEQ ID NO: 66)	NN NN NG NN NN HD NG HD NI NG NN HD HD NG NN
CCR5	TALC8	GATTTGCACAGCTCAT (SEQ ID NO: 67)	NN NI NG NG NG NN HD NI HD NI NN HD NG HD NI NG
Chr 2	SHCHR2-1	AAGCTCTGAGGAGCA (SEQ ID NO: 68)	NI NI NH HD NG HD NG NH NI NH NH NI NH HD

Chr 2	SHCHR2-2	CCCTAGCTGTCCC (SEQ ID NO: 69)	HD HD HD NG NI NK HD NG NH NG HD HD HD HD
Chr 2	SHCHR2-3	GCCTAGCATGCTAG (SEQ ID NO: 70)	NH HD HD NG NI NH HD NI NG NH HD NG NI NH
Chr 2	SHCHR2-4	ATGGGCTTCACGGAT (SEQ ID NO: 71)	NI NG NH NH NH HD NG NG HD NI HD NH NH NI NG
Chr 4	SHCHR4-1	GAAACTATGCCTGC (SEQ ID NO: 72)	NH NI NI NI HD NG NI NG NH HD HD NG NH HD
Chr 4	SHCHR4-2	GCACCATTGCTCCC (SEQ ID NO: 73)	NH HD NI HD HD NI NG NG NH HD NG HD HD HD
Chr 4	SHCHR4-3	GACATGCAACTCAG (SEQ ID NO: 74)	NH NI HD NI NG NH HD NI NI HD NG HD NI NH
Chr 6	SHCHR6-1	ACACCACTAGGGGT (SEQ ID NO: 75)	NI HD NI HD HD NI HD NG NI NH NH NH NH NG
Chr 6	SHCHR6-2	GTCTGCTAGACAGG (SEQ ID NO: 76)	NH NG HD NG NH HD NG NI NH NI HD NI NH NH
Chr 6	SHCHR6-3	GGCCTAGACAGGCTG (SEQ ID NO: 77)	NH NH HD HD NG NI NH NI HD NI NH NH HD NG NH
Chr 6	SHCHR6-4	GAGGCATTCTTATCG (SEQ ID NO: 78)	NH NI NH NH HD NI NG NG HD NG NG NI NG HD NH
Chr 10	SHCHR10-1	GCCTGGAACGTTCC (SEQ ID NO: 79)	NN HD HD NG NN NI NI NI HD NN NG NG HD HD
Chr 10	SHCHR10-2	GTGCTCTGACAATA (SEQ ID NO: 80)	NN NG NN HD NG HD NG NN NI HD NI NI NG NI
Chr 10	SHCHR10-3	GTTTTGCAGCCTCC (SEQ ID NO: 81)	NN NG NG NG NG NN HD NI NN HD HD NG HD HD
Chr 10	SHCHR10-4	ACAGCTGTGGAACGT (SEQ ID NO: 82)	NI HD NI NN HD NG NN NG NN NN NI NI HD NN NG
Chr 10	SHCHR10-5	GGCTCTCTCCTCCT (SEQ ID NO: 83)	HD NI NI NN NI HD HD NN NI NN HD NI HD NG NN HD NG NN
Chr 11	SHCHR11-1	CTATCCCAAACTCT (SEQ ID NO: 84)	HD NG NI NG HD HD HD NI NI NI NI HD NG HD NG
Chr 11	SHCHR11-2	GAAAACTATGTAT (SEQ ID NO: 85)	NH NI NI NI NI NI HD NG NI NG NH NG NI NG
Chr 11	SHCHR11-3	AGGCAGGCTGGTTGA (SEQ ID NO: 86)	NI NH NH HD NI NH NH HD NG NH NH NG NG NH NI
Chr 17	SHCHR17-1	CAATACAACCACGC (SEQ ID NO: 87)	HD NI NI NG NI HD NI NI HD HD NI HD NN HD
Chr 17	SHCHR17-2	ATGACGGACTCAACT (SEQ ID NO: 88)	NI NG NN NI HD NN NN NI HD NG HD NI NI HD NG

Chr 17	SHCHR17-3	CACAACATTTGTAA (SEQ ID NO: 89)	HD NI HD NI NI HD NI NG NG NG NN NG NI NI
Chr 17	SHCHR17-4	ATTTCCAGTGCACA (SEQ ID NO: 90)	NI NG NG NG HD HD NI NN NG NN HD NI HD NI

In embodiments, TALEs for targeting human genomic safe harbor sites using any of the TALE-based targeting elements to the TTAA site in hROSA26 (*e.g.*, hg38 chr3:9,396,133-9,396,305) are shown in **TABLE 4B**:

NAME	DNA SEQUENCE (SEQ ID NO:)	RVD AMINO ACID CODE
R1	TCGCCCTCCTCAATCTTACAG (584)	HD NH HD HD HD HD NG HD NI NI NI NG HD NG NG NI HD NI NH
R2	TCAAATCTTACAGCTGCTCA (585)	HD NI NI NI NG HD NG NG NI HD NI NH HD NG NH HD NG HD NI
R3	TCTTACAGCTGCTCACTCCC (586)	HD NG NG NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD HD
R4	TACAGCTGCTCACTCCCCTG (587)	NI HD NI NH HD NG NH HD NG HD NI HD NG HD HD HD HD NG NH
R5	TGCTCACTCCCCTGCAGGGC (588)	NH HD NG HD NI HD NG HD HD HD HD NG NH HD NI NH NH NH HD
R6	TCCCCTGCAGGGCAACGCC (456)	HD HD HD HD NG NH HD NI NH NH NH HD NI NI HD NH HD HD HD
R7	TGCAGGGCAACGCCAGGGA (589)	NH HD NI NH NH NH HD NI NI HD NH HD HD HD NI NH NH NH NI
R8	TCTCGATTATGGCGGGATT (590)	HD NG HD NH NI NG NG NI NG NH NH NH HD NH NH NH NI NG NG
R9	TCGCTTCTCGATTATGGGCG (591)	HD NH HD NG NG HD NG HD NH NI NG NG NI NG NH NH NH HD NH
R10	TGTCGAGTCGCTTCTCGATT (592)	NH NG HD NH NI NH NG HD NH HD NG NG HD NG HD NH NI NG NG
R11	TCCATGTCGAGTCGCTTCTC (593)	HD HD NI NG NH NG HD NH NI NH NG HD NH HD NG NG HD NG HD
R12	TCGCCCTCCATGTCGAGTCGC (594)	HD NH HD HD NG HD HD NI NG NH NG HD NH NI NH NG HD NH HD
R13	TCGTCACTCGCTCCATGTCG (595)	HD NH NG HD NI NG HD NH HD HD NG HD HD NI NG NH NG HD NH
R14	TGATCTCGTCATCGCTCCA (596)	NH NI NG HD NG HD NH NG HD NI NG HD NH HD HD NG HD HD NI

In embodiments, TALEs for targeting human genomic safe harbor sites using any of the TALE-based targeting elements to the AAVS1 (*e.g.*, hg38 chr19:55,112,851-55,113,324) are shown in **TABLE 4C**:

NAME	DNA SEQUENCE (SEQ ID NO:)	RVD AMINO ACID CODE
AAV1c	TGGCCGGCCTGACCACTGGG (597)	NH NH HD HD NH NH HD HD NG NH NI HD HD NI HD NG NH NH NH
AAV2c	TGAAGGCCTGGCCGGCCTGA (598)	NH NI NI NH NH HD HD NG NH NH HD HD NH NH HD HD NG NH NI
AAV3c	TGAGCACTGAAGGCCTGGCC (599)	NH NI NH HD NI HD NG NH NI NI NH NH HD HD NG NH NH HD HD
AAV4c	TCCACTGAGCACTGAAGGCC (600)	HD HD NI HD NG NH NI NH HD NI HD NG NH NI NI NH NH HD HD

AAV5c	TGGTTTCCACTGAGCACTGA (601)	NH NH NG NG NG HD HD NI HD NG NH NI NH HD NI HD NG NH NI
AAV6	TGGGGAAAATGACCCAACAG (602)	NH NH NH NH NI NI NI NI NG NH NI HD HD HD NI NI HD NI NH
AAV7	TAGGACAGTGGGGAAAATGA (603)	NI NH NH NI HD NI NH NG NH NH NH NI NI NI NI NG NH NI
AAV8	TCCAGGGACACGGTGCTAGG (604)	HD HD NI NH NH NH NI HD NI HD NH NH NG NH HD NG NI NH NH
AAV9	TCAGAGCCAGGAGTCCTGGC (605)	HD NI NH NI NH HD HD NI NH NH NI NH NG HD HD NG NH NH HD
AAV10	TCCTTCAGAGCCAGGAGTCC (606)	HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI NH NG HD HD
AAV11	TCCTCCTTCAGAGCCAGGAG (607)	HD HD NG HD HD NG NG HD NI NH NI NH HD HD NI NH NH NI NH
AAV12	TCCAGCCCTCCTCCTTCAG (608)	HD HD NI NH HD HD HD HD NG HD HD NG HD HD NG NG HD NI NH
AAV13c	TCCGAGCTTGACCCCTGGAA (462)	HD HD NH NI NH HD NG NG NH NI HD HD HD NG NG NH NH NI NI
AAV14c	TGGTTTCCGAGCTTGACCT (112)	NH NH NG NG NG HD HD NH NI NH HD NG NG NH NI HD HD HD NG
AAV15c	TGGGGTGGTTTCCGAGCTTG (609)	NH NH NH NH NG NH NH NG NG NG HD HD NH NI NH HD NG NG NH
AAV16c	TCTGCTGGGGTGGTTTCCGA (610)	HD NG NH HD NG NH NH NH NG NH NH NG NG NG HD HD NH NI
AAV17c	TGCAGAGTATCTGCTGGGGT (611)	NH HD NI NH NI NH NG NI NG HD NG NH HD NG NH NH NH NH NG

In embodiments, TALEs for targeting human genomic safe harbor sites using any of the TALE-based targeting elements to Chromosome 4 (e.g., hg38 chr4:30,793,534-30,875,476 or hg38 chr4:30,793,533-30,793,537 (9677); chr4:30,875,472-30,875,476 (8948)) are shown in **TABLE 4D**:

NAME	DNA SEQUENCE (SEQ ID NO:)	RVD AMINO ACID CODE
TALE4-R001	TCCTCCTAGTATTAAAGT (612)	HD NG NG HD HD NG NI NH NG NI NG NG NI NI NI NH NG
TALE4-R002	TCCTTAATATTACAGT (613)	HD HD NG NG NI NI NG NI NG NG NI HD HD NI NH NG
TALE4-F003	TACCAAGCTGAAATGACACAAAAGT (614)	NI HD HD NI NI NH HD NG NH NI NI NI NG NH NI HD NI HD NI NI NI NI NH NG
TALE4-F004	TGCTGTGTGCACATACCAGCAGAAT (615)	NH NH HD NG NH NG NH NG HD NI HD NI NG NI HD HD NI NH HD NI NH NI NI NG
TALE4-F005	TGTTAATTTGAATACAATCACT (616)	NH NG NG NI NI NG NG NG NH NI NI NG NI HD NI NI NG HD NI HD NG
TALE4-F006	TGTGTGCATACCAGCAGAAT (617)	NH NG NH NG HD NI HD NI NG NI HD HD NI NH HD NI NH NI NI NG
TALE4-R007	TGGTAACACTAATTT (618)	NH NH NG NI NI HD NG NI HD NG NI NI NG NG NG
TALE4-F008	TGTCACATACCAGCAGAAT (619)	NH NG HD NI HD NI NG NI HD HD NI NH HD NI NH NI NI NG
TALE4-R009	TGTGACACAGCCATCAACAAT (620)	NH NG NH NI HD NI HD NI NH HD HD NI NG HD NI NI HD NI NI NG
TALE4-F010	TCCTTTGATGAACAGT (621)	HD HD NG NG NG NH NI NG NH NI NI HD NI NH NG
TALE4-F011	TGTGTGCAATAGCGTTAAAGGAACAT (622)	NH NG NH NG NH HD NI NI NG NI NH HD NH NG NG NI NI NI NH NH NI NI HD NG NI HD NI NG
TALE4-F012	TCCTTCAATAGCCCACT (623)	HD NG NG NG HD NI NI NG NI NH HD HD HD NI HD NG

TALE4-R013	TCTCAAATGACAAGAGCACAGT (624)	HD NG HD NI NI NI NG NH NI HD NI NI NH NI NH HD NI HD NI NH NG
TALE4-F014	TACCAGTTAATTAGCACT (625)	NI HD HD NI NH NG NG NI NI NG NG NI NH HD NI HD NG
TALE4-F015	TGTTGTGACCTAAGCCAT (626)	NH NG NG NH NG NH NI HD HD NG NI NI NH HD HD NI NG
TALE4-R016	TCTCATGTTTTAAAGTCAAGAAT (627)	HD NG HD NI NG NH NG NG NG NG NI NI NI NH NG HD NI NI NH NI NI NG
TALE4-F017	TCCTGAATTGACAACAGAT (628)	HD HD NG NH NI NI NG NG HD NI NH NI NI HD NI NH NI NG
TALE4-F018	TAGCATGATGTTTCATGTTGACCT (629)	NI NH HD NI NG NH NI NG NH NG NG NG HD NI NG NH NG NG NH NG NH NI HD HD NG
TALE4-F019	TGTTTCATGTTGTGACCTAAGCCAT (630)	NH NG NG NG HD NI NG NH NG NG NH NG NH NI HD HD NG NI NI NH HD HD NI NG
TALE4-F020	TACAACAGTCTATTTTCAT (631)	NI HD NI NI HD NI NH NG HD NG NI NG NG NG HD NI NG

In embodiments, TALEs for targeting human genomic safe harbor sites using any of the TALE-based targeting elements to Chromosome 22 (e.g., hg38 chr22:35,370,000-35,380,000 or hg38 chr22:35,373,912-35,373,916 (861); chr22:35,377,843-35,377,847 (1153)) are shown in **TABLE 4E**:

NAME	DNA SEQUENCE (SEQ ID NO:)	RVD AMINO ACID CODE
TALE22F-R001	TCCTCCTAGTCCTCTCTACCCAGT (632)	HD NG NG HD HD NG NI NH NG HD NG HD NG NG HD NG HD NG NI HD HD HD NI NH NG
TALE22-F002	TACACTCCAGCCTGGGAACAGAGT (633)	NI HD NI HD NG HD HD NI NH HD HD NG NH NH NH NI NI NI HD NI NH NI NH NG
TALE22-F003	TCCTTCCTTAGGACGGCT (634)	HD NG NG NG NG HD HD NG NG NI NH NH NI HD NH NH HD NG
TALE22-F004	TCGCTCAGGCCTGTCAT (635)	HD NH HD NG HD NI NH NH HD HD NG NH NG HD NI NG
TALE22-F005	TCCATATGGAAGACTT (636)	HD HD NI NG NI NG NH NH NI NI NH NI HD NG NG
TALE22-F006	TACCCAGTTAACCACCCT (637)	NI HD HD HD NI NH NG NG NI NI HD HD NI HD HD HD NG
TALE22-F007	TGGCGCATGCCCTGTAATCCAGCTACT (638)	NH NH HD NH HD NI NG NH HD HD NG NH NG NI NI NG HD HD HD NI NH HD NG NI HD NG
TALE22-F008	TATACGAGGAGAAAATTAGCATTCCT (639)	NI NG NI HD NH NI NH NH NI NH NI NI NI NI NG NG NI NH HD NI NG NG HD HD NG
TALE22-R009	TCTGCCTCCAGGTTCAAGCAAT (640)	HD NG NH HD HD NG HD HD HD NI NH NH NG NG HD NI HD NH HD NI NI NG
TALE22-F010	TGCCITGTACAGTTTTCACAGT (641)	NH HD HD NG NG NH NG HD NI HD NH NG NG NG NG HD NI HD NI NH NG
TALE22-F001A	TGTCACCTTCTGTATGTGCAACCAT (642)	NH NG HD NI HD HD NG NG HD NG NH NG NI NG NH NG NH HD NI NI HD HD NI NG
TALE22-F002A	TCTGTATGTGCAACCAT (643)	HD NG NH NG NI NG NH NG NH HD NI NI HD HD NI NG
TALE22-R03A	TAGTCAAGCAACAGGAT (644)	NI NH NG HD NI NI NH HD NI NI HD NI NH NH NI NG
TALE22-F004A	TCCAAGATAATTCCCCAT (645)	HD HD NI NI NH NI NG NI NI NG NG HD HD HD HD NI NG
TALE22-F005A	TCTGCAAGATCCTTTT (646)	HD NG NH HD NI NI NH NI NG HD HD NG NG NG NG
TALE22-F006A	TGCTATGTAAGGTACCAAAAAGGTAACCT (647)	NH HD NG NI NG NH NG NI NI NH NH NG NI NH HD NI NI NI NI NI NH NH NG NI NI HD HD NG

TALE22-R007A	TCTCTCTCCTCCTGCT (648)	HD NG HD NG HD NG HD HD NG HD HD NG NH HD NG
TALE22-R008A	PCCAATGCTATTCTCTCT (649)	HD HD NI NI NI NG NH HD NG NI NG NG HD NG HD NG HD NG
TALE22-R009A	TGCTGATTCAGCCTCCT (650)	NH HD NG NH NI NG NG HD NI NH HD HD NG HD HD NG
TALE22-F010A	TAGAACAGCCCCACACAGT (651)	NI NH NI NI HD NI NH HD HD HD HD HD HD NI HD NI HD NI NH NG

In embodiments, TALEs for targeting human genomic safe harbor sites using any of the TALE-based targeting elements to Chromosome X (e.g., hg38 chrX:134,419,661-134,541,172 or hg38 chrX:134,476,304-134,476,307 (85); chrX:134,476,337-134,476,340 (51)) are shown in **TABLE 4F**:

NAME	DNA SEQUENCE (SEQ ID NO:)	RVD AMINO ACID CODE
TALE F002	TTTAGCAGATGCATCAGC (652)	NG NG NI NH HD NI NH NI NG NH HD NI NG HD NI NH HD
TALE F003	TGACCAGGGGCATGTCTGG (653)	NH NI HD HD NI NH NH NH NH HD NI NG NH NG HD HD NG NH NH
TALE F004	FGGTCCACCTACCTGAAAATG (654)	HD NI NI NH NH NI NH NG NG HD NG NH NH HD NG NH NH NH NG HD
TALE F007	TGTCCACAGGTATTACGGGC (655)	NH NG HD HD HD NI HD NI NH NH NG NI NG NG NI HD NH NH NH HD
TALE F008	TACGGGCCAACCTGACAATAC (656)	NI HD NH NH NH HD HD NI NI HD HD NG NH NI HD NI NI NG NI HD
TALE F009	TGAGCTTTGGGGACTGAAAGA (657)	NH NI NH HD NG NG NG NH NH NH NH NI HD NG NH NI NI NI NH NI
TALE R002	CTGGCATAATCTTTTCCCCCA (658)	NH NH NH NH NH NI NI NI NI NH NI NG NG NI NG NH HD HD NI NH
TALE R003	CCAGCTCCTGGCCATGTGCA (659)	NH HD NI HD NI NG NH NH HD HD NI NH NH NI NH NH HD NG NH NH
TALE R004	GGCCATGTGCACAGGGGCTGA (660)	HD NI NH HD HD HD HD NG NH NG NH HD NI HD NI NG NH NH HD HD
TALE R005	CTGATATGTGAAGGTTTAGCA (661)	NH HD NG NI NI NI HD HD NG NG HD NI HD NI NG NI NG HD NI NH
TALE R007	TCACCAGGCGTGGTGGCTCAC (662)	NH NI HD HD NI NH NH HD NH NG NH NH NG NH NH HD NG HD NI HD
TALE F020*	TATAGACATTTTCACT (663)	NI NG NI NH NI HD NI NG NG NG NG HD NI HD NG
TALE F021*	TCTACATTTAACTATCAACCT (664)	HD NG NI HD NI NG NG NG NI NI HD NG NI NG HD NI NI HD HD NG
TALE F030*	TCGTGCAAACGTTTGAT (665)	HD NH NG NH HD NI NI NI HD NH NG NG NG NH NI NG
TALE F031*	TACATCAATCCTGTAGGT* (666)	NI HD NI NG HD NI NI NG HD HD NG NH NG NI NH NH NG
TALE F034*	TCTATTTTAGTGCACCAAGT (667)	HD NG NI NG NG NG NG NI NH NG NH NI HD HD HD NI NI NH NG
TALE F036*	TAGAGTCAAAGCATGTACT (668)	NI NH NI NH NG HD NI NI NI NH HD NI NG NH NG NI HD NG
TALE F037*	TCCTACCCATAAGCTCCT (669)	HD HD NG NI HD HD HD NI NG NI NI NH HD NG HD HD NG
TALE F040*	TCCCCATCCCCATCAGT (670)	HD HD HD HD NI NG HD HD HD HD NI NG HD NI NH NG
TALE R022*	TCTTTAATTCAGCAAGACTTTAACAAGT (671)	HD NG NG NG NI NI NG NG HD NI NI NH HD NI NI NH NI HD NG NG NG NI NI HD NI NI NH NG
TALE R033*	TGCAGTCCCCTTTCT (672)	NH HD NI NH NG HD HD HD HD NG NG NG HD NG NG
TALE R035*	TCTGCACAAATCCCCAAGAT (673)	HD NG NH HD NI HD NI NI NI NG HD HD HD HD NI NI NI NH NI NG
TALE R038*	TACATGCTTTGACTCT (674)	NI HD NI NG NH HD NG NG NG NH NI HD NG HD NG

TALE R039*	TGGCCAGTTATACTGCCAGCAGCTATAAT (675)	NH NH HD HD NI NH NG NG NI NG NI HD NG NH HD HD NI NH HD NI NH HD NG NI NG NI NI NG
------------	--	--

In embodiments, the mobile element enzyme is capable of inserting a donor DNA at a TA dinucleotide site. In embodiments, the mobile element enzyme is capable of inserting a donor DNA at a TTAA (SEQ ID NO: 440) tetranucleotide site.

Illustrative DNA binding codes for human genomic safe harbor in areas of open chromatin via ZNFs, encompassed by various embodiments are provided in **TABLE 5A-5E**. In embodiments, there is provided a variant of the ZNFs, encompassed by various embodiments are provided in **TABLE 5A-5E**, e.g., having a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity to any of the sequences in **TABLE 5A-5E**.

In embodiments, ZNFs for targeting human genomic safe harbor sites using any of the ZNF-based targeting elements to the TTAA site in hROSA26 (e.g., hg38 chr3:9,396,133-9,396,305) are shown in **TABLE 5A**:

hROSA 26 TTAA	NAME	TARGET (SEQ ID NO: __)	SCO RE	ZFP AMINO ACID CODE (SEQ ID NO: __)
5'	ZnF3 a	TGG GAA GAT AAA CTA (676)	58. 64	LEPGKPYKCECGKSFQNMSTLTHQRTHTGKPYKCECGKSFQSRANLRAHQ RTHTGKPYKCECGKSFSTSGNLVRHQRTHTGKPYKCECGKSFQSSNLVRH QRTHTGERPYKCECGKSFSSRSEHLTHQRTHTGKXTS (677)
5'	ZnF5 a	ACT CCC CTG CAG GGC AAC (678)	56. 25	LEPGERPYKCECGKSFSDSGNLVRHQRTHTGERPYKCECGKSFSDPGHLVRHQ RTHTGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSPADNLTEH QRTHTGKPYKCECGKSFSSKXHLADHQRTHTGKPYKCECGKSFSTHLDLIR HQRTHTGKXTS (679)
5'	ZnF5 b	CCC CTG CAG GGC AAC GCC (680)	56. 25	LEPGKPYKCECGKSFSDCRDLARHQRTHTGKPYKCECGKSFSDSGNLVRHQ RTHTGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSPADNLTEH QRTHTGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSSKXHLAE HQRTHTGKXTS (681)
5'	ZnF5 c	CTG CAG GGC AAC GCC CAG (682)	60. 58	LEPGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSDCRDLARHQ RTHTGKPYKCECGKSFSDSGNLVRHQRTHTGKPYKCECGKSFSDPGHLVRH QRTHTGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSPADNLTE HQRTHTGKXTS (683)
5'	ZnF5 d	CAG GGC AAC GCC CAG GGA (684)	58. 08	LEPGKPYKCECGKSFQRAHLERHQRTHTGKPYKCECGKSFSPADNLTEHQ RTHTGKPYKCECGKSFSDCRDLARHQRTHTGKPYKCECGKSFSDSGNLVRH QRTHTGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSPADNLTE HQRTHTGKXTS (685)
5'	ZnF5 e	GGC AAC GCC CAG GGA CCA (686)	57. 32	LEPGKPYKCECGKSFSTSHSLTEHQRTHTGKPYKCECGKSFQRAHLERHQ RTHTGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSDCRDLARH QRTHTGERPYKCECGKSFSDSGNLVRHQRTHTGKPYKCECGKSFSDPGHLVR HQRTHTGKXTS (687)
5'	ZnF5 f	AAC GCC CAG GGA CCA AGT (688)	54. 99	LEPGKPYKCECGKSFSHRTTLTHQRTHTGKPYKCECGKSFSTSHSLTEHQ RTHTGKPYKCECGKSFQRAHLERHQRTHTGKPYKCECGKSFSPADNLTEH QRTHTGKPYKCECGKSFSDCRDLARHQRTHTGKPYKCECGKSFSDSGNLVR HQRTHTGKXTS (689)
5'	ZnF5 g	GCC CAG GGA CCA AGT TAG (690)	55. 31	LEPGKPYKCECGKSFSDNLHTEHQRTHTGKPYKCECGKSFSHRTTLTHQ RTHTGKPYKCECGKSFSTSHSLTEHQRTHTGKPYKCECGKSFQRAHLERH QRTHTGKPYKCECGKSFSPADNLTEHQRTHTGKPYKCECGKSFSDCRDLAR HQRTHTGKXTS (691)
5'	ZnF5 h	CAG GGA CCA AGT TAG CCC (692)	50. 76	LEPGKPYKCECGKSFSSKXHLAEHQRTHTGKPYKCECGKSFSDNLHTEHQ RTHTGKPYKCECGKSFSHRTTLTHQRTHTGKPYKCECGKSFSTSHSLTEH QRTHTGKPYKCECGKSFQRAHLERHQRTHTGKPYKCECGKSFSPADNLTE HQRTHTGKXTS (693)
3'	ZnF1 2a	GCC TAG GCA AAA GAA (694)	59. 09	LEPGKPYKCECGKSFQSSNLVFEHQRTHTGKPYKCECGKSFQSRANLRAHQ RTHTGKPYKCECGKSFQSGDLARHQRTHTGKPYKCECGKSFSPADNLTEH QRTHTGKPYKCECGKSFSDCRDLARHQRTHTGKXTS (695)

3'	ZnF1 3a	CGC GAG GAG GAA AGG AGG (696)	57. 19	LEPGEKPYKCECGKSFSRSDHLTNHQRTHTGKPYKCECGKSFSRSDHLTNHQ RTHTGKPYKCECGKSFSQSSNLVRHQRTHTGKPYKCECGKSFSRSDNLVRH QRTHTGERPYKCECGKSFSRSDNLVRHQRTHTGERPYKCECGKSFSHTGHLLE HQRTHTGKTS (697)
3'	ZnF1 3b	GAG GAG GAA AGG AGG GAG (698)	57. 80	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECGKSFSRSDHLTNHQ RTHTGKPYKCECGKSFSRSDHLTNHQRTHTGKPYKCECGKSFSQSSNLVRH QRTHTGERPYKCECGKSFSRSDNLVRHQRTHTGERPYKCECGKSFSRSDNLVR HQRTHTGKTS (699)
3'	ZnF1 3c	GAG GAA AGG AGG GAG GGC (700)	57. 61	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSRSDHLTNHQRTHTGKPYKCECGKSFSRSDHLTNH QRTHTGERPYKCECGKSFSQSSNLVRHQRTHTGKPYKCECGKSFSRSDNLVR HQRTHTGKTS (701)

In embodiments, ZNFs for targeting human genomic safe harbor sites using any of the ZNF-based targeting elements to the AAVS1 (e.g., hg38 chr19:55,112,851-55,113,324) are shown in TABLE 5B:

AAVS1 TTAA	NAME	TARGET (SEQ ID NO: _)	SCO RE	ZFP AMINO ACID CODE (SEQ ID NO: _)
5'	ZnF1 1a	TAG GAC AGT GGG GAA AAT GAC CCA ACA GCC (702)	57. 08	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSPPADLTHQRTHTGKPYKCECGKSFSSTSHSLTEHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSSTGNLTVHQRTHTGKPYKCECGKSFSQSSNLVRHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSHRTTTLNKHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKTS (703)
5'	ZnF1 0a	AGA GGG AGC CAC GAA AAC AGA (704)	56. 91	LEPGEKPYKCECGKSFSQSLAHLRAHQRTHTGKPYKCECG KSFSRSDNLVRHQRTHTGKPYKCECGKSFSQSSNLVRHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRSHLREHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSQSLAHLRAHQRTHTGKTS (705)
3'	ZnF1 2b	GCA GAT AGC CAG GAG (706)	59. 97	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSSTGNLVRHQRTHTGKPYKCECG KSFSQSGDLRRHQRTHTGKTS (707)
3'	ZnF1 3b	AGA TAG CCA GGA GTC CTT (708)	56. 80	LEPGEKPYKCECGKSFSSTGALTEHQRTHTGKPYKCECG KSFSRSDNLVRHQRTHTGKPYKCECGKSFSRSHLREHQ RTHTGKPYKCECGKSFSSTSHSLTEHQRTHTGKPYKCECG KSFSRSDNLTHQRTHTGKPYKCECGKSFSQSLAHLRAHQ RTHGKTS (709)
5'	ZnF1 4a	CCC AGT GGT CAG GCC GGC CAG GCC (710)	61. 78	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSSTSHSLTEHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRSDNLTHQRTHTGKPYKCECGKSFSQSLAHLRAHQ RTHGKTS (711)
5'	ZnF1 5a	GGC CGG CCA GGC CTT CAG (712)	58. 15	LEPGEKPYKCECGKSFSRADNLTEHQRTHTGKPYKCECG KSFSSTGALTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSSTSHSLTEHQRTHTGKPYKCECG KSFSRSDNLTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHGKTS (713)
5'	ZnF1 6a	AGT GCT CAG TGG AAA CCA CGA AAG GAC (714)	58. 65	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSQSGHLTEHQ RTHTGKPYKCECGKSFSSTSHSLTEHQRTHTGKPYKCECG KSFSQSLAHLRAHQRTHTGKPYKCECGKSFSRSDNLTHQ RTHTGKPYKCECGKSFSRADNLTEHQRTHTGKPYKCECG KSFSSTSGELVRHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHGKTS (715)
5'	ZnF1 7a	TGG CCC CCA GCC CCT CCT GCC (716)	60. 89	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSSTGNLTVHQRTHTGKPYKCECGKSFSSTGNLTVHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSSTGNLTVHQRTHTGKPYKCECGKSFSSTGNLTVHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKTS (717)
5'	ZnF1 8a	AGA GCC AGG AGT CCT GGC CCC CAG CCC (718)	57. 23	LEPGEKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHTGKPYKCECGKSFSRSDNLVRHQRTHTGKPYKCECG KSFSRADNLTEHQRTHTGKPYKCECGKSFSRSDNLVRHQ RTHGKTS (719)

				KSFSTKNSLTHQRTHTGKPYKCECGKSFSSHRTTLTNHQRTHTGKPYKCECGKSFSSRDHLTNHQRTHTGKPYKCECGKSFSDCDRLARHQRTHTGKPYKCECGKSFSQLAHLRAHQRTHTGKETS (719)
3'	ZnF1 9a	GCA GGA GGG GCT GGG GGC CAG GAC (720)	59. 93	LEPGKPYKCECGKSFSDPGNLVRHQRTHTGKPYKCECGKSFSSRADNLTHQRTHTGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSSDKLVRHQRTHTGKPYKCECGKSFSTGGELVRHQRTHTGKPYKCECGKSFSSDKLVRHQRTHTGKPYKCECGKSFSQLAHLERHQRTHTGKPYKCECGKSFSSQSGDLRRKQRTHTGKETS (721)
3'	ZnF2 0b	ATA GCC CTG GGC CCA CGG CTT CGT (722)	59. 53	LEPGKPYKCECGKSFSSRRRTCRARHQRTHTGKPYKCECGKSFSTGALTEHQRTHTGKPYKCECGKSFSSDKLTHQRTHTGKPYKCECGKSFSTSHSLTEHQRTHTGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSSRDALTEHQRTHTGKPYKCECGKSFSDCDRLARHQRTHTGKPYKCECGKSFSSQSSSLAHLQRTHTGKETS (723)
3'	ZnF2 1b	GAA GGA CCT GGC TGG (724)	55. 22	LEPGKPYKCECGKSFSSRDHLTHQRTHTGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSTKNSLTHQRTHTGKPYKCECGKSFSQLAHLERHQRTHTGKPYKCECGKSFSSQSSNLVRHQRTHTGKETS (725)
5'	ZnF2 2a	GCA GGA ACG AAG CCG TGG GCC CAG GGC (726)	56. 47	LEPGKPYKCECGKSFSDPGHLVRHQRTHTGKPYKCECGKSFSSRADNLTHQRTHTGKPYKCECGKSFSDCDRLARHQRTHTGKPYKCECGKSFSSRDHLTHQRTHTGKPYKCECGKSFSSRDNLTEHQRTHTGKPYKCECGKSFSSRDNLKLNHQRTHTGKPYKCECGKSFSSRDNLTEHQRTHTGKPYKCECGKSFSSQRAHLERHQRTHTGKPYKCECGKSFSSQSGDLRRHQRTHTGKETS (727)
5'	ZnF2 3a	GGA AAC CAC CCC AGC AGA (728)	52. 63	LEPGKPYKCECGKSFSQLAHLRAHQRTHTGKPYKCECGKSFSSERSHLREHQRTHTGKPYKCECGKSFSSKKALTEHQRTHTGKPYKCECGKSFSDSGNLVRHQRTHTGKPYKCECGKSFSQLAHLERHQRTHTGKETS (729)
5'	ZnF2 4a	AAG GGT CAA GCT CGG AAA CCA CCC CAG CAG ATA (730)	55. 09	LEPGKPYKCECGKSFSSQSSSLAHLQRTHTGKPYKCECGKSFSSPADNLTEHQRTHTGKPYKCECGKSFSSPADNLTEHQRTHTGKPYKCECGKSFSSKKHLAHLQRTHTGKPYKCECGKSFSTSHSLTEHQRTHTGKPYKCECGKSFSSQANLRAHQRTHTGKPYKCECGKSFSSDKLTHQRTHTGKPYKCECGKSFSTGGELVRHQRTHTGKPYKCECGKSFSSQSGNLTEHQRTHTGKPYKCECGKSFSSGHLVRHQRTHTGKPYKCECGKSFSSRDNLKLNHQRTHTGKETS (731)

In embodiments, ZNFs for targeting human genomic safe harbor sites using any of the ZNF-based targeting elements to Chromosome 4 (e.g., hg38 chr4:30,793,534-30,875,476 or hg38 chr4:30,793,533-30,793,537 (9677); chr4:30,875,472-30,875,476 (8948)) are shown in **TABLE 5C**:

Chr4 TTAA	NAME	TARGET (SEQ ID NO: __)	SCORE	ZFP AMINO ACID CODE (SEQ ID NO: __)
5'	ZnF3 1F	CTTTGATGAACAGTCACA (732)	58. 41	LEPGKPYKCECGKSFSSPADLTHQRTHTGKPYKCECGKSFSDPGALVRHQRTHTGKPYKCECGKSFSSPADLTHQRTHTGKPYKCECGKSFSSQAGHLASHQRTHTGKPYKCECGKSFSTGALTEHQRTHTGKETS (733)
5'	ZnF3 2F	CTTCCAATTAGTCCTACC (734)	55. 84	LEPGKPYKCECGKSFSSDKDLTRHQRTHTGKPYKCECGKSFSTKNSLTHQRTHTGKPYKCECGKSFSSHRTTLTNHQRTHTGKPYKCECGKSFSSKNALQNRQRTHTGKPYKCECGKSFSTGKSLTEHQRTHTGKPYKCECGKSFSTGALTEHQRTHTGKETS (735)
5'	ZnF3 3F	ATACTAGGAAGAAATACAATA (736)	57. 27	LEPGKPYKCECGKSFSSQSSSLAHLQRTHTGKPYKCECGKSFSSPADLTHQRTHTGKPYKCECGKSFSTGNLTVHQRTHTGKPYKCECGKSFSQLAHLRAHQRTHTGKPYKCECGKSFSQLAHLERHQRTHTGKPYKCECGKSFSSQNSLTHQRTHTGKPYKCECGKSFSSQSSSLAHLQRTHTGKETS (737)
5'	ZnF3 4F	GCTCTTGTCATTGAGAT (738)	57. 38	LEPGKPYKCECGKSFSSGNLVRHQRTHTGKPYKCECGKSFSSQAGHLASHQRTHTGKPYKCECGKSFSSKNALQNRQRTHTGKETS (739)

				THTGEKPYKCPCECGKSFSDPGALVRHQRTHTGEKPYKCPCEGKSFSTTGALTEHQRTHTGEKPYKCPCECGKSFSTSGELVRHQRTHTGKKTS (739)
5'	ZnF3 5F	CCAAGCTGAAATGACACAAAAGTTAA ACAAAG (740)	58. 23	LEPGEKPYKCPCECGKSFSDPGNLVRHQRTHTGEKPYKCPCEGKSFSSPADLTRHQRTHTGEKPYKCPCECGKSFSSQANLRAHQRTHTGEKPYKCPCECGKSFSTSGSLVRHQRTHTGEKPYKCPCEGKSFSSQANLRAHQRTHTGEKPYKCPCECGKSFSSPADLTRHQRTHTGEKPYKCPCECGKSFSDPGNLVRHQRTHTGEKPYKCPCEGKSFSTTGNLTVHQRTHTGEKPYKCPCECGKSFSSQAGHLASHQRTHTGEKPYKCPCECGKSFSSERSHLREHQRTHTGEKPYKCPCEGKSFSTSHSLTEHQRTHTGKKTS (741)
5'	ZnF3 6F	CTTATACCAGTTAATTAGCAC (742)	49. 93	LEPGEKPYKCPCECGKSFSSKKALTEHQRTHTGEKPYKCPCEGKSFSSREDNLNTHQRTHTGEKPYKCPCECGKSFSTTGNLTVHQRTHTGEKPYKCPCECGKSFSTSGSLVRHQRTHTGEKPYKCPCEGKSFSTSHSLTEHQRTHTGEKPYKCPCECGKSFSSQKSSLIAHQRTHTGEKPYKCPCECGKSFSTTGALTEHQRTHTGKKTS (743)
3'	ZnF3 7R	AACGCTATTGCACACATAGTTACA (744)	57. 67	LEPGEKPYKCPCECGKSFSSPADLTRHQRTHTGEKPYKCPCEGKSFSTSGSLVRHQRTHTGEKPYKCPCECGKSFSSQKSSLIAHQRTHTGEKPYKCPCECGKSFSSKKALTEHQRTHTGEKPYKCPCEGKSFSSQSGDLRRHQRTHTGEKPYKCPCECGKSFSSHKNALQNHQRTHTGEKPYKCPCECGKSFSTSGELVRHQRTHTGEKPYKCPCEGKSFSDSGNLVHQRTHTGKKTS (745)
3'	ZnF3 8R	TGAATTCAGGAACAAAGTATA (746)	53. 21	LEPGEKPYKCPCECGKSFSSQKSSLIAHQRTHTGEKPYKCPCEGKSFSSHRTTLTNHQRTHTGEKPYKCPCECGKSFSSQSGNLTEHQRTHTGEKPYKCPCECGKSFSSSNLVRHQRTHTGEKPYKCPCEGKSFSSRADNLTEHQRTHTGEKPYKCPCECGKSFSSHKNALQNHQRTHTGEKPYKCPCECGKSFSSQAGHLASHQRTHTGKKTS (747)
3'	ZnF3 9R	GCTGGTATGTGACACAGCCATCAACA (748)	50. 63	LEPGEKPYKCPCECGKSFSSQSGNLTEHQRTHTGEKPYKCPCEGKSFSSQSGNLTEHQRTHTGEKPYKCPCECGKSFSTSGNLTEHQRTHTGEKPYKCPCECGKSFSSERSHLREHQRTHTGEKPYKCPCEGKSFSSKKALTEHQRTHTGEKPYKCPCECGKSFSSQAGHLASHQRTHTGEKPYKCPCECGKSFSSRRDELNVHQRTHTGEKPYKCPCEGKSFSTSGHLVRHQRTHTGEKPYKCPCECGKSFSTSGELVRHQRTHTGKKTS (749)

In embodiments, ZNFs for targeting human genomic safe harbor sites using any of the ZNF-based targeting elements to Chromosome 22 (e.g., hg38 chr22:35,370,000-35,380,000 or hg38 chr22:35,373,912-35,373,916 (861); chr22:35,377,843-35,377,847 (1153)) are shown in **TABLE 5D**:

Chr 22 TTA A	NAME	TARGET (SEQ ID NO:_)	SCO RE	ZFP (SEQ ID NO:_)
5'	ZnFla	CTTCCTGAAAGCAAGAGAT GAAAT (750)	57. 34	LEPGGKPYKCPCEGKSFSTTGNLTVHQRTHTGKPYKCPCEGKSFSSQAGHLASH QRTHTGKPYKCPCEGKSFSQLAHLRAHQRTHTGERPYKCPCEGKSFSSRKDNLK NHQRTHTGKPYKCPCEGKSFSSERSHLREHQRTHTGKPYKCPCEGKSFSSQSSN LVRHQRTHTGKPYKCPCEGKSFSTKNSLTEHQRTHTGKPYKCPCEGKSFSTT GALTEHQRTHTGKTS (751)
5'	ZnFlb	CTGAAAGCAAGAGATGAA TTCCA (752)	58. 92	LEPGKPYKCPCEGKSFSTSHSLTEHQRTHTGKPYKCPCEGKSFSSHKNALQNH QRTHTGKPYKCPCEGKSFSSQSSNLVRHQRTHTGERPYKCPCEGKSFSTSSNLV RHQRTHTGKPYKCPCEGKSFSQLAHLRAHQRTHTGKPYKCPCEGKSFSSQSGD LRHQRTHTGKPYKCPCEGKSFSSQRANLRAHQRTHTGKPYKCPCEGKSFSSRN DALTEHQRTHTGKTS (753)
5'	ZnF2a	ATACGAGGAGAAAAATTAGC AT (754)	51. 25	LEPGKPYKCPCEGKSFSTSGNLTEHQRTHTGKPYKCPCEGKSFSSREDNLHTH QRTHTGKPYKCPCEGKSFSTTGNLTVHQRTHTGKPYKCPCEGKSFSSQSSNLV RHQRTHTGKPYKCPCEGKSFSSORAHLEHQRTHTGKPYKCPCEGKSFSSQSGH LTEHQRTHTGKPYKCPCEGKSFSSQSSLIAHQRTHTGKTS (755)
5'	ZnF3a	CATCCATGGCAGGAAGTTG AAGCAAAATAAATCTG (756)	58. 67	LEPGKPYKCPCEGKSFSSRDALTEHQRTHTGKPYKCPCEGKSFSTTGNLTVH QRTHTGKPYKCPCEGKSFSSQSSSLIAHQRTHTGKPYKCPCEGKSFSSQANLR AHQRTHTGKPYKCPCEGKSFSSCDRLARHQRTHTGKPYKCPCEGKSFSSQSSN LVRHQRTHTGKPYKCPCEGKSFSTSGSLVRHQRTHTGKPYKCPCEGKSFSSQS SNLVRHQRTHTGERPYKCPCEGKSFSSADNLTEHQRTHTGKPYKCPCEGKSFSS RSDHLTHQRTHTGKPYKCPCEGKSFSTSHSLTEHQRTHTGKPYKCPCEGKSFSS FSTGNLTEHQRTHTGKTS (757)

5'	ZnF3b	ATGGCAGGAAGTTGAAGCC AAAATAAA (758)	54. 14	LEPGKPYKCECGKSFSSQANLRAHQRTHTGKPYKCECGKSFSTGNLTVH QRTHGTGKPYKCECGKSFSSQANLRAHQRTHTGKPYKCECGKSFSSERSHLR EHQRTHTGKPYKCECGKSFSSQAGHLASHQRTHTGKPYKCECGKSFSSHRTT LTNHQRTHTGKPYKCECGKSFSSQRAHLERHQRTHTGKPYKCECGKSFSSQS GDLRRHQRTHTGKPYKCECGKSFSSRRDELNVHQRTHTGKKTIS (759)
3'	ZnF5a R	GAAAAGAAGACTCAAGGAA ACAGAGCCAAACAC (760)	55. 40	LEPGKPYKCECGKSFSSSKKALTEHQRTHTGKPYKCECGKSFSSQANLRAH QRTHGTGKPYKCECGKSFSSDCRDLARHQRTHTGKPYKCECGKSFSSQLAHLR AHQRTHTGKPYKCECGKSFSSSGNLRVHQRTHTGKPYKCECGKSFSSQRAH LERHQRTHTGKPYKCECGKSFSSQSGNLTEHQRTHTGKPYKCECGKSFSSSTH LDLIRHQRTHTGKPYKCECGKSFSSRKDNLKNHQRTHTGKPYKCECGKSFSS RKDNLKNHQRTHTGKPYKCECGKSFSSQSSNLVRHQRTHTGKKTIS (761)
3'	ZnF5b R	AGGAAACAGAGCCAAACAC TTACA (762)	54. 66	LEPGKPYKCECGKSFSSPADLTRHQRTHTGKPYKCECGKSFSTTGALTEH QRTHGTGKPYKCECGKSFSSPADLTRHQRTHTGKPYKCECGKSFSSQSGNL EHQRTHTGKPYKCECGKSFSSERSHLRHQRTHTGKPYKCECGKSFSSRADN LTEHQRTHTGKPYKCECGKSFSSQANLRAHQRTHTGKPYKCECGKSFSSRS DHLTNHQRTHTGKKTIS (763)
3'	ZnF6a R	ATGCAGATTTGGACACAGA CTAGTAAACTGTGAAAACG TGACAAGGCAAGGTGGCGT GGG (764)	58. 57	LEPGKPYKCECGKSFSSRSDKLVHQRTHTGKPYKCECGKSFSSRRRCRAH QRTHGTGKPYKCECGKSFSSRSDHLTHQRTHTGKPYKCECGKSFSSRKDN NHQRTHTGKPYKCECGKSFSSQSGDLRRHQRTHTGKPYKCECGKSFSSRKDN LKNHQRTHTGKPYKCECGKSFSSDPGNLVRHQRTHTGKPYKCECGKSFSSR RTCRAHQRTHTGKPYKCECGKSFSSQANLRAHQRTHTGKPYKCECGKSFSS QAGHLASHQRTHTGKPYKCECGKSFSSRNDALTEHQRTHTGKPYKCECGKSF SSQANLRAHQRTHTGKPYKCECGKSFSSHRTTTLTNHQRTHTGKPYKCECG KSFSSHRTTTLTNHQRTHTGKPYKCECGKSFSSRADNLTEHQRTHTGKPYKCE CGKSFSSPADLTRHQRTHTGKPYKCECGKSFSSRSDHLTHQRTHTGKPYKCE CGKSFSSKKNALQNHQRTHTGKPYKCECGKSFSSRADNLTEHQRTHTGKPY KCECGKSFSSRRDELNVHQRTHTGKKTIS (765)
3'	ZnF6b R	GGACACAGAGTAGTAAAC (766)	55. 80	LEPGKPYKCECGKSFSSDQGNLVRHQRTHTGKPYKCECGKSFSSQSSSLVRH QRTHGTGKPYKCECGKSFSSQSSSLVRHQRTHTGKPYKCECGKSFSSQLAHLR AHQRTHTGKPYKCECGKSFSSSKKALTEHQRTHTGKPYKCECGKSFSSQRAH LERHQRTHTGKKTIS (767)
5'	ZnF10 F	AAAGCTAGCAGCATGGCA (768)	57. 55	LEPGKPYKCECGKSFSSQSGDLRRHQRTHTGKPYKCECGKSFSSRRDELNVH QRTHGTGKPYKCECGKSFSSERSHLRHQRTHTGKPYKCECGKSFSSERSHLR EHQRTHTGKPYKCECGKSFSSSGELVRHQRTHTGKPYKCECGKSFSSQAN LRAHQRTHTGKKTIS (769)
5'	ZnF11 F	CCTCTTATAAGGCCCAAGA CGATA (770)	52. 55	LEPGKPYKCECGKSFSSQSSSLIAHQRTHTGKPYKCECGKSFSSRSDHLTNH QRTHGTGKPYKCECGKSFSSRKDNLKNHQRTHTGKPYKCECGKSFSSSKKHLA EHQRTHTGKPYKCECGKSFSSRSDHLTNHQRTHTGKPYKCECGKSFSSQSS LIAHQRTHTGKPYKCECGKSFSTTGALTEHQRTHTGKPYKCECGKSFSSSTK NSLTEHQRTHTGKKTIS (771)
5'	ZnF12 F	CAACATCCTTGACTTAATC AC (772)	55. 00	LEPGKPYKCECGKSFSSSKKALTEHQRTHTGKPYKCECGKSFSTTGNLTVH QRTHGTGKPYKCECGKSFSTTGALTEHQRTHTGKPYKCECGKSFSSQAGHLA SHQRTHTGKPYKCECGKSFSSKNSLTEHQRTHTGKPYKCECGKSFSSSGN LTEHQRTHTGKPYKCECGKSFSSQSGNLTEHQRTHTGKKTIS (773)
5'	ZnF13 F	GGTAGCAAAAAGGTAACC (774)	46. 33	LEPGKPYKCECGKSFSSDKDLTRHQRTHTGKPYKCECGKSFSSQSSSLVRH QRTHGTGKPYKCECGKSFSSRKDNLKNHQRTHTGKPYKCECGKSFSSQANLRA AHQRTHTGKPYKCECGKSFSSERSHLRHQRTHTGKPYKCECGKSFSSSGH LVRHQRTHTGKKTIS (775)
3'	ZnF14 R	TGGGGTGCAAGAGGCCAGG CCAGAGTTGTTCTGGTC (776)	61. 28	LEPGKPYKCECGKSFSSDPGALVRHQRTHTGKPYKCECGKSFSSRNDALTEH QRTHGTGKPYKCECGKSFSSSGSLVRHQRTHTGKPYKCECGKSFSSSGSLV RHQRTHTGKPYKCECGKSFSSQLAHLRAHQRTHTGKPYKCECGKSFSSDCRD LARHQRTHTGKPYKCECGKSFSSPADNLTEHQRTHTGKPYKCECGKSFSSDP GHLVRHQRTHTGKPYKCECGKSFSSQLAHLRAHQRTHTGKPYKCECGKSFSS QSGDLRRHQRTHTGKPYKCECGKSFSSSGHLVRHQRTHTGKPYKCECGKSF SSRSDHLTHQRTHTGKKTIS (777)
3'	ZnF15 R	CGCATGCTGATTACGCTC CTGAC (778)	58. 41	LEPGKPYKCECGKSFSSDPGNLVRHQRTHTGKPYKCECGKSFSSKNSLTEH QRTHGTGKPYKCECGKSFSSKNSLTEHQRTHTGKPYKCECGKSFSSRADNLTE EHQRTHTGKPYKCECGKSFSSKKNALQNHQRTHTGKPYKCECGKSFSSRNDAL LTEHQRTHTGKPYKCECGKSFSSRRDELNVHQRTHTGKPYKCECGKSFSSHT GHLEHQRTHTGKKTIS (779)
3'	ZnF14 R	AGTCAAGCAACAGCATGA (780)	50. 89	LEPGKPYKCECGKSFSSQAGHLASHQRTHTGKPYKCECGKSFSSQRAHLERH QRTHGTGKPYKCECGKSFSSPADLTRHQRTHTGKPYKCECGKSFSSQSGDLR RHQRTHTGKPYKCECGKSFSSQSGNLTEHQRTHTGKPYKCECGKSFSSHRTT LTNHQRTHTGKKTIS (781)
3'	ZnF15 R	GTCAGCAACAGGATGATC CAATGCTATT (782)	59. 22	LEPGKPYKCECGKSFSSKKNALQNHQRTHTGKPYKCECGKSFSSSGELVRH QRTHGTGKPYKCECGKSFSTTGNLTVHQRTHTGKPYKCECGKSFSSSHSLT EHQRTHTGKPYKCECGKSFSSSGNLVRHQRTHTGKPYKCECGKSFSSSGN LVRHQRTHTGKPYKCECGKSFSSRADNLTEHQRTHTGKPYKCECGKSFSSQS

				GNLTEHQRTHTGKPYKCECGKSFSRKNLKNHQRTHTGKPYKCECGKSFS DPGALVRHQRTHTGKTS (783)
--	--	--	--	--

In embodiments, ZNFs for targeting human genomic safe harbor sites using any of the ZNF-based targeting elements to Chromosome X (e.g., hg38 chrX:134,419,661-134,541,172 or hg38 chrX:134,476,304-134,476,307 (85); chrX:134,476,337-134,476,340 (51)) are shown in **TABLE 5E**:

ChrX TTAA	NAME	TARGET (SEQ ID NO:)	SCORE	ZFP AMINO ACID CODE (SEQ ID NO:)
5'	ZnF4 1F	GTAGAACTCGCCTTATG (784)	54. 04	LEPGEKPYKCECGKSFSRDELNVHQRTHTGKPYKCECG KSFSTTGALTEHQRTHTGKPYKCECGKSFSHTGHLLEHQ THTGKPYKCECGKSFSSTHLDLIRHQRTHTGKPYKCECG KSFSQSSNLVRHQRTHTGKPYKCECGKSFSQSSSLVRHQ THTGKTS (785)
5'	ZnF4 2F	TGAATGAGTCCTGTCCATCIT (786)	55. 08	LEPGEKPYKCECGKSFSSTTGALTEHQRTHTGKPYKCECG KSFSTSGNLTEHQRTHTGKPYKCECGKSFSDPGALVRHQ THTGKPYKCECGKSFSSTKNSLTEHQRTHTGKPYKCECG KSFSHRTITLTHQRTHTGKPYKCECGKSFSRDELNVHQ THTGKPYKCECGKSFSQAGHLASHQRTHTGKTS (787)
5'	ZnF4 3F	AAGATTAGAACAAATGTCCAG (788)	60. 20	LEPGEKPYKCECGKSFSRADNLTEHQRTHTGKPYKCECG KSFSDPGALVRHQRTHTGKPYKCECGKSFSSTGNLTVHQ THTGKPYKCECGKSFSSTPADLTHQRTHTGKPYKCECG KSFSQLAHLRAHQRTHTGKPYKCECGKSFSKRNALQNHQ THTGKPYKCECGKSFSRKNLKNHQRTHTGKTS (789)
3'	ZnF4 4R	ACTCTAAGCAGCAATGTA (790)	59. 94	LEPGEKPYKCECGKSFSQSSSLVRHQRTHTGKPYKCECG KSFSTTGNLTVHQRTHTGKPYKCECGKSFSERSHLREHQ THTGKPYKCECGKSFSERSHLREHQRTHTGKPYKCECG KSFSQNSTLTHQRTHTGKPYKCECGKSFSSTHLDLIRHQ THTGKTS (791)
5'	ZnF4 5R	TGGGATAGTGAATAATGTC (792)	57. 10	LEPGEKPYKCECGKSFSDPGALVRHQRTHTGKPYKCECG KSFSTTGNLTVHQRTHTGKPYKCECGKSFSQSSNLVRHQ THTGKPYKCECGKSFSHRTITLTHQRTHTGKPYKCECG KSFSTSGNLVRHQRTHTGKPYKCECGKSFSRDELNVHQ THTGKTS (793)
5'	ZnF4 6R	AAAACCTGGGTCACATAAATAGATGAT (794)	61. 20	LEPGEKPYKCECGKSFSSTSGNLVRHQRTHTGKPYKCECG KSFSTSGNLVRHQRTHTGKPYKCECGKSFSQSSSLIAHQ THTGKPYKCECGKSFSQANLRAHQRTHTGKPYKCECG KSFSTHLDLIRHQRTHTGKPYKCECGKSFSDPGALVRHQ THTGKPYKCECGKSFSRDELNVHQRTHTGKPYKCECG KSFSTHLDLIRHQRTHTGKPYKCECGKSFSQANLRAHQ THTGKTS (795)
5'	ZnF4 7R	AAACATGGAAAAGGTCAAAAACCTGGG (796)	43. 59	LEPGEKPYKCECGKSFSRSDKLVRHQRTHTGKPYKCECG KSFSTTGALTEHQRTHTGKPYKCECGKSFSQANLRAHQ THTGKPYKCECGKSFSQSGNLTEHQRTHTGKPYKCECG KSFSTSGHLVRHQRTHTGKPYKCECGKSFSQANLRAHQ THTGKPYKCECGKSFSQRAHLERHQRTHTGKPYKCECG KSFSTSGNLTEHQRTHTGKPYKCECGKSFSQANLRAHQ THTGKTS (797)
3'	ZnF4 8R	AATGACTAGATGAAGTCCTACTG (798)	59. 44	LEPGEKPYKCECGKSFSRNDALTEHQRTHTGKPYKCECG KSFSTQNSTLTHQRTHTGKPYKCECGKSFSDPGALVRHQ THTGKPYKCECGKSFSQSSNLVRHQRTHTGKPYKCECG KSFSTTGNLTVHQRTHTGKPYKCECGKSFSRDELNVHQ THTGKPYKCECGKSFSDPGALVRHQRTHTGKPYKCECG KSFSTTGNLTVHQRTHTGKTS (799)

In embodiments, the mobile element enzyme is capable of inserting a donor DNA at a TA dinucleotide site. In embodiments, the mobile element enzyme is capable of inserting a donor DNA at a TTAA (SEQ ID NO: 440) tetranucleotide site.

In embodiments, the present disclosure relates to a system having nucleic acids encoding the enzyme and the donor DNA, respectively. **FIGs. 1A-1D** show examples of a system in accordance with embodiments of the present disclosure.

Linkers

In embodiments, the targeting element comprises a nucleic acid binding component of the gene-editing system (targeting element). In embodiments, the enzyme capable of targeted genomic integration (e.g., without limitation, a chimeric mobile element enzyme) and the targeting element, e.g., nucleic acid binding component of the gene-editing system are fused or linked to one another. For example, in embodiments, the mobile element enzyme and the targeting element, e.g., nucleic acid binding component of the gene-editing system are fused or linked to one another. In embodiments, the mobile element enzyme and the targeting element, e.g., nucleic acid binding component of the gene-editing system are connected via a linker.

In embodiments, the linker is a flexible linker. In embodiments, the flexible linker is substantially comprised of glycine and serine residues, optionally wherein the flexible linker comprises $(\text{Gly}_4\text{Ser})_n$, where n is from about 1 to about 12. In embodiments, the flexible linker is of about 20, or about 30, or about 40, or about 50, or about 60 amino acid residues.

In embodiments, the flexible linker is about 50, or about 100, or about 150, or about 200 amino acid residues in length. In embodiments, the flexible linker comprises at least about 150 nucleotides (nt), or at least about 200 nt, or at least about 250 nt, or at least about 300 nt, or at least about 350 nt, or at least about 400 nt, or at least about 450 nt, or at least about 500 nt, or at least about 500 nt, or at least about 600 nt. In embodiments, the flexible linker comprises from about 450 nt to about 500 nt.

In embodiments, the mobile element enzyme and the targeting element, e.g., nucleic acid binding component of the gene-editing system are encoded on a single polypeptide.

In embodiments, the donor DNA comprises a gene encoding a complete polypeptide. In embodiments, the donor DNA comprises a gene which is defective or substantially absent in a disease state.

Inteins

Inteins (INTervening proTEINS) are mobile genetic elements that are protein domains, found in nature, with the capability to carry out the process of protein splicing. See Sarmiento & Camarero (2019) *Current Protein & Peptide Science*, 20(5), 408–424, which is incorporated by reference herein in its entirety. Protein splicing is a post-translation biochemical modification which results in the cleavage and formation of peptide bonds between precursor polypeptide segments flanking the intein. Inteins apply standard enzymatic strategies to excise themselves post-translationally from a precursor protein via protein splicing. Nanda *et al.*, *Microorganisms* vol. 8,12 2004. 16 Dec. 2020, doi:10.3390/microorganisms8122004. An intein can splice its flanking N- and C-terminal domains to become a mature protein and excise itself from a sequence. For example, split inteins have been used to control the delivery of

heterologous genes into transgenic organisms. See Wood & Camarero (2014) *J Biol Chem.* 289(21):14512-14519. This approach relies on splitting the target protein into two segments, which are then post-translationally reconstituted *in vivo* by protein trans-splicing (PTS). See Aboye & Camarero (2012) *J. Biol. Chem.* 287, 27026–27032. More recently, an intein-mediated split-Cas9 system has been developed to incorporate Cas9 into cells and reconstitute nuclease activity efficiently. Truong *et al.*, *Nucleic Acids Res.* 2015, 43 (13), 6450–6458. The protein splicing excises the internal region of the precursor protein, which is then followed by the ligation of the N-extein and C-extein fragments, resulting in two polypeptides – the excised intein and the new polypeptide produced by joining the C- and N-exteins. Sarmiento & Camarero (2019).

In embodiments, intein-mediated incorporation of DNA binders such as, without limitation, dCas9, dCas12j, or TALEs, allows creation of a split-enzyme system such as, without limitation, split-MLT mobile element enzyme system, that permits reconstitution of the full-length enzyme, *e.g.*, MLT mobile element enzyme, from two smaller fragments. This allows avoiding the need to express DNA binders at the N- or C-terminus of an enzyme, *e.g.*, MLT mobile element enzyme. In this approach, the two portions of an enzyme, *e.g.*, MLT mobile element enzyme, are fused to the intein and, after co-expression, the intein allows producing a full-length enzyme, *e.g.*, MLT mobile element enzyme, by post-translation modification. Thus, in embodiments, a nucleic acid encoding the enzyme capable of targeted genomic integration comprises an intein. In embodiments, the nucleic acid encodes the enzyme in the form of first and second portions with the intein encoded between the first and second portions, such that the first and second portions are fused into a functional enzyme upon post-translational excision of the intein from the enzyme.

In embodiments, an intein is a suitable ligand-dependent intein, for example, an intein selected from those described in U.S. Patent No. 9,200,045; Mootz *et al.*, *J. Am. Chem. Soc.* 2002; 124, 9044-9045; Mootz *et al.*, *J. Am. Chem. Soc.* 2003; 125, 10561-10569; Buskirk *et al.*, *Proc. Natl. Acad. Sci. USA.* 2004; 101, 10505-10510; Skretas & Wood. *Protein Sci.* 2005; 14, 523-532; Schwartz, *et al.*, *Nat. Chem. Biol.* 2007; 3, 50-54; Peck *et al.*, *Chem. Biol.* 2011; 18 (5), 619-630; the entire contents of each of which are hereby incorporated by reference herein.

In embodiments the intein is NpuN (Intein-N) (SEQ ID NO: 423) and/or NpuC (Intein-C) (SEQ ID NO: 424), or a variant thereof, *e.g.*, a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

SEQ ID NO: 423: nucleotide sequence of NpuN (Intein-N)

```
GGCGGATCTGGCGGTAGTGCTGAGTATTGTCTGAGTTACGAAACGGAAATACTCACGGTTGAGTATGGGCTTCTTCC
AATTGGCAAAATCGTTGAAAAGCGCATAGAGTGTAACGGTGTATTCCGTCGATAACAACGGTAATATCTACACCCAGC
CGGTAGCTCAGTGGCACCACCGAGGCGAACAGGAAGTGTTCGAGTATTGCTTGGAGATGGCTCCCTTATCCGCGCC
ACTAAAGACCATAAGTTTATGACGGTTGACGGGCAGATGCTGCCTATAGACGAAATATTTGAGAGAGAGCTGGACTT
GATGAGAGTCGATAATCTGCCAAAT
```

SEQ ID NO: 424: nucleotide sequence of NpuC (Intein-C)

GGCGGATCTGGCGGTAGTGGGGGTTCCGGATCCATAAAGATAGCTACTAGGAAATATCTTGGCAAACAAAACGTCTA
 TGACATAGGAGTTGAGCGAGATCACAATTTTGCTTTGAAGAATGGGTTTCATCGCGTCTAATTGCTTCAACGCTAGCG
 GCGGGTCAGGAGGCTCTGCTGCAAGC

Nucleic Acids of the Disclosure

In embodiments, a nucleic acid encoding the enzyme is RNA. In embodiments, a nucleic acid encoding the transgene is DNA.

In embodiments, the enzyme (e.g., without limitation, the mobile element enzyme) is encoded by a recombinant or synthetic nucleic acid. In embodiments, the nucleic acid is RNA, optionally a helper RNA. In embodiments, the nucleic acid is RNA that has a 5'-m7G cap (cap0, or cap1, or cap2), optionally with pseudouridine substitution (e.g., without limitation n-methyl-pseudouridine), and optionally a poly-A tail of about 30, or about 50, or about 100, or about 150 nucleotides in length. In embodiments, the poly-A tail is of about 30 nucleotides in length, optionally 34 nucleotides in length. In embodiments, a nuclear localization signal is placed before the enzyme start codon at the N-terminus, optionally at the C-terminus.

In embodiments, the nucleic acid that is RNA has a 5'-m7G cap (cap 0, or cap 1, or cap 2).

In embodiments, the nucleic acid comprises a 5' cap structure, a 5'-UTR comprising a Kozak consensus sequence, a 5'-UTR comprising a sequence that increases RNA stability *in vivo*, a 3'-UTR comprising a sequence that increases RNA stability *in vivo*, and/or a 3' poly(A) tail.

In embodiments, the enzyme (e.g., without limitation, a mobile element enzyme) is incorporated into a vector or a vector-like particle. In embodiments, the vector is a non-viral vector.

In embodiments, a nucleic acid encoding the enzyme in accordance with embodiments of the present disclosure, is DNA.

In various embodiments, a construct comprising a donor DNA is any suitable genetic construct, such as a nucleic acid construct, a plasmid, or a vector. In various embodiments, the construct is DNA, which is referred to herein as a donor DNA. In embodiments, sequences of a nucleic acid encoding the donor DNA is codon optimized to provide improved mRNA stability and protein expression in mammalian systems.

In embodiments, the enzyme and the donor DNA are included in different vectors. In embodiments, the enzyme and the donor DNA are included in the same vector.

In various embodiments, a nucleic acid encoding the enzyme capable of targeted genomic integration (e.g., without limitation, a mobile element enzyme which is a chimeric mobile element enzyme) is RNA (e.g., helper RNA), and a nucleic acid encoding a donor DNA is DNA.

As would be appreciated in the art, a donor DNA often includes an open reading frame that encodes a transgene at the middle of donor DNA and terminal repeat sequences at the 5' and 3' end of the donor DNA. The translated mobile element enzyme binds to the 5' and 3' sequence of the donor DNA and carries out the transposition function.

In embodiments, donor DNA or transgene are used interchangeably with mobile elements, which are used to refer to polynucleotides capable of inserting copies of themselves into other polynucleotides. The term donor DNA is well known to those skilled in the art and includes classes of donor DNAs that can be distinguished on the basis of sequence organization, for example inverted terminal sequences at each end, and/or directly repeated long terminal repeats (LTRs) at the ends. In embodiments, the donor DNA as described herein may be described as a piggyBac like element, e.g., a donor DNA element that is characterized by its traceless excision, which recognizes TTAA (SEQ ID NO: 440) sequence and restores the sequence at the insert site back to the original TTAA (SEQ ID NO: 440) sequence after removal of the donor DNA.

In embodiments, the donor DNA is flanked by one or more end sequences or terminal ends. In embodiments, the donor DNA is or comprises a gene encoding a complete polypeptide. In embodiments, the donor DNA is or comprises a gene which is defective or substantially absent in a disease state.

In embodiments, the donor DNA includes a MLT mobile element enzyme (e.g., without limitation, a MLT mobile element enzyme having at least about 90% identity to the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 10, or SEQ ID NO: 11). For example, the mobile element enzyme can act on a left terminal end having a nucleotide sequence of SEQ ID NO: 431 or a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto. In embodiments, the donor DNA can act on a right terminal end having a nucleotide sequence of SEQ ID NO: 432 or a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto. In embodiments, the donor DNA acts on both MLT left donor DNA end and MLT right donor DNA end, having nucleotide sequences of SEQ ID NO: 431 and of SEQ ID NO: 432 respectively, or a sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

In embodiments, a MLT left donor DNA end (5' to 3') is as follows

TTAACACTTGGATTGCGGGAAACGAGTTAAGTCGGCTCGCGTGAATTGCGCGTACTCCGCGGGAGCCGTC
TTAACTCGGTTTCATATAGATTTGCGGTGGAGTGC GGAAACGTGTAAACTCGGGCCGATTGTAACTGCGT
ATTACCAAATATTTGTT (SEQ ID NO: 431)

In embodiments, a MLT right donor DNA end (5' to 3') is as follows

```

AATTATTTTATGTACTGAATAGATAAAAAAATGTCTGTGATTGAATAAATTTTCATTTTACACAAGAAA
CCGAAAATTTCAATTCGAACCCATACTTCAAAAGATATAGGCATTTTAACTAACTCTGATTTTGC
GCGGGAAACCTAAATAATTGCCCGCGCCATCTTATATTTTGGCGGGAAATTCACCCGACACCGTGGTGT
AA (SEQ ID NO: 432).

```

In embodiments, a transgene is associated with various regulatory elements that are selected to ensure stable expression of a construct with the transgene. Thus, in embodiments, a transgene is encoded by a non-viral vector (*e.g.*, without limitation, a DNA plasmid) that can comprise one or more insulator sequences that prevent or mitigate activation or inactivation of nearby genes. The insulators flank the donor DNA (transgene cassette) to reduce transcriptional silencing and position effects imparted by chromosomal sequences. As an additional effect, the insulators can eliminate functional interactions of the transgene enhancer and promoter sequences with neighboring chromosomal sequences. In embodiments, the one or more insulator sequences comprise an HS4 insulator (1.2-kb 5' -HS4 chicken β -globin (cHS4) insulator element) and an D4Z4 insulator (tandem macrosatellite repeats linked to Facio-Scapulo-Humeral Dystrophy (FSHD). In embodiments, the sequences of the HS4 insulator and the D4Z4 insulator are as described in Rival-Gervier *et al. Mol Ther.* 2013 Aug; 21(8):1536-50, which is incorporated herein by reference in its entirety.

In embodiments, the transgene is inserted into a GSHS location in a host genome. GSHSs is defined as loci well-suited for gene transfer, as integrations within these sites are not associated with adverse effects such as proto-oncogene activation, tumor suppressor inactivation, or insertional mutagenesis. GSHSs can be defined by the following criteria: (1) distance of at least 50 kb from the 5' end of any gene, (2) distance of at least 300 kb from any cancer-related gene, (3) distance of at least 300 kb from any microRNA (miRNA), (4) location outside a transcription unit, and (5) location outside ultra-conserved regions (UCRs) of the human genome. See Papapetrou *et al. Nat Biotechnol* 2011;29:73-8; Bejerano *et al. Science* 2004;304:1321-5.

Furthermore, the use of GSHS locations can allow stable transgene expression across multiple cell types. One such site, chemokine C-C motif receptor 5 (CCR5) has been identified and used for integrative gene transfer. CCR5 is a member of the beta chemokine receptor family and is required for the entry of R5 tropic viral strains involved in primary infections. A homozygous 32 bp deletion in the CCR5 gene confers resistance to HIV-1 virus infections in humans. Disrupted CCR5 expression, naturally occurring in about 1% of the Caucasian population, does not appear to result in any reduction in immunity. Lobritz *et al., Viruses* 2010;2:1069-105. A clinical trial has demonstrated safety and efficacy of disrupting CCR5 via targetable nucleases. Tebas *et al., HIV. N Engl J Med* 2014;370:901-10.

In embodiments, the donor DNA is under control of a tissue-specific promoter. The tissue-specific promoter is, *e.g.*, without limitation, a liver-specific promoter. In embodiments, the liver-specific promoter is an LP1 promoter that, in

embodiments, is a human LP1 promoter. The LP1 promoter is described, e.g., in Nathwani *et al. Blood* vol. 2006;107(7):2653-61, and it is constructed, without limitation, as described in Nathawani *et al.*

It should be appreciated however that a variety of promoters can be used, including other tissue-specific promoters, inducible promoters, constitutive promoters, etc.

In embodiments, the present nucleic acids include polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides, or analogs or derivatives thereof. In embodiments, there is provided double- and single-stranded DNA, as well as double- and single-stranded RNA, and RNA-DNA hybrids. In embodiments, transcriptionally-activated polynucleotides such as methylated or capped polynucleotides are provided. In embodiments, the present compositions are mRNA or DNA.

In embodiments, the present non-viral vectors are linear or circular DNA molecules that comprise a polynucleotide encoding a polypeptide and is operably linked to control sequences, wherein the control sequences provide for expression of the polynucleotide encoding the polypeptide. In embodiments, the non-viral vector comprises a promoter sequence, and transcriptional and translational stop signal sequences. Such vectors may include, among others, chromosomal and episomal vectors, e.g., vectors bacterial plasmids, from donor DNAs, from yeast episomes, from insertion elements, from yeast chromosomal elements, and vectors from combinations thereof. The present constructs may contain control regions that regulate as well as engender expression.

In embodiments, the construct comprising the enzyme and/or transgene is codon optimized. Transgene codon optimization is used to optimize therapeutic potential of the transgene and its expression in the host organism. Codon optimization is performed to match the codon usage in the transgene with the abundance of transfer RNA (tRNA) for each codon in a host organism or cell. Codon optimization methods are known in the art and described in, for example, WO 2007/142954, which is incorporated by reference herein in its entirety. Optimization strategies can include, for example, the modification of translation initiation regions, alteration of mRNA structural elements, and the use of different codon biases.

In embodiments, the construct comprising the enzyme and/or transgene includes several other regulatory elements that are selected to ensure stable expression of the construct. Thus, in embodiments, the non-viral vector is a DNA plasmid that can comprise one or more insulator sequences that prevent or mitigate activation or inactivation of nearby genes. In embodiments, the one or more insulator sequences comprise an HS4 insulator (1.2-kb 5'-HS4 chicken β -globin (cHS4) insulator element) and an D4Z4 insulator (tandem macrosatellite repeats linked to Facio-Scapulo-Humeral Dystrophy (FSHD). In embodiments, the sequences of the HS4 insulator and the D4Z4 insulator are as described in Rival-Gervier *et al. Mol Ther.* 2013 Aug; 21(8):1536-50, which is incorporated herein by reference in its entirety. In embodiments, the gene of the construct comprising the enzyme and/or transgene is capable of transposition in the presence of a mobile element enzyme. In embodiments, the non-viral vector in accordance with embodiments

of the present disclosure comprises a nucleic acid construct encoding a mobile element enzyme. The mobile element enzyme is an RNA mobile element enzyme plasmid. In embodiments, the non-viral vector further comprises a nucleic acid construct encoding a DNA mobile element enzyme plasmid. In embodiments, the mobile element enzyme is an *in vitro*-transcribed mRNA mobile element enzyme. The mobile element enzyme is capable of excising and/or transposing the gene from the construct comprising the enzyme and/or transgene to site- or locus-specific genomic regions.

In embodiments, the enzyme and the donor DNA are included in the same vector.

In embodiments, the enzyme is disposed on the same (*cis*) or different vector (*trans*) than a donor DNA with a transgene. Accordingly, in embodiments, the enzyme and the donor DNA encompassing a transgene are in *cis* configuration such that they are included in the same vector. In embodiments, the enzyme and the donor DNA encompassing a transgene are in *trans* configuration such that they are included in different vectors. The vector is any non-viral vector in accordance with the present disclosure.

In some aspects, a nucleic acid encoding the enzyme capable of targeted genomic integration (e.g., a mobile element enzyme or a chimeric mobile element enzyme) in accordance with embodiments of the present disclosure is provided. The nucleic acid is or comprises DNA or RNA. In embodiments, the nucleic acid encoding the enzyme is DNA. In embodiments, the nucleic acid encoding the enzyme capable of targeted genomic integration (e.g., a chimeric mobile element enzyme) is RNA such as, e.g., helper RNA. In embodiments, the chimeric mobile element enzyme is incorporated into a vector. In embodiments, the vector is a non-viral vector.

In embodiments, a nucleic acid encoding the transgene in accordance with embodiments of the present disclosure is provided. The nucleic acid is or comprises DNA or RNA. In embodiments, the nucleic acid encoding the transgene is DNA. In embodiments, the nucleic acid encoding the transgene is RNA such as, e.g., helper RNA. In embodiments, the transgene is incorporated into a vector. In embodiments, the vector is a non-viral vector.

In embodiments, the present enzyme can be in the form of an RNA or DNA and have one or two N-terminus nuclear localization signal (NLS) to shuttle the protein more efficiently into the nucleus. For example, in embodiments, the present enzyme further comprises one, two, three, four, five, or more NLSs. Examples of NLS are provided in Kosugi et al. (*J. Biol. Chem.* (2009) 284:478-485; incorporated by reference herein). In a particular embodiment, the NLS comprises the consensus sequence K(K/R)X(K/R) (SEQ ID NO: 348). In an embodiment, the NLS comprises the consensus sequence (K/R)(K/R)X₁₀₋₁₂(K/R)_{3/5} (SEQ ID NO: 349), where (K/R)_{3/5} represents at least three of the five amino acids is either lysine or arginine. In an embodiment, the NLS comprises the c-myc NLS. In a particular embodiment, the c-myc NLS comprises the sequence PAAKRVKLD (SEQ ID NO: 350). In a particular embodiment, the NLS is the nucleoplasmin NLS. In embodiments, the nucleoplasmin NLS comprises the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 351). In embodiments, the NLS comprises the SV40 Large T-antigen NLS. In embodiments, the SV40 Large T-antigen NLS comprises the sequence PKKKRKV (SEQ ID NO: 352). In a particular

embodiment, the NLS comprises three SV40 Large T-antigen NLSs (e.g., DPKKKRKVDPKKKKRKVDPKKKKRKV (SEQ ID NO: 353). In embodiments, the NLS may comprise mutations/variations in the above sequences such that they contain 1 or more substitutions, additions or deletions (e.g., about 1, or about 2, or about 3, or about 4, or about 5, or about 10 substitutions, additions, or deletions).

In some aspects, a host cell comprising the nucleic acid in accordance with embodiments of the present disclosure is provided.

Lipids

In embodiments, at least one of the first nucleic acid and the second nucleic acid is in the form of a lipid nanoparticle (LNP). In embodiments, a composition comprising the first and second nucleic acids is in the form of an LNP.

In embodiments, a nucleic acid encoding the enzyme and a nucleic acid encoding the transgene are contained within the same lipid nanoparticle (LNP). In embodiments, the nucleic acid encoding the enzyme and the nucleic acid encoding the donor DNA are a mixture incorporated into or associated with the same LNP. In embodiments, the nucleic acid encoding the enzyme and the nucleic acid encoding the donor DNA are in the form of a co-formulation incorporated into or associated with the same LNP.

In embodiments, the LNP is selected from 1,2-dioleoyl-3-trimethylammonium propane (DOTAP), a cationic cholesterol derivative mixed with dimethylaminoethane-carbamoyl (DC-Chol), phosphatidylcholine (PC), triolein (glyceryl trioleate), and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[carboxy(polyethylene glycol)-2000] (DSPE-PEG), 1,2-dimyristoyl-rac-glycero-3-methoxypolyethyleneglycol – 2000 (DMG-PEG 2K), and 1,2 distearol -sn-glycerol-3-phosphocholine (DSPC) and/or comprising of one or more molecules selected from polyethylenimine (PEI) and poly(lactic-co-glycolic acid) (PLGA), and N-Acetylgalactosamine (GalNAc).

In embodiments, an LNP is as described, e.g., in Patel *et al.*, *J Control Release* 2019; 303:91-100. The LNP can comprise one or more of a structural lipid (e.g., DSPC), a PEG-conjugated lipid (CDM-PEG), a cationic lipid (MC3), cholesterol, and a targeting ligand (e.g., GalNAc).

In embodiments, a nanoparticle is a particle having a diameter of less than about 1000 nm. In embodiments, nanoparticles of the present disclosure have a greatest dimension (e.g., diameter) of about 500 nm or less, or about 400 nm or less, or about 300 nm or less, or about 200 nm or less, or about 100 nm or less. In embodiments, nanoparticles of the present disclosure have a greatest dimension ranging between about 50 nm and about 150 nm, or between about 70 nm and about 130 nm, or between about 80 nm and about 120 nm, or between about 90 nm and about 110 nm. In embodiments, the nanoparticles of the present disclosure have a greatest dimension (e.g., a diameter) of about 100 nm.

In some aspects, a cell in accordance with the present disclosure is prepared via an *in vivo* genetic modification method. In embodiments, a genetic modification in accordance with the present disclosure is performed via an *ex vivo* method.

In some aspects, a cell in accordance with the present disclosure is prepared by contacting the cell with an enzyme capable of targeted genomic integration (e.g., without limitation, a mammalian mobile element enzyme) *in vivo*. In embodiments, the cell is contacted with the enzyme *ex vivo*.

In embodiments, the present method provides reduced insertional mutagenesis or oncogenesis as compared to a method with a non-chimeric mobile element enzyme.

Therapeutic Applications

In embodiments, the transgene of interest in accordance with embodiments of the present disclosure can encode various genes.

In embodiments, the enzyme (e.g., without limitations, a mobile element enzyme), and the donor DNA are included in the same pharmaceutical composition.

In embodiments, the enzyme (e.g., without limitations, a mobile element enzyme) and the donor DNA are included in different pharmaceutical compositions.

In embodiments, the enzyme and the donor DNA are co-transfected.

In embodiments, the enzyme and the donor DNA are transfected separately.

In embodiments, a method of producing a molecule with the transfected cell is provided, wherein the transfected cell is generated using a method of making a viral packaging or producer cell line in accordance with embodiments of the present disclosure. The molecule is optionally a protein. In embodiments, the protein is an antibody.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with REP expression or substantially enhanced REP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression and REP expression or substantially enhanced REP expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with REP expression or substantially enhanced VP1 expression.

In embodiments, the method of making a viral packaging or producer cell line of the present disclosure provides a stable cell with substantially reduced or ablated CAP expression.

In embodiments, the viral packaging or producer cell line of the present disclosure is suitable for providing substantially reduced empty or cargo-free capsid.

In embodiments, a transfecting cell for gene therapy is provided, wherein the transfecting cell is generated using a method of making a viral packaging or producer cell line in accordance with embodiments of the present disclosure.

In embodiments, a method of delivering a cell therapy is provided, comprising administering to a patient in need thereof the transfecting cell generated using a method of making a viral packaging or producer cell line in accordance with embodiments of the present disclosure.

In embodiments, a method of treating a disease or condition using a cell therapy, comprising administering to a patient in need thereof the transfecting cell generated using a method of making a viral packaging or producer cell line in accordance with embodiments of the present disclosure.

In embodiments, a method of treating a disease or condition using an antibody therapy, comprising administering to a patient in need thereof the transfecting cell generated using a method of making a viral packaging or producer cell line in accordance with embodiments of the present disclosure.

In embodiments, the disease or condition is or comprises cancer. In embodiments, the cancer is or comprises an adrenal cancer, a biliary tract cancer, a bladder cancer, a bone/bone marrow cancer, a brain cancer, a breast cancer, a cervical cancer, a colorectal cancer, a cancer of the esophagus, a gastric cancer, a head/neck cancer, a hepatobiliary cancer, a kidney cancer, a liver cancer, a lung cancer, an ovarian cancer, a pancreatic cancer, a pelvis cancer, a pleura cancer, a prostate cancer, a renal cancer, a skin cancer, a stomach cancer, a testis cancer, a thymus cancer, a thyroid cancer, a uterine cancer, a lymphoma, a melanoma, a multiple myeloma, or a leukemia.

In embodiments, the cancer is selected from one or more of the basal cell carcinoma, biliary tract cancer; bladder cancer; bone cancer; brain and central nervous system cancer; breast cancer; cancer of the peritoneum; cervical cancer; choriocarcinoma; colon and rectum cancer; connective tissue cancer; cancer of the digestive system; endometrial cancer; esophageal cancer; eye cancer; cancer of the head and neck; gastric cancer; glioblastoma; hepatic carcinoma; hepatoma; intra-epithelial neoplasm; kidney or renal cancer; larynx cancer; leukemia; liver cancer; lung cancer; melanoma; myeloma; neuroblastoma; oral cavity cancer; ovarian cancer; pancreatic cancer; prostate cancer; retinoblastoma; rhabdomyosarcoma; rectal cancer; cancer of the respiratory system; salivary gland carcinoma; sarcoma; skin cancer; squamous cell cancer; stomach cancer; testicular cancer; thyroid cancer; uterine or endometrial cancer; cancer of the urinary system; vulval cancer; Hodgkin's lymphoma; non-Hodgkin's lymphoma; B-cell lymphoma; small lymphocytic (SL) NHL; intermediate grade/follicular NHL; intermediate grade diffuse NHL; high grade

immunoblastic NHL; high grade lymphoblastic NHL; high grade small non-cleaved cell NHL; bulky disease NHL; mantle cell lymphoma; AIDS-related lymphoma; Waldenstrom's Macroglobulinemia; chronic lymphocytic leukemia (CLL); acute lymphoblastic leukemia (ALL); and Hairy cell leukemia.

In embodiments, the cancer is selected from one or more of basal cell carcinoma, biliary tract cancer; bladder cancer; bone cancer; brain and central nervous system cancer; breast cancer; cancer of the peritoneum; cervical cancer; choriocarcinoma; colon and rectum cancer; connective tissue cancer; cancer of the digestive system; endometrial cancer; esophageal cancer; eye cancer; cancer of the head and neck; gastric cancer (including gastrointestinal cancer); glioblastoma; hepatic carcinoma; hepatoma; intra-epithelial neoplasm; kidney or renal cancer; larynx cancer; leukemia; liver cancer; lung cancer (e.g., small-cell lung cancer, non-small cell lung cancer, adenocarcinoma of the lung, and squamous carcinoma of the lung); melanoma; myeloma; neuroblastoma; oral cavity cancer (lip, tongue, mouth, and pharynx); ovarian cancer; pancreatic cancer; prostate cancer; retinoblastoma; rhabdomyosarcoma; rectal cancer; cancer of the respiratory system; salivary gland carcinoma; sarcoma; skin cancer; squamous cell cancer; stomach cancer; testicular cancer; thyroid cancer; uterine or endometrial cancer; cancer of the urinary system; vulvar cancer; lymphoma including Hodgkin's and non-Hodgkin's lymphoma, as well as B-cell lymphoma (including low grade/follicular non-Hodgkin's lymphoma (NHL); small lymphocytic (SL) NHL; intermediate grade/follicular NHL; intermediate grade diffuse NHL; high grade immunoblastic NHL; high grade lymphoblastic NHL; high grade small non-cleaved cell NHL; bulky disease NHL; mantle cell lymphoma; AIDS-related lymphoma; and Waldenstrom's Macroglobulinemia; chronic lymphocytic leukemia (CLL); acute lymphoblastic leukemia (ALL); Hairy cell leukemia; chronic myeloblastic leukemia; as well as other carcinomas and sarcomas; and post-transplant lymphoproliferative disorder (PTLD), as well as abnormal vascular proliferation associated with phakomatoses, edema (e.g. that associated with brain tumors), and Meigs syndrome.

In embodiments, the disease or condition is or comprises an infectious disease. In embodiments, the infectious disease is a coronavirus infection, optionally selected from infection with SAR-CoV, MERS-CoV, and SARS-CoV-2, or variants thereof.

In embodiments, the infectious disease is or comprises a disease comprising a viral infection, a parasitic infection, or a bacterial infection. In embodiments, the viral infection is caused by a virus of family *Flaviviridae*, a virus of family *Picornaviridae*, a virus of family *Orthomyxoviridae*, a virus of family *Coronaviridae*, a virus of family *Retroviridae*, a virus of family *Paramyxoviridae*, a virus of family *Bunyaviridae*, or a virus of family *Reoviridae*.

In embodiments, the virus of family *Coronaviridae* comprises a betacoronavirus or an alphacoronavirus, optionally wherein the betacoronavirus is selected from SARS-CoV-2, SARS-CoV, MERS-CoV, HCoV-HKU1, and HCoV-OC43, or the alphacoronavirus is selected from a HCoV-NL63 and HCoV-229E. In embodiments, the infectious disease comprises a coronavirus infection 2019 (COVID-19).

In embodiments, the method requires a single administration. In embodiments, the method requires a plurality of administrations.

Isolated Cell

In some aspects of the present disclosure, an isolated cell is provided that comprises the transfected cell in accordance with embodiments of the present disclosure.

In some aspects, the present disclosure provides an *ex vivo* gene therapy approach. Accordingly, in embodiments, the method that is used to treat an inherited or acquired disease in a patient in need thereof comprises (a) contacting a cell obtained from a patient (autologous) or another individual (allogeneic) with a transfected cell in accordance with embodiments of the present disclosure; and (b) administering the cell to a patient in need thereof.

One of the advantages of *ex vivo* gene therapy is the ability to "sample" the transduced cells before patient administration. This facilitates efficacy and allows performing safety checks before introducing the cell(s) to the patient. For example, the transduction efficiency and/or the clonality of integration can be assessed before infusion of the product. The present disclosure provides transfected cells and methods that can be effectively used for *ex vivo* gene modification.

In embodiments, a composition comprising transfected cells in accordance with the present disclosure comprises a pharmaceutically acceptable carrier, excipient, or diluent.

Methods of formulating suitable pharmaceutical compositions are known in the art, see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005; and the books in the series Drugs and the Pharmaceutical Sciences: a Series of Textbooks and Monographs (Dekker, N.Y.). For example, pharmaceutical compositions suitable for injectable use can include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile, and the fluid should be easy to draw up by a syringe. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, and sodium chloride in the composition.

Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle, which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying, which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Therapeutic compounds can be prepared with carriers that will protect the therapeutic compounds against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as collagen, ethylene vinyl acetate, polyanhydrides (e.g., poly[1,3-bis(carboxyphenoxy)propane-co-sebacic-acid] (PCPP-SA) matrix, fatty acid dimer-sebacic acid (FAD-SA) copolymer, poly(lactide-co-glycolide)), polyglycolic acid, collagen, polyorthoesters, polyethyleneglycol-coated liposomes, and polylactic acid. Such formulations can be prepared using standard techniques, or obtained commercially, e.g., from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811. Semisolid, gelling, soft-gel, or other formulations (including controlled release) can be used, e.g., when administration to a surgical site is desired. Methods of making such formulations are known in the art and can include the use of biodegradable, biocompatible polymers. See, e.g., Sawyer *et al.*, *Yale J Biol Med.* 2006; 79(3-4): 141-152.

In embodiments, there is provided a method of transforming a cell using the construct comprising the enzyme and/or transgene described herein in the presence of a mobile element enzyme to produce a stably transfected cell which results from the stable integration of a gene of interest into the cell. In embodiments, the stable integration comprises an introduction of a polynucleotide into a chromosome or mini-chromosome of the cell and, therefore, becomes a relatively permanent part of the cellular genome.

In embodiments, there is provided a transgenic organism that may comprise cells which have been transformed by the methods of the present disclosure. In embodiments, the organism may be a mammal or an insect. When the organism is a mammal, the organism may include, but is not limited to, a mouse, a rat, a monkey, a dog, a rabbit, and the like. When the organism is an insect, the organism may include, but is not limited to, a fruit fly, a mosquito, a bollworm, and the like.

In embodiments, the cells produced in accordance with embodiments of the present disclosure, and/or components for generating cells, is included in a container, kit, pack, or dispenser together with instructions for administration.

Also provided herein are kits comprising: one or more genetic constructs encoding the present enzyme and donor DNA and instructions and/or reagents for the use of the same.

Also provided herein are kits comprising: i) a transfected cell in accordance with embodiments of the present disclosure, ii) instructions for the use of the transfected cell.

Furthermore, in embodiments, a kit is provided for creating an AAV of packaging cell line or an AAV library of packaging cell lines, and instructions for creating a library. A subsequent user-defined transfection of the viral genome can be performed to generate a producer cell line. In embodiments, the kit includes instructions for transfection of an AAV packaging cell line with a desired transgene.

In embodiments, a kit includes an AAV library of packaging cell lines and instructions for creating a library

In some aspects, a kit is provided that comprises an enzyme (e.g., without limitation, a recombinant mammalian mobile element enzyme) or a nucleic acid in accordance with embodiments of the present disclosure, and instructions for introducing DNA and/or RNA into a cell using the enzyme.

Definitions

The following definitions are used in connection with the disclosure disclosed herein. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of skill in the art to which this invention belongs.

As used herein, "a," "an," or "the" can mean one or more than one.

Further, the term "about" when used in connection with a referenced numeric indication means the referenced numeric indication plus or minus up to 10% of that referenced numeric indication. For example, the language "about 50" covers the range of 45 to 55.

An "effective amount," when used in connection with medical uses is an amount that is effective for providing a measurable treatment, prevention, or reduction in the rate of pathogenesis of a disease of interest.

The term "*in vivo*" refers to an event that takes place in a subject's body.

The term "*ex vivo*" refers to an event which involves treating or performing a procedure on a cell, tissue and/or organ which has been removed from a subject's body. Aptly, the cell, tissue and/or organ may be returned to the subject's body in a method of treatment or surgery.

As used herein, the term "variant" encompasses but is not limited to nucleic acids or proteins which comprise a nucleic acid or amino acid sequence which differs from the nucleic acid or amino acid sequence of a reference by way of one or more substitutions, deletions and/or additions at certain positions. The variant may comprise one or more conservative substitutions. Conservative substitutions may involve, e.g., the substitution of similarly charged or uncharged amino acids.

"Carrier" or "vehicle" as used herein refer to carrier materials suitable for drug administration. Carriers and vehicles useful herein include any such materials known in the art, e.g., any liquid, gel, solvent, liquid diluent, solubilizer, surfactant, lipid, or the like, which is nontoxic, and which does not interact with other components of the composition in a deleterious manner.

The phrase "pharmaceutically acceptable" refers to those compounds, materials, compositions, and/or dosage forms that are, within the scope of sound medical judgment, suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problems or complications commensurate with a reasonable benefit/risk ratio.

The terms "pharmaceutically acceptable carrier" or "pharmaceutically acceptable excipient" are intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and inert ingredients. The use of such pharmaceutically acceptable carriers or pharmaceutically acceptable excipients for active pharmaceutical ingredients is well known in the art. Except insofar as any conventional pharmaceutically acceptable carrier or pharmaceutically acceptable excipient is incompatible with the active pharmaceutical ingredient, its use in the therapeutic compositions of the disclosure is contemplated. Additional active pharmaceutical ingredients, such as other drugs, can also be incorporated into the described compositions and methods.

As referred to herein, all compositional percentages are by weight of the total composition, unless otherwise specified. As used herein, the word "include," and its variants, is intended to be non-limiting, such that recitation of items in a list is not to the exclusion of other like items that may also be useful in the compositions and methods of this technology. Similarly, the terms "can" and "may" and their variants are intended to be non-limiting, such that recitation that an embodiment can or may comprise certain elements or features does not exclude other embodiments of the present technology that do not contain those elements or features.

Although the open-ended term "comprising," as a synonym of terms such as including, containing, or having, is used herein to describe and claim the invention, the present invention, or embodiments thereof, may alternatively be described using alternative terms such as "consisting of" or "consisting essentially of."

As used herein, the words "preferred" and "preferably" refer to embodiments of the technology that afford certain benefits, under certain circumstances. However, other embodiments may also be preferred, under the same or other circumstances. Furthermore, the recitation of one or more preferred embodiments does not imply that other embodiments are not useful and is not intended to exclude other embodiments from the scope of the technology.

The amount of compositions described herein needed for achieving a therapeutic effect may be determined empirically in accordance with conventional procedures for the particular purpose. Generally, for administering therapeutic agents for therapeutic purposes, the therapeutic agents are given at a pharmacologically effective dose. A "pharmacologically effective amount," "pharmacologically effective dose," "therapeutically effective amount," or "effective amount" refers to an amount sufficient to produce the desired physiological effect or amount capable of achieving the desired result, particularly for treating the disorder or disease. An effective amount as used herein would include an amount sufficient to, for example, delay the development of a symptom of the disorder or disease, alter the course of a symptom of the disorder or disease (e.g., slow the progression of a symptom of the disease), reduce or eliminate one or more symptoms or manifestations of the disorder or disease, and reverse a symptom of a disorder or disease. Therapeutic benefit also includes halting or slowing the progression of the underlying disease or disorder, regardless of whether improvement is realized.

Effective amounts, toxicity, and therapeutic efficacy can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to about 50% of the population) and the ED₅₀ (the dose therapeutically effective in about 50% of the population). The dosage can vary depending upon the dosage form employed and the route of administration utilized. The dose ratio between toxic and therapeutic effects is the therapeutic index and can be expressed as the ratio LD₅₀/ED₅₀. In embodiments, compositions and methods that exhibit large therapeutic indices are preferred. A therapeutically effective dose can be estimated initially from in vitro assays, including, for example, cell culture assays. Also, a dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ as determined in cell culture, or in an appropriate animal model. Levels of the described compositions in plasma can be measured, for example, by high performance liquid chromatography. The effects of any particular dosage can be monitored by a suitable bioassay. The dosage can be determined by a physician and adjusted, as necessary, to suit observed effects of the treatment.

As used herein, "methods of treatment" are equally applicable to use of a composition for treating the diseases or disorders described herein and/or compositions for use and/or uses in the manufacture of a medicaments for treating the diseases or disorders described herein.

SELECTED SEQUENCES

In embodiments, the present disclosure provides for any of the sequence provided herein, including the below, and a variant sequence having at least about 90%, or at least about 93%, or at least about 95%, or at least about 97%, or at

least about 98%, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

SEQ ID NO: 1: MLT mobile element enzyme protein (amino acid sequence of a variant of the hyperactive mobile element enzyme with S at position 8 and C at position 13 (572 amino acids)

MAQHSDYSDDEFKADKLSNYSCDSLEENASTSDEDSSDDEVMVRPRTLRRRISSSSDSESDIEGGREEWSHV
DNPPVLEDFLGHQGLNTDAVINNIEDAVKLFIGDDFFEFLVEESNRYYNQNRNFKLSKSLKWKDITPQEMKK
FLGLIVLMGQVRKDRRDDYWTTEPWTETPYFGKTMTRDRFRQIWKAWHFNNNADIVNESDRLCKVRPVLDFVFP
KFINIYKPHQQLSLDEGIVPWRGRLLFRVYNAGKIVKYGILVRLLCESDTGYICNMEIYCCEGKRLLLETIQTVV
SPYTDWYHIYMDNYNSVANCEALMKNKFRICGTIRKNRGIKPDFQTISLKKGETKFIRKNDILLQVWQSKKP
VYLIISSIHSAEMEESQNIIDRTSKKKIVKPNALIDYNKHMKGVDRAQYLSYYSILRRTVKWKRLAMYMINCAL
FNSYAVYKSVRQRKMGFKMFLKQTAIHWLTDIDIPEDMDIVPDLQVPVSTSGMRAKPPTSPPCRLSMDMRKHTL
QAIVGSGKKKNILRRRCRVCSVHKLRSETRYMCKFCNIPLHKGACFEKYHTLKNY

SEQ ID NO: 2: Codon-optimized MLT mobile element enzyme DNA. Nucleotide sequence of SEQ ID NO: 1 (1719 bp)

ATGGCCAGCACAGCGACTACAGCGACGACGAGTTCTGTGCCGATAAGCTGAGTAACTACAGCTGCCAGACGCGA
CCTGGAACACGCCAGCACATCCGACGAGGACAGCTCTGACGACGAGGTGATGGTGCGGCCAGAACCTGAGAC
GGAGAAGAATCAGCAGCTCTAGCAGCGACTCTGAATCCGACATCGAGGGCGGCCGGAAGAGTGGAGCCACGTG
GACAACCTCCTGTTCTGGAAGATTTCTGGGCCATCAGGGCCTGAACACCGACGCCGTGATCAACAACATCGA
GGATGCCGTGAAGCTGTTTATAGGAGATGATTTCTTTGAGTTCTGGTTCGAGGAATCCAACCGCTATTACAACC
AGAATAGAAACAACCTTCAAGCTGAGCAAGAAAAGCCTGAAGTGAAGGACATCACCCCTCAGGAGATGAAAAAG
TTCCTGGGACTGATCGTTCTGATGGGACAGGTGCGGAAGGACAGAAGGGATGATTACTGGACAACCGAACCTTG
GACCGAGACCCCTTACTTTGGCAAGACCATGACCAGAGACAGATTCAGACAGATCTGGAAAGCCTGGCACTTCA
ACAACAATGCTGATATCGTGAACGAGTCTGATAGACTGTGTAAAGTGGCGCCAGTGTGGATTACTTTCGTGCCT
AAGTTCATCAACATCTATAAGCCTCACCAGCAGCTGAGCCTGGATGAAGGCATCGTGCCCTGGCGGGGCGAGACT
GTTCTTCAGAGTGTACAATGCTGGCAAGATCGTCAAATACGGCATCCTGGTGGCCTTCTGTGCGAGAGCGATA
CAGGCTACATCTGTAATATGGAATCTACTGCGGCGAGGGCAAAAGACTGCTGGAAACCATCCAGACCGTCTGTT
TCCCCTTATACCGACAGCTGCTACCATCTACATGGACAACCTACTACAATTCTGTGGCCAACCTGCGAGGCCCT
GATGAAGAACAAGTTTAGAATCTGCGGCACAATCAGAAAAAACAGAGGCATCCCTAAGGACTTCCAGACCATCT
CTCTGAAGAAGGGCGAAACCAAGTTCATCAGAAAGAACGACATCCTGCTCCAAGTGTGGCAGTCCAAGAAACCC
GTGTACCTGATCAGCAGCATCCATAGCGCCGAGATGGAAGAAAGCCAGAACATCGACAGAACAAAGCAAGAAGAA
GATCGTGAAGCCCAATGCTCTGATCGACTACAACAAGCACATGAAAGGCGTGGACCGGGCCGACCAGTACCTGT
CTTATTACTCTATCCTGAGAAGAACAGTGAATGGACCAAGAGACTGGCCATGTACATGATCAATTGCGCCCTG
TTCAACAGCTACGCCGTGTACAAGTCCGTGCGACAAAGAAAAATGGGATTCAAGATGTTCTGAAGCAGACAGC
CATCCACTGGCTGACAGACGACATTCCCTGAGGACATGGACATTGTGCCAGATCTGCAACCTGTGCCCAGCACCT
CTGGTATGAGAGCTAAGCCTCCACCAGCGATCCTCCATGTAGACTGAGCATGGACATGCGGAAGCACACCCTG
CAGGCCATCGTGGCAGCGGCAAGAAGAAGAACATCCTTAGACGGTGCAGGGTGTGCAGCGTGCACAAGCTGCG
GAGCGAGACTCGGTACATGTGCAAGTTTTTGCAACATTCCTTGCACAAGGGAGCCTGCTTCGAGAAGTACCACA
CCCTGAAGAATTACTAG

SEQ ID NO: 3: PGBD4 Amino Acid Sequence (585 Amino Acids).

MSNPKRSIP	MRDSNTGLEQ	LLAEDSFDES	DFSEIDSDN	FSDSALEADK	50
IRPLSHLESD	GKSSTSSDSG	RSWKWSARAM	IPRQRYDFTG	TPGRKVDVSD	100
ITDPLQYFEL	FFTEELVSKI	TRETNQAAL	LASKPPGPKG	FSRMDKWKDT	150
DNDELKVFFA	VMLLQGIYQK	PELEMFWSTR	PILDTPYLRLQ	IMTGERFLLL	200
FRCLHFNNS	SISAGQSKAQ	ISLQKIKPVF	DPLVNKFSTV	YTPNRNIAVD	250
ESLMLFKGPL	AMKQYLPYTKR	VRFLKLYVL	CESQSGYVWN	ALVHTGPGMN	300
LKDSADGLKS	SRIVLTLVND	LLQGQYCVFL	DNFNISPMLF	RELHQNRDTA	350
VGTARLNRRQ	IPNDLKKRIA	KGTTVARFCG	ELMALKWCDG	KEVTMLSTFH	400

NDTVIEVNNR NGKKTKRPRV IVDYNNMGA VDSADQMLTS YPSEKRRHKV 450
 WYKKFFHLL HITVLNSYIL FKKDNPEHTM SHINFRLALI ERMLEKHHKP 500
 GQQHLRGRPC SDDVTPLRLS GRHFPKSIPA TSGKQNPTGR CKICCSQYDK 550
 LGKKIRKETR YFCAECDVPL CVVPCFEIYH TKNKY 585

SEQ ID NO: 4: PGBD4 Hyperactive Mutant (S8P, G17R, K134K) Amino Acid Sequence (585 Amino Acids).

MSNPRRPIP MRDSNTRLEQ LLAEBSFDES DPSEIDSDN FSDSALEADK 50
 IRPLSHLES D KGSSTSSDSG RSMKWSARAM IPRQRYDFTG TPGRKVDVSD 100
 ITDFLQYFEL FTFEBLVSKI TRETNQAAL LASKPPGPKG FSRMDKWKDT 150
 DNDELKVFFA VMLLQGIVQK PELEMFWSTR PLLDTPYLRLQ IMTGERFLLL 200
 FRCLHFNNS SISAGQSKAQ ISLQIKPVF DFLVNKFSTV YTPNRNIAVD 250
 ESLMLFKGPL AMKQYLPTRK VRFLGLKLYVL CESQSGYVWN ALVHTGPGMN 300
 LKDSADGLKS SRIVLTLVND LLGGQYCVFL DNFNISPMLF RELHQNRD 350
 VGTARLNRKQ IPNDLKKRIA KGTTVARFCG ELMALKWCDG KEVTMLSTFH 400
 NDTVIEVNNR NGKKTKRPRV IVDYNNMGA VDSADQMLTS YPSEKRRHKV 450
 WYKKFFHLL HITVLNSYIL FKKDNPEHTM SHINFRLALI ERMLEKHHKP 500
 GQQHLRGRPC SDDVTPLRLS GRHFPKSIPA TSGKQNPTGR CKICCSQYDK 550
 LGKKIRKETR YFCAECDVPL CVVPCFEIYH TKNKY 585

SEQ ID NO: 5: PGBD4 Hyperactive Mutant (S8P, G17R, K134K) Nucleotide Sequence (1758 bp).

ATGTCAAATC CTAGAAAACG TCCCATTCCT ATGCGTGATA GTAATACCCG TCTCGAACAG 60
 TTGTTGGCTG AAGATTCATT TGATGAATCT GATTTTTCGG AAATAGATGA TTCTGATAAT 120
 TTTTCGGATA GTGCTTTAGA AGCCGATAAG ATCAGGCCTC TGTCCCATTT AGAATCTGAT 180
 GGAAAGAGCT CTACATCAAG TGACTCAGGG CGCTCCATGA AATGGTCAGC TCGTGCTATG 240
 ATTCCACGTC AAAGGTATGA CTTTACCGGC ACACCTGGCA GAAAAGTCGA TGTCACTGAT 300
 ATCAGTAGCC CATTGCAGTA TTTTGAACGT TTCTTTACTG AGGAATTAGT TTCAAAAATT 360
 ACTAGAGAAA CAAATGCCCC AGCTGCCTTG TTGGCTTCAA AGCCACCGGG TCCGAAAGGA 420
 TTTTCGGGAA TGGATAAATG GAAAGACACT GACAATGACG AGCTCAAAGT CTTTTTTGCA 480
 GTAATGTTAC TGCAAGGTAT TGTGCAGAAA CCTGAGCTGG AGATGTTTGG GTCAACAAGG 540
 CCTCTTTTGG ATACACCTTA TCTCAGGCAA ATTATGACTG GTGAAAGATT TTTACTTTTG 600
 TTTTCGGTGCC TGCATTTTGT CAACAATTCCT TCTATATCTG CTGGTCAATC AAAGGCCAG 660
 ATTTCAATTC AGAAGATCAA ACCTGTGTTT GACTTTCTTG TAAATAAATT TTCCACTGTA 720
 TATACTCCAA ACAGAAACAT TGCAGTTGAT GAATCACTGA TGCTGTTCAA GGGGCCATTA 780
 GCTATGAAGC AGTACCTCCC GACAAAACGA GTACGATTTG GTCTGAAGCT ATATGTACTT 840
 TGTGAAAGTC AGTCTGGTTA TGTGTGGAAT GCGCTTGTTT ACACAGGGCC TGGCATGAAT 900
 TTGAAAGATT CAGCGGATGG CCTGAAATCA TCACGCATTG TTCTTACCTT GGTCAATGAC 960
 CTTCTTGGCC AAGGGTATTG TGTCTTCCTC GATAACTTTA ATATATCTCC CATGCTTTTC 1020
 AGAGAATTAC ATCAAAATAG GACTGATGCA GTTGGGACAG CTCGTTTGAA CAGAAAACAG 1080
 ATTCCAAATG ATCTGAAAAA AAGGATTGCA AAGGGGACGA CTGTAGCCAG ATTCTGTGGT 1140
 GAACCTATGG CACTGAAATG GTGTGACGGC AAGGAGGTGA CAATGTTGTC AACATTCAC 1200
 AATGATACTG TGATTGAAGT AAACAATAGA AATGGAAAGA AAACATAAAG GCCACGTGTC 1260
 ATTGTGGATT ATAACGAGAA TATGGGAGCA GTGGACTCGG CTGATCAAAT GCTTACTTCT 1320
 TATCCATCTG AGCGCAAAAG ACACAAGGTT TGGTATAAGA AATTCTTTCA CCATCTTCTA 1380
 CACATTACAG TGCTGAAGTC CTACATCCTG TTCAAGAAGG ATAATCCTGA GCACACGATG 1440
 AGCCATATAA ACTTCAGACT GGCATTGATT GAAAGAATGC TGGAAAAGCA TCACAAGCCA 1500
 GGGCAGCAAC ATCTTCGAGG TCGTCCTTGC TCCGATGATG TCACACCTCT TCGTCTGTCT 1560
 GGAAGACATT TCCCCAAGAG CATACCAGCA ACGTCCGGGA AACAGAATCC AACTGGTCGC 1620
 TGCAAAATTT GCTGCTCCCA ATACGACAAG GATGGCAAGA AGATCCGGAA AGAAACGCGC 1680
 TATTTTGTG CCGAATGTGA TGTCCGCTT TGTGTTGTTT CGTGCTTTGA AATTACCAC 1740
 ACGAAAAAAA ATTATTAA 1758

SEQ ID NO: 6: PGBD1 Amino Acid Sequence (809 Amino Acids).

MYEALPGPAP ENEDGLVKVK EEDPTWEQVC NSQEGSSHTQ EICRLRFRHF CYQEAHGPQE 60
 ALAQLRELCH QWLRPEMHTK EQIMELLVLE QFLTILPKEL QPCVKTYPLE SGEEAVTVLE 120
 NLETGSGDTG QQASVYIQG DMHPMVAEYQ QVSLQCQLQ LLPGITTLKC EPPQRPQGNP 180
 QEVSGPVPHG SAHLQEKNP DRKAVVPVFN VRSQTLVKTE EETAQAVAAE KWSHLSLTRR 240
 NLCCNSAQET VMSLSPMTEE IVTKDRLFKA KQETSEEMEQ SCEASGKPNR ECAPQIPCST 300
 PIATERTVAH LNTLKDHRHPG DLWARMHISS LEYAAGDITR KGRKKDKARV SELLQGLSFS 360

GDSDEVKEDNE	PEIQPAQKKL	KVSCFPPEKSW	TKRDIKPNFP	SWSALDSGLL	NLKSEKLNVP	420
ELFELFFDDE	TFNLIVNETN	NYASQKNVSL	EVTVQEMRCV	PGVLLLSGFM	RHPRREMYWE	480
VSDTDQNLVR	DAIRRDREFL	IFSNLHFADN	GHLQKDKFT	KLRPLIKQMN	KNFLLYAPLE	540
EYYCFDKSMC	ECFDSQQLN	GKPIRIGYKI	WCGTTTQGYL	VWFEPYQRES	TMKVDEDEPD	600
GLGGNLVMNF	ADVLLERGQY	PYHLCFDSFF	TSVKLLSALK	KKGVRATGTI	RENRTKCPPL	660
MNVEHMKMK	RGYDFRIEE	NNELILCRWY	GDGIISLCSN	AVGIEPVNEV	SCCDADNEEI	720
PQISQPSIVK	VYDECKEGVA	KMDQIISKYR	VRIRSKKWYS	ILVSYMIDVA	MNNAWQLHRA	780
CNPGASLDPL	DFRRFVAHFY	LEHNAHLS				809

SEQ ID NO: 7: PGBD2 Amino Acid Sequence (592 Amino Acids).

MASTRDVIA	GRGIHVKVS	AKLLEVLNAM	EEEEENNRE	EIFIAPPDNA	AGEFTDEDSG	60
DEDSQRGAHL	PGSVLHASVL	CEDSGTGEDN	DDLELQPAKK	RQKAVVKPQR	IWTKRDIRPD	120
PGSWTASDPH	IEDLSQELS	PVGLFELEFD	BGTINFIVNE	TNRYAWQKNV	NLSLTAQELK	180
CVLGILILSG	YISYPRRRMF	WETSPDSHHH	LVAADAIRDR	FELIFSYLHF	ADNNELDASD	240
KFAKVRPLII	RMNCFQKHA	PLEEFYSFGE	SMCEYFGHRG	SKQLHRGKPV	RLGYKIWCGT	300
TSRGYLVWFE	PSQGTFTKP	DRSLDLGGS	VIKFDALQE	RGFLPYHIF	DKVFTSVKLM	360
SILRKKGVKA	TGTVREYRTE	RCPLKDPKEL	KMKRGSEFY	KVDESEIIV	CRWHDSSVNV	420
ICSNAVGLIEP	VRLTSRHS	AKTRTQVHQP	SLVKLYQEKV	GGVGRMDQNI	AKYVKIRGM	480
KWYSSFIGYV	IDAALNNAWQ	LHRIQCQDAQ	VDLAFRRYI	ACVYLESNAD	TTSQGRRSRR	540
LETESRPDMI	CHWIIHQDKR	TRCALCHSQT	NTRCEKCQKG	VHAKCFREYH	IR	592

SEQ ID NO: 8: PGBD3 Amino Acid Sequence (593 Amino Acids).

MPRTLSEHEI	TDLLETDDSI	EASAIVIQPP	ENATAPVSDE	ESGDEEGGTI	NNLPGLLHT	60
AAYLIQDGS	AESDSDPSY	APKDDSPDEV	PSTFTVQPP	PSRRRKMTKI	LCKWKKADLT	120
VQPVAGRFTA	PFNDFFTVMR	TPTEILELFL	DDEVIELIVK	YSNLYACSKG	VHLGLTSSEF	180
KCFLGIIFLS	GYVSVPRRRM	FWEQRTDVHN	VLVSAAMRRD	RFETIFSNLH	VADNANLDPV	240
DKFSKLRPLI	SKLNERCMKF	VPNETYFSFD	EFMVPIYFGRH	GCKQFIRGKP	IRFGYKFWCG	300
ATCLGYICWF	QPYQGNPNT	KHEEYGVGAS	LVLQFSEALT	EAHPGQYHFV	FNNFPTSIAL	360
LDKLSSMGHQ	ATGTVRKDH	DRVPLESDVA	LKKKERGTFD	YRIDGKGNIV	CRWHDNSVVT	420
VASSGAGIHF	LCLVSRYSQK	LKKKIQVQPF	NMIKVYNQFM	GGVDRADENI	DKYRASIRGK	480
KWYSSPLLF	FELVLQNAWQ	LHKTIDEKPV	DFLEFRRRVV	CHYLETHGHP	PEPGQKGRPQ	540
KRNIDSRYDG	INHVIKQKGK	QTRCAECHKN	TTFRCEKCDV	ALHVKCSVEY	HTE	593

SEQ ID NO: 9: PGBD5 Amino Acid Sequence (524 Amino Acids).

MAEGGGGARR	RAPALLEAAR	ARYESLHISD	DVFGESGPD	GGNPFYSTSA	ASRSSSAASS	60
DEREPPPGPP	GAAPPPPRAP	DAQPEEDEEA	GAGWSAALRD	RPPPRFEDTG	GPTRKMPPSA	120
SAVDFQQLFV	PDNVLKMMV	QTNMYAKKFO	BRFGSDGAWV	EVTLTETMAF	LYMISTSSIS	180
HCESVLSIWS	GGFYSNRSLA	LVMQARFEK	ILKYFHVVA	RSSQTHGLY	KVQPFLLSLQ	240
NSFDSAPRPS	QTQVLHEPLI	DEDPVFIATC	TERELRKRKK	RKFSLWVRQC	SSTGPPIQIY	300
VHLKEGGGPD	GLDALKNKPQ	LHSMVARSLC	RNAAGKNYII	FTGPSITSLT	LFEEFEKQGI	360
YCCGLLRARK	SDCTGLPLSM	LTNPATPPAR	GQYQIKMKGN	MSLICWYNKG	HFRFLTNAYS	420
PVQQGVIIKR	KSGEIIPCPLA	VEAFAAHLISY	ICRYDDKYSK	YFISHKPNKT	WQQVPWFALS	480
IAINNAYIILY	KMSDAYHVKR	YSRAQFGERL	VRELLGLEDA	SPTH		524

SEQ ID NO: 10: *Myotis lucifugus* (Wild-type) Amino Acid Sequence with Hyperactive Mutations (S8P, C13R, N125K) 572 Amino Acids.

MSQHSQVDD	EFRADEKLSNY	SCDSLEENAS	TSDEDSDDDE	VMVRFTLR	RRISSSSSDS	60
ESDIEGGREE	WSHVDNPFVL	EDFLGHQGLN	TDVAVINNIED	AVKLFIGDDF	FEFLVESNR	120
YYNQKRNFK	LSKSLKWKD	ITPQEMKKFL	GLIVLMGQVR	KDRDDYWT	EPWTETPYFG	180
KTMTRDRFRQ	IWKAWHFNNN	ADIVNESDRL	CKVRPVDYF	VPKFINIYKP	HQQLSLDEGI	240
VPWRGRLLFFR	VYNAGKIVKY	GILVRLLCES	DTGYICNMEI	YCGEGKRLLE	TIQTVVSPYT	300
DSWYHIYMDN	YNSVANCEA	LMKNKFRICG	TIRKNRGIPK	DFQTISLKKG	ETKFKRNDI	360
LLQVWQSKKP	VYLISSTHSA	EMEESQNIDR	TSKKKIVKPN	ALIDYNKHMK	GVDRADQYLS	420
YYSILRRITVK	WTKRLAMYMI	NCALFNYSYAV	YKSVRQRKMG	FKMFLKQTAI	HWLTDIPED	480
MDIVPDLQFV	PSTSGMRAPK	PTSDPPCRLS	MDMRKHTLQA	IVGSGKKKNI	LRRCRVCSVH	540
KLRSETRYMC	KFCNIPLHKG	ACEKYHTLK	NY			572

SEQ ID NO: 11: *Myotis lucifugus* Corrected Amino Acid Sequence with Hyperactive Mutations (S8P, C13R) 572 Amino Acids.

MAQHSDDYPDEFRADKLSNYSKSDLENASTSDESDSDEVMPVRPTLRRRRISSSSSDSESDIEGGREWSHVDNPPVLEDPLGH
QGLNTDAVINNIEDAVKLFIGDDFFFLVEBSNRYYNQNRNNFKLSKSLKWKDITPQEMKKFLGLIIVLMGQVRKDRRDDYWTTEP
WTETPYFGKTMTRDRFRQIWKAWHPNNNADIYNESDRLCKVRPVLDFVPEKFINIYKPHQQLSLDEGIVPWRGRLFFRVYNAGKIV
KYGILVRLLCESDTGYICNMEIYCGEGKRLLLETIQTVVSPYTDWYHIYMDNYNSVANCEALMKNKFRICGTIRKNRGIPKDFQT
ISLKKGETKFIRKNDILLQVWQSKKPVYLISSISHAEMERSQNIIDRTSKKKIVKPNALIDYNKMHKMGVDRADQYLSYYSILRRTVK
WTKRLAMYMINCALFNSYAVYKSVRQRKMGFKMFLKQTAIHWLTDDIPEDMDIVPDLQPVPSSTSGMRAPPTSDPPCRLSMDMRKH
TLQAIVSGGKKKNILRRCRVCSVHKLRSETRYMCKFCNIPLHKGACFEKYHTLKNY

SEQ ID NO: 12: *Pteropus vampyrus* Left End Sequence

Sequence 381 bp.

TTAACCCATT	TCCTGTTTGC	CCCGAGAATA	CTCACCAGCG	GCACTTGCAG	CTGCAGCGTT	60
TACCCCGAGA	TAACCTCGTCG	ATTACAGTCC	TAACCTTACC	CCCAAAGTTT	GCCATGAAAT	120
ATCTCGCTTT	TATTATTATT	TTCGCATCGC	TCTAGTATAT	CGATAGTCTT	TGGAAACAAA	180
TGACATCATT	CTATTACAG	CATTCTGTTT	TTAGTAGTGG	TATTTCATT	TACAAAATAT	240
AGTAATTTTC	TATCGCTGAA	AATGTCAAAT	CCTAGAAAAC	GTAGCATTCC	TACATGTGAT	300
GTTAACTTCG	TTCTCGAACA	GTTGTAGCC	GAAGATTCAT	TTGATGAATC	CGATTTTTC	360
GAAATAGACG	ATTCTGATGA	T				381

SEQ ID NO: 13: PGBD4 Left End Nucleotide Sequence

Sequence 373 bp.

TTAACTCAT	TCTCCTTAGC	CCCGAGATTA	CGCGCTGCTG	TGCCTGCGAC	TGCAGCGTTT	60
ACGCCGAGAT	AACTCGTGA	TTACAGTGCC	AACCTTACTC	CCAAAGTTTG	CCACGAAATA	120
TCTCGCTTCT	GTTATTTTCG	CATGGTTCTG	GTATATTGAC	TTTTGAAACA	AAAGACATCA	180
TTCTGTTTAT	AGCATTCTGT	TTTTAGTAGT	GGGATTTCCA	TCTACAAAAT	ATAGTAATTC	240
TCGATCGCTG	AAATGCAAAA	TCCTAGAAAA	CGTAGCATTC	CTATGCGTGA	TAGTAATACC	300
GGTCTCGAAC	AGTTGTTGGC	TGAAGATTCA	TTTGATGAAT	CTGATTTTTC	GGAAATAGAT	360
GATTCTGATA	ATT					373

SEQ ID NO: 14: MER75 Left End Nucleotide Sequence

Sequence 344 bp.

TTAACCCTTT	TCCCCTTTGC	CCCGAGAATA	CTCGCCGGCG	GCGCTTGCGG	CTGCAGCGTT	60
TACCCCGAGA	TAACCTTGCC	ACGAAATATC	TCGCTTTTAT	TATTATTTTC	GCATCGCTCT	120
AGTATATCGA	CTTTGAAAC	AAAAGACATC	ATTCTATTTA	TAGCATTCTG	TTTTTAGTAG	180
TGGTATTTTC	ATTTACAAA	TATAGTAATT	CTCGATCGCT	GAAAATGTCA	AATCCTAGAA	240
AACGTAGCAT	TCCTACGCGT	GATGTTAACA	TCGTTCTCGA	ACAGTTGTTG	GCCGAAGATT	300
CATTTGATGA	ATCCGATTTT	TCCGAAATAG	ACGATTCTGA	TGAT		344

SEQ ID NO: 15: MER75B Left End Nucleotide Sequence

Sequence 91 bp.

TTAACCCTTT	TCCCCTTTGC	CCCGAGAATA	CTCTGTCTC	TAATCCTAAT	GTAACATCAT	60
ATACATTTCT	GTTACATTAG	GATTAGAGAC	A			91

SEQ ID NO: 16: MER75A Left End Nucleotide Sequence

Sequence 32 bp.

TTAACCCTTT	TCCCCTTTGC	CCCGAGAATA	CT		32
------------	------------	------------	----	--	----

SEQ ID NO: 17: *Pteropus vampyrus* Right End Sequence

Sequence 171 bp.

TAGGATTAGA	GACAAGTTCT	GTTTAGAAAT	AACTCCAAGA	ACAGTTTTTA	TATTTTATTT	60
TCACATTGAA	AACGAGTCAG	ATTGCTTCA	GCCTCAAAGA	GCATGTTTAT	GTAAATTA	120

ATTAACGCTG GCAGCGAGCT GCACTTTTTT TCTAACGGG AAATGGGTTA A 171

SEQ ID NO: 18: PGBD4 Right End Nucleotide Sequences

Sequence 176 bp.

CCTGGGATTA	TAGGCATGAG	CCACTGCGCC	TAGCACCAAG	AACAGTTTTT	ATATTTTATT	60
TTTACATTGA	AAATCAGTCA	GATTTGCTTC	AGCCTCAAAG	AGGGTGTTTA	TGTAAAACTA	120
AATGAGTGCA	GGCAGCGAGC	TACACTTTTT	TTTTTCCTAA	ATGGAAAATG	GGTTAA	176

SEQ ID NO: 19: MER75 Right End Nucleotide Sequences

Sequence 178 bp.

TCAGACGATT	CTGATGTTAG	TTCTGTTTAG	AAATAACTCC	AAGAACAGTT	TTTATATTTT	60
ATTTTCACAT	TGAAAATCAG	TCAGATTTGC	TTTACGCTCA	AAGAGCGTGT	TTATGTAAAA	120
TTAAATGAGC	GCTGGCAGCG	AGCTGCACCT	TTTTTTTTCT	AAACGGGAAA	AGGGTTAA	178

SEQ ID NO: 20: MER75B Right End Nucleotide Sequences

Sequence 160 bp.

AGTTCTGTTT	AGAAAATACT	CCAAGAACAG	TTTTTATATT	TTATTTTCAC	ATTGAAAATC	60
AGTCAGATTT	GCTTCAGCCT	CAAAGAGCGT	GTTTATGTAA	AATTAAATGA	GCGCTGGCAG	120
CGAGCTGCAC	TTTTTTTTTT	CTAAACGGGA	AAAGGGTTAA		160	

SEQ ID NO: 441: MER75A Right End Nucleotide Sequences

Sequence 46 bp.

CGCTGGCAGC	GAGCTGCACT	TTTTTTCTAA	ACGGGAAATG	GGTTAA	46
------------	------------	------------	------------	--------	----

SEQ ID NO: 429: Extended *Pteropus vampyrus* Nucleotide Sequence* 2210 BP

CCCATTTCCCT	GTTTGCCCCG	AGAATACTCA	CCAGCGGCAC	TTCAGCTGCG	AGCGTTTACC	60
CCGAGATAAC	TCGYCGATTA	CAGTCCTAAC	CTTACCCCCA	AAGTTTGCCA	TGAAATATCT	120
CGCTTTTATT	ATTATTTTCG	CATCGCTCTA	GTATATCGAT	AGTCCTTGGA	AACAAATGAC	180
ATCATTTNAT	TTACAGCATT	CTGTTTTTAN	TAGTGGTATT	TCCATTTACA	AAATATAGTA	240
ATTTTCTATC	GCTGAAAATG	TCAAATCCTA	GAAAACGTAG	CATTCCTACA	TGTGATGTTA	300
ACTTCGTTCT	CGAACAGTTG	TTAGCCGAAG	ATTCATTTGA	TGAATCCGAT	TTTTCCGAAA	360
TAGACGATTC	TGATGATTTT	TCGGATAGTG	CTTCGGAAGA	CTATACGGTC	AGGCCTCCGT	420
CCGATTCGGA	ATCTGATGGA	AATAGCCCTA	CATCAGCTGA	CTCGGGTCGC	GCTCTGAAAT	480
GGTCAACTCG	TGTTATGATT	CCACGTCAAA	GCTATGACTT	TACCGGCACA	CCTGGCAGAA	540
AAGTTGATGT	CAGTGATAAC	ACTGACCCAC	TGCAGTATTT	TGAAGTGTTC	TTTACTGAGG	600
AATTAGTTTC	AAAAATTACC	AGTGAAATGA	ATGCCCAAGC	TGCCTTGTTC	GCTTCAAAGC	660
CACCTGGTCC	GAAAGGATTT	TCGCGAATGG	ATAAATGGAA	AGACACTGAC	AATGATGAAC	720
TGAAAGTCTT	TTTTGCAGTA	ATGTTACTGC	AAGGTATTGT	GCAGAAACCT	GAGCTGGAGA	780
TGTTTTGGTC	GACAAGGCCT	CTTTTGGATA	TACCTTATCT	CAGGCAAATT	ATGACTGGTG	840
AAAGATTTTT	ACTTTTGCTT	CGGTGCCTGC	ATTTTGTCAA	CAATTCCTCC	ATATCCGCTG	900
GTCAATCAAA	GGCCCAGATT	TCATTGCAGA	AGATCAAACC	TGTGTTCCGAC	TTTCTGTGTA	960
ATAAGTTTTC	AACGTATAT	ACTCCAAACA	GAAACATTGC	AGTCGATGAA	TCACGTGATG	1020
TGTTCAAGGG	GCGGTAGCT	ATGAAGCAGT	ACATCCCGAC	GAAATGCGCA	CGATTGGGTC	1080
TCAAGCTNTA	TGTACTTTGT	GAAAGTCAAT	CTGGTTACGT	GTGGAATGCG	CTTGTTCACA	1140
CAGGGCCCG	TATGAATTTG	AAAGATTCAG	CTGATGGTCT	GAAATCGTCA	TGCATTGTTC	1200
TTACCTTGGT	CAATGACCTT	CTTGGCCCAAG	GATATTGTGT	CTTCCTCAAT	AACTTTTATA	1260
CATCTCCCAT	GCTTTTCAGA	GAATTACATC	AAAACAGGAC	TGATGCAGTT	GGGACAGCTC	1320
GTTTGAACAG	AAAACAGATG	CCAAATGATC	TGAAAAAAG	GATTGCAAAG	GGGACGACTG	1380
TAGCCAGATT	CTGTGGTGAA	CTTATGGCAC	TGAAATGGTG	TGACAAGAAG	GAGGTGACAA	1440
TGTTGTCAAC	ATTCCACAAT	GATACTGTGA	TTGAAGTAGA	CAACAGAAAT	GGAAAGAAAA	1500
CTAAGAAGCC	ATGTGTCAAT	GTGGATTATA	ACGAGAATAT	GGGAGCAGTG	GACTCGGCTG	1560
ATCAGATGCT	CACCTCTTAT	CCAACTGAGC	GCAAAAGGCA	CAAGTTTTCG	TATAAGAAAT	1620
TCTTTCGCCA	CCTTCTAAAC	ATTACAGTGC	TGAACCTCTA	CATCCTGTTC	AAGAAGGACA	1680
ATCCTGAGCA	CACGATCAGC	CATGTAAACT	TCAGACTGAC	GTTGATTGAA	AGAATGCTGG	1740
AAAAGCATCA	CAAGCCAGGG	CAGCAACGTC	TTCGAGGTGC	TCCGTGCTCT	GATGATGTCA	1800

```

CACCTCTTCG CCTGTCTGGA AGACATTTCC CCAAGAGCAT ACCACCAACA TCAGGGAAAC 1860
AGAATCCAAC TGGTCGCTGC AAAGTTTGCT GCTCGCACGA CAAGGATGGC AAGAAGATCC 1920
GGAGAGAAAC GTTATATTTT TGTGCGGAAT GTGATGTTCC GCTTTGTGTT GTTCCGTGCT 1980
TTGAAATTTA CCACACGAAA AAAAATTATT AAATACTGAT CATCATATAC ATTTCTGTTA 2040
CATTAGGATT AGAGACAAGT TCTGTTTAGA AATAACTCCA AGAACAGTTT TTATATTTTA 2100
TTTTACATT GAAAACCAAGT CAGATTTGCT TCAGCCTCAA AGAGCATGTT TATGTAAAT 2160
TAAATTAACG CTGGCAGCGA GCTGCACTTN TTTTCTAAAC GGGAAATGGG 2210

```

SEQ ID NO: 430: Extended *Pteropus vampyrus* Amino Acid Sequence 584 Amino Acids.

```

MSNPKRSIP TCDVNFVLEQ LLAEDSFDES DFSEIDDSDD FSDSASEDYT VRPPSDSESD 60
GNSPTSADSG RALKWSTRVM IPRQRYDFTG TPGRKVDVSD TTDPLQYFEL FFTEELVSKI 120
TSEMNAQAAL LASKPPGPKG FSRMDKWKDT DNDELKVFFA VMLLQGIVQK PELEMFWSTR 180
PLLDIPYLEQ IMTGERFLLL LRCLHFVNNS SISAGQSKAQ ISLQKIKPVF DFLVNKFSTV 240
YTPNRNIAVD ESLMLFKGRL AMKQYIPTKC ARFGLKLYVL CESQSGYVWN ALVHTGPSMN 300
LKDSADGLKS SCIVLTLYND LLGQGYCVFL NNFYTSPLMF RELHQNRDA VGTARLNRKQ 360
MPNDLKKRIA KGTTVARFCG ELMALKWCDK KEVTMLSTFH NDTVIEVDNR NGKTKKPCV 420
IVDYNENMGA VDSADQMLTS YPTERRHKF WYKFPFRHLL NITVLNSYIL FKKDNPEHTI 480
SHVNFRLTLI ERMLEKHKP GQQLRGRPC SDDVTPLRLS GRHFPKSIPP TSGKQNPGR 540
CKVCCSHDKD GKKIRRETLY FCAECDVPLC VVPCFEIYHT KKNY

```

SEQ ID NO: 431: A MLT left end (5' to 3') is as follows:

TTAACACTTGGATTGCGGGAAACGAGTTAAGTCGGCTCGCGTGAATTGCGCGTACTCCGCGGGAGCCGTC
TTAACTCGGTTTATATAGATTTGCGGTGGAGTTCGGGAAACGTGTAAACTCGGGCCGATTGTAAGTTCGCT
ATTACCAAATATTTGTT

SEQ ID NO: 432: A MLT right end (5' to 3') is as follows:

AATTATTTTATGTACTGAATAGATAAAAAAATGCTGTGATTGAATAAAATTTTCATTTTTTACAC
AAGAAACCGAAAATTTTCAATTCGAACCCATACTTCAAAAGATATAGGCATTTTAAACTAA
CTCTGATTTTTCGCGGGAAACCTAAATAATTGCCCGCGCCATCTTATATTTTGGCGGGAAATTC
ACCCGACACCGTGGTGTTAA

SEQ ID NO: 433: *Trichnoplusia ni*

```

1 MGSSLDDEHI LSALLQSDDE LVGEDSDSEI SDHVSEDDVQ SDTEEAFIDE VHEVQPTSSG
61 SEILDEQNVI EQPGSSLASN KILTLPQRTI RGKNKHCWST SKSTRRSRVS ALNIVRSQRG
121 PTRMCRNIYD PLLCFKLFFT DEIISEIVKW TNAEISLKRR ESMTGATFRD TNEDEIYAFF
181 GILVMTAVRK DNHMSTDDL FDRSLSMVYVS VMSRDRFDFL IRCLEMDDKS IRPTLRENDV
241 FTPVRKIWDL FIHQCIQNYT PGAHLTIDEQ LLGFRGRCPF RMYIPNKPSK YGIKILMMCD
301 SGT KYMINGM PYLGRGTQTN GVPLGEYYVK ELSKPVRGSC RNITCDNWFT SIPLAKNLLQ
361 EPYKLTI VGT VRSNKREIPE VLKNSRSRPV GTSMFCFDGP LTLVSYKPKP AKMVYLLSSC
421 DEDASINEST GKPQVMYYN QTKGGVDTLD QMCSVMTCR KTNRWPMALL YGMINIACIN
481 SFIIYSHNVS SKGEKVQSRK KPMRNLYMSL TSSFMRKRLE APTLKRYLRD NISNILPNEV
541 PGTSDDSTEE PVTKKRTYCT YCP SKIRKA NASCKCKKV ICREHNIDMC QSCF

```

SEQ ID NO: 434: *Pteropus vampyrus*

```

1 MSNPKRSIP TCDVNFVLEQ LLAEDSFDES DFSEIDDSDD FSDSASEDYT VRPPSDSESD
61 GNSPTSADSG RALKWSTRVM IPRQRYDFTG TPGRKVDVSD TTDPLQYFEL FFTEELVSKI
121 TSEMNAQAAL LASKPPGPKG FSRMDKWKDT DNDELKVFFA VMLLQGIVQK PELEMFWSTR
181 PLLDIPYLEQ IMTGERFLLL LRCLHFVNNS SISAGQSKAQ ISLQKIKPVF DFLVNKFSTV
241 YTPNRNIAVD ESLMLFKGRL AMKQYIPTKM NLKDSADGLK

```

SEQ ID NO: 435: *Myotis myotis* ("2a")

1 MDLRCQHTVL SIRESRGLLP NLKMKTSRMK KGDIIFSRKG DILLLAWKDK RVVRMISIH
 61 TSVSTTGKKN RKTGENIVKP ACIKEYNAHM KGVDRADQFL SCCSILRKMM KWTKKVVLVYL
 121 INCGLFNSFR VYNVLNPQAK MKYKQFLLSV ARDWIMDDNN EGSPEPETNL SSPSPGGARR
 181 APRKDPKRL SGDMKQHEPT CIPASGKKKF PTRACRVCAH GKRSESRYLK KFCLVPLHRG
 241 KCFTQYHTLK KY

SEQ ID NO: 436: *Myotis myotis* ("1")

1 MKAFLGVILN MGVLNHPNLQ SYWSMDFESH IPFFRSVFKR ERFLQIFWML HLKNDQKSSK
 61 DLRTREKVN CFLSVLEMKF RERFCPGREI AVDEAVVGFK GKIHFITYNP KKPTKWGIRL
 121 YVLSDSKCGY VHSFVPYYGG ITSETLVPRD LPFTSRIVLE LHERLKNSVP GSQGYHFFTD
 181 RYYTSVTLAK ELFKETHLT GTIMPNRKN PPVIKHQKLG KGEIVAFRDE NVMLLAWKDK
 241 RIVTLSTWDS ETESVERRVG GGKEIVLKP VVTNYTKFMG GVDIADYTST YCFMRKTLKW
 301 WRTLFFWGLE VSVVNSYILY KECQKRKNEK PITHVKFIRK LVHDLVGEFR DGTLTSGRL
 361 LSTNLEQRLD GKLHIITPHP NKKHKDCVVC SNRKIKGGRR ETIYICETCE CKPGLHVGE
 421 FKKYHTMKNY RD

SEQ ID NO: 437: *Myotis lucifugus* ("2")

1 MPSLRKRKET NETDTLPEVF NDNLS DIPSE IEDADDCFDD SGDDSTDSTD SEIIRPVRRK
 61 KVAVLSSDS TDEATDNCWS EIDTPPRLQM FEHAGVTTT PSQCDSVPSV TNLFFGDELF
 121 EMLCKELSNY HDQTAMKRKT PSRTLKWSVP TQDKIKKFLG LIILMGQTRK DSLKDYWSTD
 181 PLICTPIFPQ TMSRHRFEQI WTFWHFNDNA KMDSRSGRLF KIQPVLDYFL HKFRTIYKPK
 241 QQLSLDEGMI PWRGRFKPRT YNPAKITKYG LLVRMVCESD TGYICSMEIY TAEGRLQET
 301 VLSVLGPYLG IWHHIYQDNY YNATSTAELL LQNKTRVCGT IRESRGLPPN LEMKTSRMKK
 361 GDIIFSRKGD ILLLAWKDKR VVRMISTIH TSVSTTGKKN RKTGENIVKP TCIKEYNAHM
 421 KGVDRADQFL SCCSILRKTMM KWTKKVVLVYL INCGLFNSFR VYNVLNPQAK MKYKQFLLSV
 481 ARDWITDDNN EGSPEPETNL SSPSPGGARR APRKDPKRL SGDMKQHEPT CIPASGKKKF
 541 PTRACRVCAH HGRSESRYL CKFCLVPLHR GKFTQYHTL KKYMDLRCQH TVLSTVGRGY
 601 SVLARFKPRT NERTGSSHCH VQVPAGGQGP PSTIIANGCG CKLEPMVRTR SPTCLVIEFG
 661 CM

SEQ ID NO: 438: *Myotis myotis* ("2")

1 MPSLRKRKET NETDTLPEVF NDNLS DIPSE IEDADDCFDD SGDDSTDSTE SEIIRPVRRK
 61 KVAVLSSDSN TDEATDNCWS EIDTPPRLQM FEHAGVTTT PSQCDSVPSV TNLFFGDELF
 121 EMLCKELSNY HDQTAMKRKT PSRTLKWSVP TQDKIKKFLG LIILMGQTRK DSWKDYWSTD
 181 PLICTPIFPQ TMSRHRFEQI WTFWHFNDNA KMDSRSGRLF KIQPVLDYFL HKFRTIYKPK
 241 QQLSLDEGMI PWRGRFKPRT YNPAKITKYG LLVRMVCESD TGYICSMEIY TAEGRLQET
 301 SVLGPYLG IWHHIYQDNYN ATSTAELLQ NKTRVCGTIR ESRGLPPNLK MKTSRMKKGD
 361 IIFSRKGDIL LLAWKDKRVV RMISTIHDT VSTTGKKNRK TGENIVKPTC IKEYNAHMKG
 421 VDRADQFLSC CSILRKTTKW TTKVVLVYLIN CGLFNSFRVY NILNPQAKMK YKQFLLSVAR
 481 DWITDDNNEG SPEPETNLSS PSSGARRAP RKDQPKRLSG DMKQHEPTCI PASGKKKFPT
 541 ACRVCAAHGK RSESRYLKRF CFVPLRGKCF MYHTLKKYSE LFSLVVSKI QNVIIYKTTK
 601 VYMRVYMRSH CPLSFLVFAP SVKDRSRVFS FFTRHLLWTL DVNTLSCPHR MKRSHWWKPC
 661 RSIYEXLYNC TNP

SEQ ID NO: 439: *Myotis myotis* ("2b")

1 MDLRCQHTVL SIRESRGLPP NLKMKTSRMK KGDIIFSRKG DILLLAWKDK RVVRMISTIH
 61 DTSVSTTGKK NRKTGENIVK PACIKEYNAH MGVDRADQF LSCCSILRKT MKWTKKVVLVYL
 121 LINCGLFNSF RVYNVLNPQA KMKYKQFLLS VARDWITDDN NEGSPEPETN LSSPSPGGAR
 181 RAPRKDPKRL LSGDMKQHEP TCIPASGKKK FPTRACRVCA AHGKRSESRYL LCKFCLVPLH

241 RGKCFYQYHT LKKY

TTAA (SEQ ID NO: 440)

SEQ ID NO: 816: Tet-operator-containing AAV2 P40 promoter with a modified AAV2 P40 TATA element

gaaggtaaccaagcaggaagtaaacagacttttccgggtgggcaaaaggatcacgtgggtgaggaggagcatgaattctacgcaaaaaggg
tggagccaagaaaagaccgccccagtgacgcaggtatataagtgagcccCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATA
GAGACCaaCgggtgagcagagtcagtl ggcagccat cgCcgtaagagcggaagct togtatcaact tgcagacag

SEQ ID NO: 817: Modified AAV2 P40 intron_2

GggTgaGTaCc aaCgtgttc TaCgtgtttc cgtcaAaca Cgogaaga Ttgaatcaga attcaaatCt cgtgttact caccgacCga gactgtt
gtgttt cccgtgtcag aatcaaacccgtttctgtc gtcaagg cgtatcagaa actgtgtac attcatcata toTgggaaagggtccagac gctgtcactg
cctgcatct ggtcaCgtg gatttggCtaacgtcatcttccacCaiTT aCCa tttT tcaggt

SEQ ID NO: 818: Modified AAV2 P40 intron_1 (a at position 276 substituted for C), in which aataaa is changed to Cataaa:

gtacaaaaac aaatgttctc gtaacgtgggcaatgaatctg atctgtttc cctgcagaca atgogagaga atgaatcaga attcaaatat cgtgttact
caccgacaga aagactgttt agagtgttt cccgtgtcag aatcaaacccgtttctgtc gtcaaaaagg cgtatcagaa actgtgtac attcatcata
tcaigggaagggtccagac gctgtcactg cctgcatct ggtcaatgtg gatttggatg actgcatct
tgaacCataa atga ttaaaa tcaggt atgggtgc cgtgtgtatctccag

SEQ ID NO: 819: Modified AAV2 P40 intron_2, which contains a-C change and the potential translation start site ATGs have been changed to either Ctg, aCg, or Ttg:

gtacaaaaac aaCgtgttc gtaacgtgggCtgaatctg aCgtgtttc cctgcagaca Ctgogagaga Ttgaatcaga attcaaatCt cgtgttact
caccgacaga aagactgttt agagtgttt cccgtgtcag aatcaaacccgtttctgtc gtcaaaaagg cgtatcagaa actgtgtac attcatcata
toTgggaaagggtccagac gctgtcactg cctgcatct ggtcaCgtg gatttggCtg actgcatct
tgaacCataa aCga ttaaaa tcaggt atgggtgc cgtgtgtatctccag

SEQ ID NO: 820: tetO-containing P40 promoter with a modified TATA element (AAV2)

GAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATGAA
TTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCAGTGACGCAGGTATATAAGTGAGCCCCCTCTC
CCTATCAGTGATAGAGATCTCCCTATCAGTGATGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGA
CGTCAGACGCGGAAGCTTCGATCAACT ACGCAGACAG

SEQ ID NO: 821: AAV9 P40 intron

gtac aaaaacaaat gttctgtca cgggggcatg ctcagatgc tgcctccctg caaaacgtgc gagagaatga atcagaattt caacatttgc ttcacacag
gggtcagaga ctgtcagag tgttccctg gctgtcaga atcaaacg gtcgcagaa agaggacgta toggaaactc tgtgcattc atcatctgt
gggggggggt ccagagattg cttgtcggc ctgcatctg gtcaacgtgg acctggatga ctgtttct gagcaataaa tgacttaac caggtatggc
tgccgatgtt tatctccag attggctcga ggacaacct

SEQ ID NO: 822: tetO-containing P40 promoter with a modified TATA element (upper case) and AAV9 P40 intron (lower case):

GAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATGAA
TTCTACGTCAAAAAGGG
TGGAGCCAAGAAAAGACCCGCCCCCAGTGACGCAGGTATATAAGTGAGCCCCCTCTCCCTATCAGTGATAGAGAT
CTCCCTATCAGTGATGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGAGTCAGACGCGGAAGCTT
CGATCAACT ACGCAGACAG gtac caaaacaaat gttctgtca cggggcatg ctcagatgc tcttccctg caaaacgtgc gagagaatga
atcagaattt caacatttgc ttacacacg gggtcagaga ctgtcagag tgttccctg ggtgtcaga atctcaacg gtgtcagaa agaggacgt
tcggaaactc tgtgcattc atcatctgt ggggggggt ccagagattg cttgtctggc ctgcgatcg gtcaacgtgg acctggatga ctgtgttct
gagcaataaa tgacttaac caggtatggc tgcgatgtt tatcttcag attggtctga ggacaacct

SEQ ID NO: 823: Tet-operator-containing AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_2

gaaggtcaccaagcaggaagtc aaagaacttttccgggtgggcaaaggatcacgtgggtgagggtggagcatgaattctacgtcaaaaaggg
tgagccaagaaaagacccgccccagtgacgcaggtatataagtgagcccCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATA
GAGACCaaCgggtgtcgagtgatgtt ggcagccat ogCgtcagagcggaagct tcatcaact tgcagacag GggtGaGTaCc
aaCtgtctc TaCgtgttgc cctgcaAaca Cgcaaga Ttgaatcaga atcaaatCt ctgttcaat caggacCga gactgtti gtgtti ccgtgtcag
aatctcaaccctgtctgc gtcaagg cgtacagaa actgtctac attcatcata lcTgggaaagggtccagac gctgcactg cctgcgatct ggtcaCtgt
gatttggCtaactgcacttttccacCatTT aCCa tttT tcagg

SEQ ID NO: 824: AAV2 CAP under the control of the tetO-containing AAV2 P40 promoter with a modified AAV2 P40 TATA element and AAV9 P40 intron and AAV9 poly A

5'-

gaaggtcaccaagcaggaagtc aaagaacttttccgggtgggcaaaggatcacgtgggtgagggtggagcatgaattctacgtcaaaaaggggtggagccaagaaaa
gacccgccccagtgacgcaggtatataagtgagcccCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGACCaaacgg
gtgcgagtgatgttgcagccatgcagtcagacgggaagcttgcataactacgcagacaggtacaaaacaaatgttctgtcagcggggcatgtctcag
atgtcttccctgcaaaacgtgcgagagaatgaatcagaatttcaacattgtctcacacacgggtcagagactgtctcagagtttccctggcgtgtcagaatctc
aacgggtgtcagaaaggaggtatcggaaactcgtgcgattcatctgtctgggggggtcccgagattgtctgtcggcctgtcgtcaacgtggacc
tggaigtgtgttctgagcaataaatgaacttaaacagggtatggtgcgagtggtatcttcagattgtctcagagacactctctgaaggaaataagacagtggtg
gaagtcacaaactggccaccaccaccacaaagccgcagagcggcataaggacgacagcaggggtctgtgtcttccgtgacaagtaactcggacccitcaacg
gactgcacaaaggagagccgggtcaacgaggaagacgcggccctgcagacacaaagcctacgacggcagctgcagcggagacaacccgtacct
caagtacaaccaacgcgcaggttccaggagcgtttaaagaagatacgtcttgggggcaacctcggacgagcagcttccaggcgaagaaagggttct
tgaacctctgggctgttgggaacctgttaagacggctccgggaaaaaaggagccggtagagcactctctgtggagccagactctctcgggaacccgaaa
ggggggccagcagcctgcaagaaaaagattgaatttggtagactggagacgagactcagtaacctgacccacagcctcggacagccaccagcagccct
ctgtctgtggaaactaatcgtatggtacaggcagtggtgcacaaatggcagacaataacgaggggcgcagcaggtgggttaattctcgggaaattggcattgog
attccacatggatgggcagagatcatcaccaccagcaccgaacctgggcccctgccacctacaacaaccacctctacaaacaaattccagccaatcagga
gcctgaacgacaatcactacttgggtacagcacccttgggggtatttgactcaacagattccactgccactttcaccacgtgactgtgcaaaagactcatcaaaa
caactggggtatccgacccaagagactcaactcaagctttaaacttaacatcaagtcacaaagggtacgcagaaatgacggtacgacagcaggtgcaataaacctacc
agcaoggttcaggtgttactgactcggagtaacagctccctcagctcctcggcgcacaaaggatgcctcccgccgttccagcagagcgttctatggtgccac
agtatggatccctcaccctgaacaacgggagtcagcagtaggacgctctcatttactgcttgagtaacttctctcagatgtgtgtaaccggaacaaacttacct

131

[illegible]

SEQ ID NO: 826: SFA2-EP40CAP2WZ2__SB: AAV2 CAP under the control of the tetO-containing hCMV enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and MODIFIED AAV2 P40 intron_2 and WPRE/modified ICP27 poly A followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

5.

CGCGCCGCT atacagtgaagtgggaagtttacatacacacttaagttggagtcataaaaclogttttcaactaclocacaaaattctgttaacaaacaatagtttg
gcaagtcagttaggacatctactttgtgcgatgacacaagtcattttccaacaaatgittacagacagattatttcaactataatcacgtatcacaaattccagtgggtcaga
agtttacatacactaaGATATCTCAACGGGACTTTCCAAAATcccgccGGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGGCCAACGACCCCGGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT
CCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCACGGTAAATGGCCCGCCTGGCCACTGGGCGTGG
ATAGCGGTTTGACTCACGGGGATTTCAAAATGTCGTAACAACTCCGCCCCATTGACGTCAATGGGCGGTTTGT
TTTGGCACC AAAATGGCGGGGaaggtcaccaagcaggaagtc aaagacttttccgggtgggcaaggatcacgtggttgagggtgagcatgaattct
acgtcaaaaagggtggagccaagaaaagaccccgccccagtgagcgaggtatataagtgagcccCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGACCaaaagggtgcgcgagtcagtt ggcagccat cgaogtcagacgcggaagot tcatcaact acgcagacag
gtacaaaac aaCtgttctc gtacagtgggcCtgaatcig aCgtgttfc cctgcagaca Ctgcagagaga Ttgaatcaga attcaaatCt ctgttcaact
caaggacaga aagactgttt agagtgtt cccgtgtcag aatctcaaccogtttctgc gtcaaaaagg cgtatcagaa actgtgtac attcatata
lcTgggaaagggtgcagac gcttgcactc cctgcgatct ggicaCtgtg gatttggCtg actgcatotttgaacCataa aCga ttlaaa tcaggi atggctgc
cgatgttattctocagatt ggctcgagga cactctctct gaaggaataa gacagtgttg gaagctcaaacctggccac caccacaaa gccgcgag
cggcataagg acgacagcag ggttctgtgttctctgggt acaagtaact cgaaccttc aacggactcg acaaaggaga gccggtcaacgaggcagacg
ccgcgccct ogagcaogac aaagctacg accgcagct cgacagcggagacaaccgt acotcaagta caaccacgcc gaogcggagt
tcaggagcg ccftaagaagatactctt ttgggggcaa cctcggaoga gcagcttcc agcgcaaaaa gagggttctgaacctctgg gctgtgtga
ggaacctgtt aagaoggtc cgggaaaaaa gaggcgggtagagcactct ctgtggagcc agactctcc tggggaaccg gaaagcggg
ccagcagcctgaagaaaaa gattgaatt ttgtcagact ggagacgag actcagtaac tgacccccagctctcggac agccaccagc agccccct
ggctgggaa ctaatacgal ggtacaggcagtgccgcac caatggcaga caataacgag ggcgcgacg gattgggttaa ttctcgggaaattggcatt
ccgattccac atggatggc gacagactca taccaccag cccccgaacctggccctgc ccactagaa caaccaccc tacaacaaa ttccagcc

[illegible]

SEQ ID NO: 841: Plasmid VB220803-2044esq_pSFP5TO1-AAVCDH: pSFP5TO1-AAVCDH is an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP 78/68 ORFs under the control of P5TO1 promoter, 2) Ad5 E2A_IRES_Ad5 E4ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-2 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

5.

ogtacggcgogcgogcgogcgogccaacttctatagaaaagttgttaacacttggattgcgggaaacgagttaaagtcggctcogxtgaattgcgcgtactccgcggga
 gcgcgttctaactcgggtcatatagatttgcgggtgagtcgggaaacgtgtaaactcgggcgcgattgtlaactgcgtattaccaaatatttctatagctctagcgataagc
 tcatatogaaatcgaggctgactcaattgtlaattaatgtggcgctggatagcgggttgactcacggggatttccaaagctccaccccaatgacgtcaatgggagttgtttg
 gcacaaaaaicaacgggactttccaaaaatgctgaacaaactcgcgcacatgacgcaaatggcgggggttcctgtattagaggtcacgtgagttgtttgcacatttgc
 gacaccatgttggtcacgctgggtatataagcaacgggtctccatcacgtgatagagatctccatcacgtgatagagatcggcccgagtgagcagcaggggtctcc
 gcggcgcgcgcgagccaacatgcgcgggttttacgagattgtgattaaggctccacagcgaacttgaagagcatctgcgcggcatttctgacagcttctgaactgggt
 ggccgagaaggaatgggagttgcgcgcagattctgacatgcatctgaatctgattgagcagcgaacccctgacccgtggccgagaagctgcagcgcgacttctgaag
 gaattgcgcgcgtgtgagtgaaggccccggaggccctttcttctgtcaatttgagaaggagagagatctccacatgcacgtgtctcgtggaaacccacgggggtgaa
 atccatgggtttggacgttctcgtagtcagattcgcgaaaaaactgattcagagaatttaccgogggatcgagccgacttctccaaaactgggttcgggtcacaaagacc
 agaaaatggcccgaggccgggaacaaagtgctgagtgatctacalcccaaatcttctcctcccaaaaacccacccctgagctccagtgggcggtggaactaatatg

[illegible]

[illegible]

SEQ ID NO: 842: Plasmid with pSF-ITRP5T01-AAV2H: pSF-ITRP5T01-AAV2H is an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5T01 promoter and CAPs under the native P40 promoter, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

5.

ogtaacggggcgcgccggccgccaaacttctatagaaaagttgttaacacttggatfggggaaaacaggttaagtgggctcgogtgaattggcgctactcgcgggga
 gccgttctaactcgggtcatatagattgcggtggagtgcggggaaaacgtgtaaacctggggcogatfctgaactcgctattaccaaatattgttagatccctagcgataagc
 ttgatacgaattcgaggctgacittaatcaacccattgacgcaalggcgggcggtggatagcggttgaactacgggggattccaagtctccaccocattgacgtcaat
 gggagttgtttggccaccaaaatcaacgggactttccaaaagtctgaacaaactcgccocattgaacgcaaatggcggggggctcgtattagaggctcaactgagtggtt

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

SEQ ID NO: 830: Synthetic hCMV-derived Enhancer Element-3 (contains 2 copies of 21 bp repeat, 3 copies of 18 bp repeats, 3 copies of 19 bp repeats and 2 additional SP1 sites):

TCAACGGGACTTTCCAAAATccggccGGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACC
 CCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGG
 TGGACTATTTACGGTAACTGCCACGGTAAATGGCCCCGCTGGCCACTGGGCGTGGATAGCGGTTTGACTCAC
 GGGGATTTCCAAAATGTCGTAACAACCTCCGCCCATTTGACGTCAATGGGCGGTTTGTGTTTGGCACCAAAATGGC
 GGG

SEQ ID NO: 831: Sleeping Beauty LE ITR Sequence (IR/DR(L) Lmut44) (231 bp including TATA):

5'-

tatacagtgaagtcggaagttacatacaactaagttggagtcattaaaactcgttttcaactactccacaaattctgttaacaaacaatagtttggcaagtcagttag
 gacatctactttgtgcagacacaagtcattttccaacaattgtttacagacagattattcacttataattcactgtatcacattccagtggtcagaagttacatacaact
 aa- 3'

SEQ ID NO: 832: Sleeping Beauty RE ITR Sequence (IR/DR (R) Rmut13, Δ130, 143, 150) (232 bp including TATA):

5'-

ttgagtgatgtaaactctgacccactgggaatgtgatgaagaaataaaagctgaaatgaatcattctctactatttctgatatcttcaactcttaaaataaagtggt
 gatctactgacctaagacagggaaattttactaggattaaatgtcaggaaattgtgaaaaagtgagttaaatgtatttggctaagggtatgtaaactccgactcaac
 tgtata -- 3'

SEQ ID NO: 833: P5TO1 promoter sequence:

Ttaattaatgggctggatagcgggttgactcaggggatttccaaagtcctcccccattgacgcaatgggagttgtttggcaccacaaatcaacgggactttccaaa
 atgtcgttaacaactccgccccattgacgcaaatgggctcctgtatttagaggtaacgtgagtggtttgcgacattttgcgacaccattgttggtcagctgggtatata
 agcaacoggtctccctatcagtgatagagatctccctatcagtgatagagatcggcccgagtgagcagcagggctctccgogggcgc

SEQ ID NO: 834: SV40TO promoter sequence:

Cccgggcaagtttgacaaaaagcaggctctgtggaatgtgtcagttaggggtgtggaagtcocaggctccccagcaggcagaagatgcaaacatgcacat
 ctcaattagtcagcaaccagggtgtggaagtcocaggctccccagcaggcagaagatgcaaacatgcacatctcaattagtcagcaaccatagtcocggccctaa
 ctccggccatcccgccctaaactccggccagttccggccattctccggccatggctgacaggcgccgggtatataagcagagctctccctatcagtgatagagatc
 ccctatcagtgatagagatcagctcgggtgtgtgagggccggagggccgctc

SEQ ID NO: 835: HSV-2 ICP4 promoter sequence followed by VA RNAs and ICP27 poly A

5'-

gggcccgggggggccaacgggagcggggggcgggcactcattaccaogaaacccggaagggcaggggagcggagcccgcccgacgagggtctcattagc
 atcggggggggaagcgggaagcggcccgccggggcgctaatgagatgcggcgggggcggagcggcgggcgccgaccaacggggccggccacgggaog
 cggagcggggggcggggggggggcggcgaataatgcgggtccacctggggggcgggaaccccgggcagcggggggcggcgggcgctcgcctcctccg
 cgtctcctctcttcccccggccggcgcccgaggactatagacagggcgaacgggggatcgggcactctcgggtggtcgtggataaattcgcaagggtat
 catggcggaacgacgggggttagcggccgtatccggcggcggcgtgacatgcgggttaccggccgggtgtgaacccagggtgtgcgaogtcagacaacggg
 ggagtgctccttctggctcctccaggcggggggcgtcgtcgtcgtagctttttggccactggccggcgccagcgtgaacgggttaggctggaagcgaagcgaatgaag

tgctctgctccctgtagccggaggggtattttccaaagggtagctcgogggacccccgggttagagctcggacccggccggactcggcgaaacgggggttgctcccc
gtcatgcaagaccccgcttgcaaatctccggaaacagggacgagcccttttcaaatatttattgcaactccctgttttagglacaataaaaaacaaaacatttcaa
acaaatcgccctcgtgttgctctcttgcctcatggccggcggaacgogtaccggt -- 3'

SEQ ID NO: 836: TK Promoter – hygro_resistant gene_synthetic poly A:

5-

caaaaccccgcccagcgtctgtcatgtggcgaattogaaacacgcagatgcagtcggggcgggcgcgaggtccacttcgtatataagggtgacggtgtggcctc
gaacacccgagcgaacctgcagcgaacctgttaacagcgccacctgaaaaagccctgaactcacccgcgacgtctgtcagagaagttctgatcgaaaagttcgac
agcgtctccgaccgtatgcagctctcggaggggcgaagaatctcgtgttcagcttcgatgtaggaggggcgtgggataatgtccgcggglaaalagctgcgcgcgafggt
tctacaaagalcgttatgttatcggcactttgcatcggccgcgcgtcccgatccgggaagtgttgacattgggggaattcacgcgagagccctgacctatgtcatctccgc
gtgcacaggggtgtcacgttgcaagacctgcctgaaacccgaactgcgcgcgtgtctgcagccgggtcgcggaggccatggatgcgatcgtcgcgcgacttagcca
gacgagcgggttcggcccatcggacccgcaaggaaatcgggtcaatatactacatggcgtgattcatatgcgcgattgtcgtatccccatgtgtatcactggcaaaactgtg
atggacgacacccgtcagtcgcgtccgcgcgaggtctcgtatgagctgaigtcttggggcgcgaggacgtcccccgaagtcgggcacctcgtgcacgcggatttcggctc
caacaatgttcgtacgggacaatggccgcataaacagcgggtcattgactggagcgcgaggtgatgttgggggaltcccaatacagaggtgcgcaacatctctcttggaggc
cgtggttggctgttatggagcagcagacgcgcacttcgtagcgcgaggcatccggagcttgcaggatcgcgcggctccggcggtatatagtctccgcatttggtctgacc
aacctatcagagcttgggtgacggcaatttgcgtatgcagcttgggcgcgaggttcgatgcgcgcaatcgtccgataccggagccgggacgtcggcggtacacaa
atcgcgcgcgagaagcgcgcgcgcgtcgtggaacgatggtctgtgtagaagtaactcgcgcgatalgtggaaaccgcgcgcgcagcactcgtccggatcgggagatggggg
aggctaactgaacacaataaaaagatcttattttcattagatctgtgttgggtttttgtgtg -- 3'

SEQ ID NO: 837: Zeocin Resistant Gene under the control of a modified HSV-1 ICP0 promoter with modified ICP0 poly A

5-

[illegible]

SEQ ID NO: 838: WPRE REGULATORY ELEMENTS PLUS BGH POLY A:

Aatcaacctctggattacaaaattgtgaaagattgaactgtattcttaactaigtgtctcttttaagctatgtggalaagctgtttaatgoccttgtatcatgtctattgtctcc
gtatggctttcatcttctctctgtataaatcccggttgctgtctttatgaggagtggtggcccggtgtcaggcaacgtggcggtgggtgtgcaactgtgtttgtgtaagcaaaccc
ccactggttggggcaatgccaccacactgtcagctcccttcggggaccttgccttccccctccctattgccacgggggaactcatgcgcgcctgccttgcgcgcgtctgtgga
caggggctcgggtgttgggcaactgacaattccgtggtgtgtcgggggaagctgaagctccttccatggctgtctgcctgtgtgtgccacctggattctgcgcggggacgtcct
ctgtcactgccttgcggccctcaatccagggaccttccctcccgggcctgtcgcgggtctcggggcctcttcggcgctctgccttgcgcctcagacgagtcggatctc
ccittgggcgcgcctccccgcctggaattcgagctcgtgaagatcagcctogaCtGtgcctctagttgccagccatctgtgttggccctcccccgctgccttcttgaccct
ggaagggtgcactccactgtccttctctaataaaaaatgaggaaattgcacatgcattgtctgagtagggtcattctattctgggggggtgggggtggggcaggacagcaag
ggggaggattgggaagacaatagcaggcatgc

SEQ ID NO: 839: AAV2 REPs coding sequence plus its native poly A tail:

gogcagccaccatgcgggggttttaacgagattgtgattaaaggtocccagcgacotfgacgagcatctgcgggcattctgacacgtttgtgaactgggtggccgaga
aggaaatgggagttgcggccagattctgacatggatctgaatctgattgagcagggcaocctgacccgtggccgagaagctgcagcgogacittctgacggaaatggcg
ccgtgtgaglaaggcccccggaggccctttcttctgtcaatttgagaagggagagagctactccacatgcacgtgtcctgtggaaaccacccggggigaaatccatggt
tttgggaagtttctgagtcagattcgggaaaaactgattcagagaatttaccgcggtatcgagcgacitttgccaaaactgtgtcgggtcacaaaggaccagaaatgg
cgccggaggcggggaacaagggtgggtgatgagtgtacatccccaattactgtctcccaaaaaccagccctgagctccagtgggcgctggactaataatggaacagtat
ttaagcgccgtgttgaaatccacggagcgtaaacgggttggtggcgagcatctgacgcacgtgtgcgacagcaggagcagaacaaaggagaatcagaalcccaat
ctgcatgcggccgggtgacagatcaaaaacttcagccagggtacatggagctgtgtgggtgtgtgcacaaaggggattaccctggagaagcagtggtgacccaggag
gaccaggccctacatcatctccttcaatgcggccctccaaactcgggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgccc
ccgactaccctgtggggccagcagcccggtggaggacatttccagcaatcggatttataaaattttggaactaaacgggtacgacccccaatatgcggcttccgttctctg
ggatggggccacgaaaaagttcggcaagagggaacaccatctggtgttttggccctgcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgc
ccttctacgggtgtcgttaaactggaccaatgagaacttcccttcaacagactgtgtcgacaagatggtgatctgtgtgggaggagggaagatgacogccaagggtcgtg
gagtcggccaaagccattctcggagggaagcagggtgcgcgtggaccagaatatgcaagtctcggccagatagaccgactcccggtgatctgcacotccaacac
caacatgtgcgcggtgattgaocgggaactcaacgaccttcgaacaccagcagccgttcgaagaccggatgttcaaaattgaaactacccgcggctcgtgacatgactt
tggaagggtaccacagcagggaagtcacaaagacttttccggtgggcaaaaggatcacgtggttgagggtggagcatgaattctacgtcaaaaagggtggagccaagaa
aagaccgcgccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcgggaagcttgcataactacgca
gacagggtaccacaaacaaatgttctcgtcacgtgggcgatgaatctgaigtgtttccctgcagacaatgcgagagaatgaatcagaatcacaatctcgttccactcag
gacagaaagactgttagagtgcttcccggtgcagaatctcaaccogttctgtgtgcacaaaaggcggtatcagaaactgtgtacattcactatcatatcgtggaaagggtg
ccagacogctfgcacigccctgcgatctggtcaatgtggatttggatgactgcacttctgaacaataaaagtattaaatcagggtatggctgcogattgttalcaccagattggc
tcgaggacactctctgaagggaataagataatctgaattgtgtgtaatcaataaacogtttaaltcgttccagttgaacttggctctcgttattcttctatctagtttccat
ggctacgta

SEQ ID NO: 840: Ad5 E2A coding sequence followed by IRES and Ad5 ORF 6 coding sequence and SV40 poly A

EXAMPLES

Hereinafter, the present disclosure will be described in further detail with reference to examples. These examples are illustrative purposes only and are not to be construed to limit the scope of the present invention. In addition, various modifications and variations can be made without departing from the technical scope of the present invention.

Example 1 – Design of Chimeric Mobile Element Enzymes with Transcription Activator-Like Effector (TALE) DNA Binding Domains (DBDs) or dCas9/gRNA that Target Human Genomic Safe Harbor Sites (GSHS)

FIG. 2E illustrates a non-limiting example of a system comprising a nucleic acid (e.g., helper RNA) encoding an enzyme capable of performing targeted genomic integration and a nucleic acid encoding a mobile element enzyme (donor DNA). The helper RNA is translated into a bioengineered enzyme (e.g., integrase, recombinase, or mobile element enzyme) that recognizes specific ends and seamlessly inserts the donor DNA into the human genome in a site-specific manner without a footprint.

In this example, enzymes, e.g., chimeric mobile element enzymes, are designed using human GSHS TALE or Cas9/gRNA DBD. **FIGs. 2A-D** depict representations of chimeric mobile element enzyme designed using human GSHS TALE or Cas9/gRNA DBD. **FIG. 2A.** TALEs includes nuclear localization signals (NLS) and an activation domain (AD) to function as transcriptional activators. A central tandem repeat domain confers specific DNA-binding and host specificity. Translocation signal (TD) and four cryptic repeats required for initiation of DNA binding and for the recognition of 5' -T⁰ are located at the N-terminus (checkered rectangles). Each 34 amino acid (aa) long repeat in the CRD binds to one nucleotide with specificity determined mainly by aa at position 13. One sample repeat is shown below the protein scheme. Numbers 12/13 refer to aa positions within the repeat. See Jankele *et al.*, *Brief Funct Genomics* 2014;13:409-19. **FIG. 2B.** Repeat types are shown that have specificity for one or several nucleotides. Only bases of the DNA leading strand are shown. **FIG. 2C.** A chimeric mobile element enzyme having a TALE DNA-binding protein fused thereto by a linker that is greater than 23 amino acids in length (top). See Hew *et al.*, *Synth Biol (Oxf)* 2019;4:ysz018. **FIG. 2D.** Binding of the TALE to GSHS physically sequesters the mobile element enzyme to the same location and promotes transposition to the nearby TTAA sequences. All RVD are preceded by a thymine (T) to bind to the NTR shown in **FIG. 2A**. All of these GSHS regions are in open chromatin and are susceptible to mobile element enzyme activity).

FIG. 2C also illustrates (bottom) a chimeric mobile element enzyme construct comprising dCas9 linked to one or more guide RNAs. An engineered chimeric mobile element enzyme may include: a guide RNA (gRNA) and an inactivated Cas protein. The gRNA is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Thus, the genomic target of the Cas protein is based upon the sequence present in the gRNA. **FIG. 3B** shows gRNA sequences that physically sequester the mobile element enzyme to GSHS and promotes transposition to the nearby TTAA sequences.

FIGs. 3A-3E depict examples of construct templates (**FIGs. 3A-3D**) and a system (**FIG. 3E**) comprising a donor DNA and helper RNA. **FIG. 3A** depicts a plasmid construct template that transcribes mobile element enzyme RNA that is later processed with a 5'- m7G cap. Other mobile element enzymes can be substituted. **FIG. 3B** depicts a donor DNA construct template with the transgene VLDLR. Other tissue specific promoters and transgenes can be substituted. **FIG. 3C** depicts a chimeric mobile element enzyme construct template with a TALE DNA binder. Other TALEs and mobile element enzymes can be substituted. **FIG. 3D** depicts a chimeric mobile element enzyme construct template with a dCas9/gRNA DNA binder. Other dCas9/gRNA and mobile element enzymes can be substituted.

FIG. 3E depicts a non-limiting example of the system of **FIG. 1A**. In **FIG. 3E**, the system comprises a nucleic acid (donor DNA), and a nucleic acid (helper RNA) encoding a mobile element enzyme (hyperactive ENGINEERED MOBILE ELEMENT ENZYME). As shown in **FIG. 3E**, panel (**A**), the donor DNA can be any gene of interest (GOI) including those replace, inactive, or provide suicide or helper functions. The GOI can be driven by a predetermined promoter and flanked by insulators to prevent gene silencing. The ITRs are specific for the mammal-derived mobile element enzyme (ENGINEERED MOBILE ELEMENT ENZYME). **FIG. 3E**, panel (**B**) depicts the helper RNA that is 5'-m7G capped (cap1) with flanking globin 5'- and 3'-UTRs, a 34 polyalanine tail region, and pseudouridine modification. The mobile element enzyme (e.g., a hyperactive mobile element enzyme in accordance with the present disclosure) is bioengineered to insert the donor DNA in a site- or locus specific manner.

Example 2 – Creation of an AAV Packaging and Producer Cell Line

A goal of this study was to assess the ability to produce high-titer recombinant AAV particles in a two-plasmid transfection protocol. **FIG. 1A** is a non-limiting representation of an AAV production strategy in accordance with embodiments of the present disclosure, using a donor vector to create a HEK293 producer cell line or another cell line including the E1a gene (e.g., engineered CHO-K1 or Vero cell line with E1a) that forms a replication deficient (*rep-*) AAV particle containing a transgene of interest. The single or dual donor comprising a transgene of interest is incorporated (e.g., by transfection such as electroporation) into a HEK293 cell line or another cell line including the E1a gene. The HEK293 cells are expanded to create a producer cell line and culturing the producer cells creates replication deficient (*rep-*) AAV particles with the transgene of interest. **FIG. 1B** is a representation of an example of an inducible Rep/Cap and helper AAV donor plasmid construct (Kana^r) used with a helper RNA or DNA to create an AAV producer cell line, in accordance with embodiments of the present disclosure. The donor DNA plasmid of **FIG. 1B** includes helper E2A, E4 and VA genes flanked by insulators and mobile element enzyme recognition ends, to create an AAV producer cell line. **FIG. 1C** is a representation of an example of a nucleic acid (plasmid) encoding a transgene (gene of interest (GOI)) included between AAV inverted terminal repeats (ITRs), in accordance with embodiments of the present disclosure. The plasmids shown in **FIGs. 1B** and **1C** can be combined or can be used separately for transfection into an E1+ cell line to produce AAV particle comprising the GOI.

FIG. 4 depicts a non-limiting representation of a conventional rAAV production system. Production in adenovirus complementation systems is usually performed as plasmid transfection processes, where AAV Rep/Cap genes, the ITR-flanked gene of interest (GOI), as well as Ad-helper genes are provided as three separate plasmids, respectively, to a E1a/E1b containing HEK293 cell line or other engineered cell line that contains E1a (e.g., engineered CHO-K1 or Vero cells).

The present strategy is to include the AAV *rep* and *cap* genes, the AAV vector DNA sequences, and the essential Ad helper genes in a single donor plasmid flanked end sequences recognized by the mobile element enzyme in accordance with embodiments of the present disclosure. The AAV *rep* and *cap* genes are included under the control of t-REx, leading to *rep* and *cap* gene amplification, the AAV vector DNA sequences, and the essential Ad helper genes in a single donor plasmid flanked by the end sequences recognized by the mobile element enzyme in accordance with embodiments of the present disclosure (**FIG. 1B**). The ability of the integrated AAV helper plasmid to direct the rescue, replication, and packaging of an AAV ITR-flanked transgene (e.g., the transgene of **FIG. 1C**) will be assessed.

The integrated AAV helper plasmid can direct the rescue, replication, and packaging of an AAV ITR-flanked transgene. This mobile element enzyme-mediated recombination system is expected to generate an integrated AAV helper plasmid that can facilitate the production of high-titer recombinant AAV particles in a simple two-plasmid transfection protocol. In addition, the inducible nature of the *rep* and *cap* genes is employed by the use of reagents (e.g., rAAV-5 plasmid reagents) useful for the construction of rAAV-5 vectors bearing other reporter and/or therapeutic transgenes.

Example 3 – Identification of Excision Positive and Integration Negative Mutants

TABLES 14A-14B depict hyperactive MLT mutants (**TABLE 14A**) and Integration-deficient (Int-) mutants (**TABLE 14B**) from MLT mobile element enzyme DNA.

TABLE 14A. Hyperactive MLT mutants from mobile element enzyme protein of SEQ ID NO: 1 and mobile element enzyme DNA of SEQ ID NO: 2

Nucleotide Change	Amino Acid Change
C41T	A14V
A1424G	D475G
C1472A	P491Q
G1681A	A561T
A898G	T300A
A880G	T294A
G1558A	A520T
G715A	G239R
T13C	S5P
C23T	S8F
G161A	S54N

A26G	D9G
A1033G	I345V
A1441G	M481V
A32G	E11G
A389C	K130T
A32G	E11G
G1280A	R427H
A106G	S36G
A29G	D10G
N125K	T375G
S8P	T22C
C13R	T37C

TABLE 14B. Integration deficient MLT mutants from mobile element enzyme protein of SEQ ID NO: 1 and mobile element enzyme DNA of SEQ ID NO: 2

Nucleotide Change	Amino Acid Change
T841G, A842C	Y281A
T844G, G845C	C282A
G848C	G283A
A851C	E284A
G854C	G285A
A928G, A929C	N310A
G989C	G330A
A991G	T331A
A994G, T995C	I332A
C997G, G998C	R333A
A1000G, A1001C	K334A
A1003G, A1004C	N335A
C1006G, G1007C	R336A
G1010C	G337A
A1012G, T1013C	I338A
C1015G	P339A
A1247C	D416A

Example 4: Schematic Illustration for Constructs Used for Establishment of AAV producer stable cell lines, AAV2-H (Plasmid shown in FIG. 5), AAV2-HB (Plasmid shown in FIG. 6), and AAV-CDH (Plasmid shown in FIG. 7) in HEK-293R-22 tetR-expressing HEK-293 cells.

Constructs pSF-ITRP5TO1-AAV2H, pSFP5TO1-AAV2HB, or pSFP5TO1-AAVCDH were used to co-transfect HEK-293R-22 cells with pCMV-Engineered mobile element enzymeB VB200927-4525erk, and pSF-mCMVR to generate AAV stable producer/packaging cell pools, AAV2-H (Plasmid shown in FIG. 5), AAV2-HB (Plasmid shown in FIG. 6), and AAV-CDH (Plasmid shown in FIG. 7), for subsequent single cell cloning, respectively.

a. Plasmid with pSF-ITRP5TO1-AAV2H (**FIG. 5**): pSF-ITRP5TO1-AAV2H is an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter with AAV2_REP52/40 ORFs and CAPs (VP1/2/3) under their native P19 or P40 promoter, respectively, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter.

b. Plasmid with pSF-ITRP5TO1-AAV2HB (**FIG. 6**): pSFP5TO1-AAV2HB is an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP78/68 ORFs under the control of P5TO1 promoter and AAV2_REP52/40 ORFs under the native P19 promoter, 2) AAV2 CAPs under the tetO-containing P40 promoter with AAV9 P40 intron, 3) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 4) VA RNAs under the control of HSV-1 ICP4 promoter, and 5) hygro-B resistant gene under the control of TK promoter.

c. Plasmid pSFP5TO1-AAVCDH (**FIG. 7**): pSFP5TO1-AAVCDH is an engineered mobile element enzyme ITR containing plasmid that encodes 1) AAV2_REP 78/68 ORFs under the control of P5TO1 promoter and AAV2_REP52/40 ORFs under the native P19 promoter, 2) Ad5 E2A_IRES_Ad5 E4 ORF6 under the control of the tetO-containing SV40 promoter, 3) VA RNAs under the control of HSV-1 ICP4 promoter, and 4) hygro-B resistant gene under the control of TK promoter. Note that the nucleotide sequence between the end of native AAV Rep ORFs and native poly A was deleted in pSFP5TO1-AAVCDH to create an AAV Cap VP1/2/3 deletion phenotype.

d. Plasmid pSFA2-EP40CAP2WZ1_SB (**FIG. 8**): This construct encodes AAV2 CAP under the control of the tetO-containing hCMV-derived enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_1 (a at position 276 substituted for C) and WPRE/modified ICP27 poly A, followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with ITRs for Sleeping Beauty.

e. Plasmid pSFA2-EP40CAP2WZ2_SB (**FIG. 9**): This construct encodes AAV2 CAP under the control of the tetO-containing hCMV enhancer element 3/AAV2 P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron_2 and WPRE/modified ICP27 poly A, followed by ICP0 promoter-Zeocin resistant gene-ICP0 poly A cassette flanked with SB ITRs.

Establishment of AAV Producer/packaging cell lines

a. Establishment of AAV2 full producer stable cell lines, AAVH cells, in HEK-293R-22 tetR-expressing HEK-293 cells: HEK-293R-22 cells were seeded at about 5×10^6 cells per 100 mm dish with normal HEK-293 cells growth

medium. Cells were co-transfected with pSF-ITRP5TO1-AAV2H, CMV-Engineered mobile element enzymeB VB200927-4525erk, and pSF-mCMVR at a DNA:Lipofectamine 2000 ratio of 1:2. Transfection medium was removed at 3 and a half hour post transfection followed by addition of normal 293 cells growth medium. Transfected cells were passed on day 3 post transfection and grew in 293 cells growth medium in the presence of hygromycin-B at a concentration of 100 µg/ml. Hygromycin-B resistant single stable AAV-H colonies were then selected and analyzed by 1) ddPCR analyses with REP, E2A, ORF6, VA RNA and hygromycin-B resistant gene-specific primers, 2) western blot analysis with anti-AAV-2 REPs and CAPs specific antibodies, and 3) their ability to support AAV genome amplification following transduction with AAV2-EGFP viral particles.

b. Establishment of second generation AAV2 full producer stable cell lines, AAVHB cells, in HEK-293R-22 tetR-expressing HEK-293 cells: HEK-293R-22 cells were seeded at about 5×10^6 cells per 100 mm dish with normal HEK-293 cells growth medium. Cells were co-transfected with VB220909-1308jwtpSF-ITRP5TO1-AAV2HB, CMV-Engineered mobile element enzymeB VB200927-4525erk, and pSF-mCMVR at a DNA:Lipofectamine 2000 ratio of 1:2. Transfection medium was removed at 3 and a half hour post transfection followed by addition of normal 293 cells growth medium. Transfected cells were passed on day 2 post transfection and grew in 293 cells growth medium in the presence of hygromycin-B at a concentration of 100 µg/ml. Hygromycin-B resistant single stable AAV-HB colonies were then selected and analyzed by 1) ddPCR analyses with REP, E2A, ORF6, VA RNA and hygromycin-B resistant gene-specific primers, 2) western blot analysis with anti-AAV-2 REPs and CAPs specific antibodies, and 3) their ability to support AAV genome amplification following transduction with AAV2-EGFP viral particles.

c. Establishment of AAV serotype adaptable (Cap minus or Cap-) stable cell lines, AAV-CDH (Plasmid shown in FIG. 7) cells, in HEK-293R-22 tetR-expressing HEK-293 cells: HEK-293R-22 cells were seeded at about 5×10^6 cells per 100 mm dish with normal HEK-293 cells growth medium. Cells were co-transfected with pSF-ITRP5TO1-AAVCDH, CMV-Engineered mobile element enzymeB VB200927-4525erk, and pSF-mCMVR at a DNA:Lipofectamine 2000 ratio of 1:2. Transfection medium was removed at 3 and a half hour post transfection followed by addition of normal 293 cells growth medium. Transfected cells were passed on day 3 post transfection and grew in 293 cells growth medium in the presence of hygromycin-B at a concentration of 100 µg/ml or 50 µg/ml. Hygromycin-B resistant single stable AAV-CDH (Plasmid shown in FIG. 7) colonies were then selected and analyzed by 1) ddPCR analyses with REP, E2A, ORF6, VA RNA and hygromycin-B resistant gene-specific primers, 2) western blot analysis with anti-AAV-2 REPs specific antibody, and 3) their ability to support AAV genome amplification following transduction with AAV2-EGFP viral particles.

d. Establishment of AAV2 REPs and CAPs expressing stable cell lines with sleeping beauty transposon-mediated integration of AAV2 CAPs-expressing cassettes into AAV2 capsid-minus stable cell lines: AAV2-capsid minus stable lines Clone CDH-48, CDH-24 and CDH-129 cells HEK-293 were seeded at about 4 to 5.5×10^6 cells per 100 mm dish with normal HEK-293 cells growth medium. Cells were co-transfected with pSFA2-EP40CAP2WZ1-SB,

pSFA2-EP40CAP2WZ2-SB, pSF-mCMVR, and sleeping beauty transposase SB100X Helper mRNA (VB210111-1032qxd) at a DNA:Lipofectamine 2000 ratio of 1:2. Transfection medium was removed at 3 and a half hour post transfection followed by addition of normal 293 cells growth medium. Transfected cells were passed on day 2 post transfection with normal HEK-293 cells growth medium. Cells were then grown in HEK-293 cells growth medium in the presence of 100 ug/ml of zeocin. Zeocin resistant single stable colonies were then selected and analyzed by 1) ddPCR analyses with REP, Cap, E2A, ORF6, VA RNA and zeocin resistant gene-specific primers, 2) western blot analysis with anti-AAV-2 REPs and CAPs specific antibodies, and 3) their ability to support AAV genome amplification following transduction with AAV2-EGFP viral particles.

ddPCR analyses for an engineered mobile element enzyme mediated Transposon Integration, AAV Genome Amplification/Rep Function, Cap Copy Insertion via Sleeping Beauty, and physical titer of rAAV particles.

AAV genome amplification is supported in either AAV2 producer cell line clone H-17 (FIG. 10A – 10B, Cap+) or AAV serotype adaptable packaging cell line clone CDH-13 (FIG. 11A – 11B, Cap-). AAV genome amplification was used as a reference for expression of functional AAV Rep proteins in the AAV-H Clone H-17 or AAV-CDH (Plasmid shown in FIG. 7) clone CDH-13 cells. Commercially available rAAV2-EGFP particles from SignaGen at 75 VGC/cell was used to transduce 3×10^6 cells/dish of Clone H-17, CDH-13, or HEK-293 (AAV Rep- control) cells. At 4 hours post-transduction, cells were switched to growth media with or without Doxycycline in the indicated concentration. GFP images (FIG. 10A – FIG. 11B) were taken ~120 hours post-transduction using Zeiss AxioVert A1 microscope and AxioCam 705 mono camera, followed by cell harvesting, genomic DNA extraction, and ddPCR analyses for quantification of the rAAV2-EGFP genome copy number per nanogram (ng) of total genomic DNA. The results clearly indicated that the rAAV-EGFP replicated in doxycycline dose responsive manner (ddPCR results at right), suggesting the inducible expression of functional AAV Rep proteins in the Clone H-17 and CDH-13 cells. Negative/low intensity EGFP image from transduced HEK-293 Rep negative cells also support this conclusion.

Expression of functional AAV2 Rep proteins in AAV2-HB (Plasmid shown in FIG. 6) clones as determined by western blot and ddPCR. FIG. 12A shows the results of an western blot analysis illustrating cells from selected AAV2-HB (Plasmid shown in FIG. 6) cell clones HB-36, HB-42, HB-118, and HB-120 were treated with 10ng/ml doxycycline for 72 hours followed by extraction with RIPA buffer in the presence of proteinase inhibitor. The resulting cell lysate supernatants were subjected to electrophoresis and western blot analyses. Lysate supernatants from HEK-293 cells with or without transfection of pSFP5TO1-AAV2HB plasmid were analyzed as positive and negative controls. Cyclophilin B western blot was performed as a loading control.

FIG. 12B shows the results of a ddPCR analysis for AAV Genome Amplification / Rep function Assay. The AAV-HB stable clone #36, 42, 118 and 120 cells were transduced with the rAAV2-EGFP particles from SignaGen at low MOI, and genomic DNA extracted/ddPCR performed for quantification of the rAAV2-GFP genome copy number per

nanogram (ng) of total genomic DNA analyzed. Western blot results clearly demonstrated that the rAAV Rep proteins are expressed and ddPCR results indicate the expressed AAV2 Rep proteins are functionally active in supporting AAV genome amplification, both in doxycycline responsive manner.

Expression of functional AAV2 Rep proteins in AAV-CDH (Plasmid shown in FIG. 7) clones as determined by western blot (FIG. 13A) and ddPCR (FIG. 13B).

Cell extraction with RIPA buffer containing proteinase inhibitors, western blot, and ddPCR analyses of selected AAV-CDH (Plasmid shown in FIG. 7) Clones of CDH-6, CDH-24, CDH-48, CDH-129, and CDH-175 were carried out as described in legend above. Like the AAVHB clones, western blot results clearly demonstrated that the rAAV Rep proteins are expressed while ddPCR results indicate the expressed AAV2 Rep proteins are functionally active in supporting AAV genome amplification, both in response to doxycycline. Note that CDH-48 and CDH-129 represent the top clones selected based on their Rep expression level, genome copy number/ng gDNA and the fold of induction.

Establishment of AAV2 full producer stable cell pools, CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB with sleeping beauty transposase-mediated integration of AAV2 CAPs-expressing cassettes into AAV-CDH (Plasmid shown in FIG. 7) serotype adaptable (Cap minus or Cap-) stable cell lines CDH-24, CDH48, and CDH-129.

Procedures for Establishment of AAV2 full producer stable cell pools, CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB, with sleeping beauty (SB) transposase-mediated integration of AAV2 CAPs-expressing cassettes into AAV-CDH clone 24, 48, and 129 have been described above. **FIG. 14A** shows ddPCR analyses of SB-mediated AAV2 Cap transposon integration in each Zeocin-selected stable pools. Average Cap copy number integration in each pool is summarized at the bottom of the panel. **FIG. 14B** shows western blot analysis of SB-mediated AAV2 Cap transposon integration in each Zeocin-selected stable pools. Cells from the CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB pools were treated with 20ng/ml doxycycline for 72 hours followed by extraction with RIPA buffer in the presence of proteinase inhibitor. Western blot analysis was performed similarly as describe above. The results show that stable integration of 25-40 copies of AAV-cap transposon enables detection of Cap expression in response to doxycycline. The lysate supernatants from HEK-293 cells with or without transient transfection of pSFA2-EP40CAP2WZ2-SB plasmid were analyzed as positive and negative controls. Cyclophilin B western blot was performed as a loading control.

Expression of functional AAV2 Rep proteins in CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB pools as determined by western blot and ddPCR.

Cell extraction with RIPA buffer containing proteinase inhibitors, western blot, and ddPCR analyses of the CDH-24/Cap-SB, CDH-48/Cap-SB, and CDH-129/Cap-SB cell pools were carried out as described previously. The western blot

results in FIG. 15A clearly demonstrated that the rAAV Rep proteins are expressed while ddPCR results in FIG. 15B indicate the expressed AAV2 Rep proteins are functionally active in supporting AAV genome amplification, both in response to doxycycline.

Expression of AAV2 Cap and Rep in CDH-48/Cap-SB stable pool increased with doxycycline dosages and over time.

Cells from CDH-48/Cap-SB stable pools were incubated either in the presence of 0, 5, 10, 25, 50, or 100ng/ml doxycycline and harvest at 72 hours post induction; or incubated with 50ng/ml doxycycline and harvested at 24, 48, 72, or 96 hours post induction (FIG. 16). Cell extraction with RIPA buffer containing proteinase inhibitors, western blot, and ddPCR analyses of the CDH-48/Cap-SB (AAV2-capsid minus stable lines Clone CDH-48) samples were carried out as described previously. The results show that expression of AAV2 Cap and Rep proteins increased over time in the presence of 50ng/ml doxycycline. The Cap and Rep expression also increased in response to increasing amounts of doxycycline.

Production of competent rAAV2 particles from stably engineered CDH-48/Cap-SB cell pools in response to doxycycline.

The cells from the CDH-48/Cap-SB stable pools were transduced with the rAAV2-EGFP particles from SignaGen at 500 VGC/cell in the presence or absence of 50ng/ml. 4 days post-transduction, culture media were removed, and cells harvested and washed twice with fresh growth media to remove the residue doxycycline. Upon resuspending in 7ml of fresh growth media without doxycycline, cells were subjected to 3 cycles of flash-freeze and thaw. rAAV-EGFP containing supernatants were collected following centrifugation at 2500rpm for 10 minutes AAV physical titer in each sample was measured by ddPCR as described above and data presented in FIG. 17A. To test the presence of the competent rAAV2-EGFP particles in dox- and dox+ samples, the unconcentrated supernatants were used for transduction of CDH-48 Cap- cells at 1.5ml/well of the 6-well tissue culture plates in duplicates. The transduction media were replaced next day with fresh growth media and cells were harvested after additional 2 days incubation followed by genomic DNA extraction. ddPCR analyses of the genomic DNA samples were performed to evaluate copies of the rAAV-GFP copies introduced into CHD-48 Cap- cells through transduction and results presented in the FIG. 17B. The ddPCR results from both panels indicated that rAAV-EGFP particles were produced in response to doxycycline induction and the produced rAAV particles were competent for transduction.

Characterization of Zeocin-resistant single clones selected from the CDH-48/Cap-SB stable pool.

Zeocin resistant single stable colonies were selected from the CDH-48/Cap-SB stable pool and analyzed by 1) ddPCR analysis for SB-mediated AAV2-Cap cassette integration (FIG. 18A); 2) western blot analysis with antibodies specific for AAV2 CAP (FIG. 18B), and AAV2 Rep (FIG. 19A); and 3) ddPCR assays to evaluate their ability to support AAV

genome amplification following transduction with low VGC/cell of AAV2-EGFP viral particles (**FIG. 19B**). Specifically, data in **FIG. 18A** indicated that SB-mediated AAV2-cap cassette integration ranged from 20-90 copies per ng of total genomic DNA. AAV2 Cap and Rep expression profile (both in expression level and ratio among the different Cap or Rep proteins) varies greatly among different clones as expected. Finally, all selected clones are capable to support AAV genome amplification to different degree and fold of induction varies as well.

Materials and Methods

a. Transposon copy number ddPCR: The Transposon Copy Number ddPCR Assay quantifies the copy numbers of five elements in the Engineered mobile element enzyme transposon per genome of HEK293R-22 based cells. These elements are AAV2 Rep, AAV2 Cap, Ad5 E2A, Ad5 E4 Orf6, and VA RNA. Genomic DNA was extracted from samples with the Zymo Research Quick-DNA MiniPrep Kit (D3025) according to the manufacturer's protocol. Sample DNA concentrations were quantified by the Thermo Qubit 1X dsDNA Broad Range Assay Kit (Q33266) according to the manufacturer's protocols and normalized to the same concentration in TE. ddPCR reactions were assembled with Manufacturer's Supermix for Probes (no dUTP), a FAM-labeled hydrolysis probe-based ddPCR assays for one of the transposon elements, a HEX-labeled hydrolysis probe-based ddPCR assay for RPP30, HindIII restriction enzyme, and water according to Manufacturer's ddPCR protocol. Sample DNA was added to reaction wells at inputs of 4 ng and 20 ng per well to ensure that at least one reaction per sample was within the range of the assay. Droplets were generated by the automated QX200 AutoDG system and thermal cycled according to Manufacturer's recommended parameters, with an annealing T_m of 60°C. Thermal cycled droplets were analyzed by the CNV algorithm on the QX200 Droplet Reader, with the reference RPP30 assay set to 3 copies per genome (RPP30 is located on Chr10, which is triploid in the HEK293 cell line). Sample data was inspected for sufficient droplets (>10,000) and properly identified positive and negative droplets prior to analysis. The QX200 droplet reader software automatically calculates the copies per μ L of reaction volume for the target (transposon element) and reference (RPP30) and reports the target copies per genome using the reference RPP30 = 3 copies per genome. Calculated copy numbers for all transposon elements were used to determine the integrity of the inserted transposon, as well as the copy number. Intact transposons had an integer copy number (e.g. 1, 2, etc.) across all elements.

b. AAV Genome Amplification / Rep function Assay: The Genomic Amplification Assay quantifies the copy number of AAV-GFP genomes per nanogram (ng) of genomic DNA analyzed. Genomic DNA was extracted from cells with the Zymo Research Quick-DNA MiniPrep Kit (D3025) according to the manufacturer's protocol. Sample DNA concentrations were quantified by the Thermo Qubit 1X dsDNA Broad Range Assay Kit (Q33266) according to the manufacturer's protocols and normalized to the same concentration in TE. ddPCR reactions were assembled with Manufacturer's Supermix for Probes (no dUTP), a FAM-labeled hydrolysis probe-based ddPCR assay for EGFP, HindIII restriction enzyme, and water according to Manufacturer's ddPCR protocol. Sample DNA was added to reaction

wells at inputs of 0.02 to 20 ng per well to ensure that at least one reaction per sample was within the range of the assay. Droplets were generated by the automated QX200 AutoDG system and thermal cycled according to Manufacturer's recommended parameters, with an annealing T_m of 60°C. Thermal cycled droplets were analyzed by Direct Quantitation on the QX200 Droplet Reader. Sample data was inspected for sufficient droplets (>10,000) and properly identified positive and negative droplets prior to analysis. The QX200 droplet reader software automatically calculates the target (EGFP) copies per μL of reaction volume. The copy number of AAV-GFP genomes/ng DNA (y) was calculated by $y = (a*b)/c$, where a = EGFP copies per μL , b = 20 μL total reaction volume, and c = ng DNA per reaction volume.

c. Cap Copy Number ddPCR Assay: The Cap Copy Number ddPCR Assay quantifies the copy number of AAV2 Cap genes per genome of HEK293-based cells. Genomic DNA was extracted from samples with the Zymo Research Quick-DNA MiniPrep Kit (D3025) according to the manufacturer's protocol. Sample DNA concentrations were quantified by the Thermo Qubit 1X dsDNA Broad Range Assay Kit (Q33266) according to the manufacturer's protocols and normalized to the same concentration in TE. ddPCR reactions were assembled with Manufacturer's Supermix for Probes (no dUTP), a FAM-labeled hydrolysis probe-based ddPCR assay for AAV2 Cap, a HEX-labeled hydrolysis probe-based ddPCR assay for RPP30, HindIII restriction enzyme, and water according to Manufacturer's ddPCR protocol. Sample DNA was added to reaction wells at inputs of 4 ng and 20 ng per well to ensure that at least one reaction per sample was within the range of the assay. Droplets were generated by the automated QX200 AutoDG system and thermal cycled according to Manufacturer's recommended parameters, with an annealing T_m of 60°C. Thermal cycled droplets were analyzed by the CNV algorithm on the QX200 Droplet Reader, with the reference RPP30 assay set to 3 copies per genome (RPP30 is located on Chr10, which is triploid in the HEK293 cell line). Sample data was inspected for sufficient droplets (>10,000) and properly identified positive and negative droplets prior to analysis. The QX200 droplet reader software automatically calculates the copies per μL of reaction volume for the target (AAV2 Cap) and reference (RPP30) and reports the target (AAV2 Cap) copies per genome using the reference RPP30 = 3 copies per genome.

d. AAV physical titer ddPCR: The AAV Physical Titer ddPCR Assay quantifies the copy number of AAV-GFP genomes per mL of AAV sample analyzed. A series of eight 10-fold serial dilutions (10E0-10E-7) of AAV samples were created to ensure that several dilutions would fall within the range of ddPCR quantitation. Serial dilutions were created in nuclease-free water with 0.05% Pluronic F-68 (Gibco 24040-032). AAV particles were lysed by incubation at 95°C for 10 minutes, followed by cooling to 4°C at a rate of 3°C per minute. ddPCR reactions were assembled with Manufacturer's Supermix for Probes (no dUTP), a FAM-labeled hydrolysis probe-based ddPCR assay for EGFP, HindIII restriction enzyme, and water according to Manufacturer's ddPCR protocol. AAV sample dilutions were added to reaction wells at 5 μL per well. Droplets were generated by the automated QX200 AutoDG system and

thermal cycled according to Manufacturer's recommended parameters, with an annealing T_m of 60°C. Thermal cycled droplets were analyzed by Direct Quantitation on the QX200 Droplet Reader. Sample data was inspected for sufficient droplets (>10,000) and properly identified positive and negative droplets prior to analysis. The QX200 droplet reader software automatically calculates the target (EGFP) copies per uL of reaction volume. The titer of AAV genomes per uL of sample (y) was calculated by $y = (20 \times A \times B \times 1000) / C$, where 20 = total uL of reaction volume, A = AAV-GFP genome copies/uL of reaction volume, B = dilution factor of AAV sample, 1000 = uL per mL (conversion factor), and C = uL of AAV sample dilution in reaction volume.

Western Blot analyses for detecting AAV Rep and Cap Expression:

a. Cell lysate preparation: Upon wash and scrape, $\sim 1 \times 10^7$ cells in 10 ml of same PBS containing 1mM PMSF and 100ug/ml TPCK or Halt™ Protease and Phosphatase Inhibitor Cocktail, EDTA-free (100X) (Cat# 78441) were pelleted down by centrifugation at 3000 rpm for 15 min at 4°C and resuspend in 600 ul of RIPA buffer containing 1mM PMSF, 100ug/ml TPCK, and 50ug/ml leupeptin or Pierce Protease and Phosphatase Inhibitor Mini Tablets (cat# A32959). Samples were Incubated on ice for 40 min -- vortex each tube for 10 seconds every 15 minutes followed by centrifugation at 13,000 rpm for 2 min at 4°C. The supernatants were quick-frozen down in 100 ul to 150 ul aliquots with dry ice and stored at -80 C.

b. Protein concentration, SDS-PAGE and western blot analyses: Protein concentration from each cell lysate supernatants were determined in triplicates using Manufacturer DC Protein Quantitation Assay (Cat no: 5000112) according to the manufacturer's instruction. Upon heat treatment in 1x laemmli buffer containing 10% BME and 0.5M UREA, equal amounts of protein samples (up to 100ug /well) were loaded onto Manufacturer Criterion TGX gradient (8-16% or 4-20%) gels, separated by electrophoresis at 85V for ~2.5hr followed by protein transfer to PVDF membrane using Manufacturer Trans-Blot Turbo system. Blocking and antibody incubation steps were performed using EveryBlot Blocking Buffer (cat# 1201002). AAV Rep, AAV Cap, and host cell Cyclophilin B (a loading control) on PVDF membranes were probed with mouse anti-AAV2 Rep (clone 303.9, Progen Catlog# 61069 at 1:100x), mouse anti-AAV VP1/VP2/VP3 (clone B1, Progen 65158, supernatant at 1:25x), or mouse anti-Cyclophilin B (Abcam, ab236760 at 1:10,000x) followed by HRP conjugated secondary goat-anti-mouse IgG H&L antibody (AbCam, ab6789 at 2000x to 10,000x). Upon adding developing solution (1:1 peroxide SuperSignal™ West Pico PLUS Chemiluminescent Substrate for REP/CAP, or 1:1 peroxide: enhancer: ECL for Cyclophilin B), the western blot images were visualized and captured using iBright Imaging System.

GFP images

a. GFP expression images were captured using Zeiss AxioVert A1 microscope, AxioCam 705 mono camera, 10x objective, 150ms GFP exposure.

EQUIVALENTS

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features herein set forth and as follows in the scope of the appended claims.

Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific embodiments described specifically herein. Such equivalents are intended to be encompassed in the scope of the following claims.

INCORPORATION BY REFERENCE

All patents and publications referenced herein are hereby incorporated by reference in their entireties.

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

As used herein, all headings are simply for organization and are not intended to limit the disclosure in any manner. The content of any individual section may be equally applicable to all sections.

CLAIMS

What is claimed is:

1. A method of making a viral particle packaging and producer cell line, the method comprising:
transfecting a cell with two or more nucleic acids, wherein the nucleic acids encode one or more of:
 - (a) an enzyme capable of performing targeted genomic integration,
 - (b) an inducible viral replication (Rep) gene,
 - (c) an inducible viral Capsid (cap) gene,
 - (d) one or more adenoviral auxiliary genes, optionally selected from one or more of E1A, E1B, E4, E2A, and VA of an AAV, optionally E1A and E1B,
 - (e) an insulator, optionally selected from HS4, D4Z4,
 - (f) one or more terminal ends recognized by the enzyme, and
 - (g) a transgene flanked by AAV inverted terminal repeats (ITRs),wherein the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof,
to thereby result in a transfected cell that expresses the transgene in association with the viral particle.
2. A method of making a viral particle packaging and producer cell line, the method comprising:
transfecting an E1A, E1B+ cell with two or more nucleic acids, wherein the nucleic acids encode one or more of:
 - (a) an enzyme capable of performing targeted genomic integration,
 - (b) an inducible viral replication (Rep) gene,
 - (c) an inducible viral Capsid (cap) gene,
 - (d) an insulator, optionally selected from HS4, D4Z4,
 - (e) one or more terminal ends recognized by the enzyme, and
 - (f) a transgene flanked by AAV inverted terminal repeats (ITRs),

wherein the inducible viral replication (Rep) and Capsid (cap) genes are controlled by at least one of a tetO-containing P5 promoter, P19 promoter, SV40 promoter, P40 promoter, or a functional variant or fragment thereof,

to thereby result in a transfected cell that expresses the transgene in association with the viral particle.

3. The method of claim 1, wherein the cell is transfected with:

(1) two nucleic acids, and wherein the first nucleic acid encodes (a), and the second nucleic acid encodes (b), (c), (d), (e), (f), and (g);

(2) three nucleic acids, and wherein the first nucleic acid encodes (a), the second nucleic acid encodes (b), (c), (d), (e), and (f), and the third nucleic acid encodes (g); or

(3) four nucleic acids, and wherein the first nucleic acid encodes (a), the second nucleic acid encodes (b), (d), (e), and (f), the third nucleic acid encodes (c), and the fourth nucleic acid encodes (g).

4. The method of claim 2, wherein the cell is transfected with:

(1) two nucleic acids, and wherein the first nucleic acid encodes (a), and the second nucleic acid encodes (b), (c), (d), (e), and (f);

(2) three nucleic acids, and wherein the first nucleic acid encodes (a), the second nucleic acid encodes (b), (c), (d), and (e), and the third nucleic acid encodes (f); or

(3) four nucleic acids, and wherein the first nucleic acid encodes (a), the second nucleic acid encodes (b), (d), and (e), the third nucleic acid encodes (c), and the fourth nucleic acid encodes (f).

5. The method of claim 1 or claim 2, wherein the cell is human embryonic kidney (HEK293), Chinese hamster ovary (CHO) E1A, E1B+ engineered CHO-K1, or *Spodoptera frugiperda* (Sf9) cell line, baby hamster kidney (BHK), vero cell.

6. The method of claim 1, wherein the inducible viral replication (Rep) and Capsid (cap) genes are controlled by a tetO-containing P40 promoter.

7. The method of claim 6, wherein the tetO-containing P40 promoter, or a functional variant or fragment thereof, further comprises a modified TATA box element and a modified P40 intron.

8. The method of claim 7, wherein the tetO-containing P40 promoter is a tetO-containing AAV2 P40 promoter.

9. The method of claim 8, wherein the tetO-containing AAV2 P40 promoter comprises a nucleotide sequence of SEQ ID NO: 816, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto or at least about 10 mutations, or at least about

9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

10. The method of any one of claims 6-9, wherein the P40 intron is an AAV2 P40 intron.

11. The method of claim 10, wherein the AAV2 P40 intron comprises a C276A substitution, or a substitution at position corresponding thereto relative to SEQ ID NO: 818.

12. The method of claim 10 or 11, wherein the AAV2 P40 intron comprises one or more mutated translation start sites (ATGs), optionally wherein the translation start sites are mutated to one of CTG, ACG, and TTG.

13. The method of any one of claims 10-12, wherein the AAV2 P40 intron comprises substitutions at one or more positions A13, A32, T42, A61, A71, A89, A203, A246, A258, and T282, or one or more positions corresponding thereto, relative to SEQ ID NO: 818.

14. The method of any one of claims 10-13, wherein the AAV2 P40 intron comprises substitutions at one or more positions A13C, A32C, T42C, A61C, A71T, A89C, A203T, A246C, A258C, and T282C corresponding to SEQ ID NO: 818.

15. The method of any one of claims 10-14, wherein the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 818, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

16. The method of any one of claims 10-15, wherein the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 819, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

17. The method of any one of claims 10-15, wherein the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 817, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at

least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

18. The method of any one of claims 6-17, wherein the modified TATA box element has the nucleotide sequence of TATATAA.

19. The method of any one of claims 6-18, wherein the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 823, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

20. The method of claim 6, wherein the tetO-containing P40 promoter is a tetO-containing AAV2 P40 promoter.

21. The method of claim 20, wherein the AAV2 P40 promoter comprises the nucleotide sequence of SEQ ID NO: 820, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

22. The method of claim 20 or 21, wherein the modified P40 intron is a AAV9 P40 intron.

23. The method of claim 22, wherein the modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 821, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

24. The method of any one of claims 20-23, wherein the modified TATA box element has the nucleotide sequence of TATATAA.

25. The method of any one of claims 20-24, wherein tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises the nucleotide sequence of SEQ ID NO: 822, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

26. The method of any one of claim 6-9 or 22-24, wherein the tetO-containing AAV2 P40 promoter or a functional variant or fragment thereof, comprising a modified TATA box element and a modified AAV9 P40 intron comprises a nucleotide sequence of SEQ ID NO: 824, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

27. The method of any one of claims 6-26, wherein the P40 promoter further comprises one or more enhancer elements.

28. The method of claim 27, wherein the enhancer element comprises one or more cis-acting elements, optionally selected from an Sp1 binding site, GC rich sequence, GCGGAAC motif, TAATGARAT element, AP1 binding site, and CCAAT box element.

29. The method of claim 27 or 28, wherein the enhancer element comprises about 1 to about 5 Sp1 binding sites, optionally about 1, or about 2, or about 3, or about 4, or about 5 Sp1 binding sites.

30. The method of any one of claims 27-29, wherein the enhancer element comprises about 1 or about 2 GC rich sequences.

31. The method of any one of claims 27-30, wherein the enhancer element is derived from an hCMV Enhancer Element-3.

32. The method of claim 31, wherein the hCMV Enhancer Element-3 comprises the nucleotide sequence of SEQ ID NO: 830 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9

mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

33. The method of claim 1, wherein the inducible promoter is an antibiotic-dependent promoter, optionally a tetracycline-dependent promoter or a variant thereof, or a cumate or coumermycin/novobiocin promoter or a variant thereof.

34. The method of any one of claims 1-33, wherein the viral particle is an AAV, and optionally wherein the AAV is selected from AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9.

35. The method of any one of claims 1-34, wherein any one or more of the first, second, and third nucleic acids is encoded by a single nucleic acid.

36. The method of any one of claims 1-35, wherein the second nucleic acid and/or the third nucleic acid are included in a single expression vector.

37. The method of any one of claims 1-36, wherein the first and second nucleic acids are included in a single expression vector, and the third nucleic acid is included in an expression vector that is different from the expression vector including the first and second nucleic acids.

38. The method of claim 36 or claim 37, wherein the expression vector is or comprises a plasmid.

39. The method of claim 38, wherein the plasmid comprises the nucleotide sequence of SEQ ID NO: 825 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

40. The method of claim 38, wherein the plasmid is substantially of the form as show in **FIG. 8**, or a functional equivalent thereof.

41. The method of claim 38, wherein the plasmid comprises a AAV CAP under the control of an adeno-associated viral (AAV) system encoding AAV2 CAP comprising: a tetO-containing P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron, wherein the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 818 or a functional variant or fragment thereof, an hCMV-derived enhancer element, a Zeocin resistance gene, a WPRE regulatory element, and a cassette flanked with Sleeping beauty ITRs

42. The method of claim 38, wherein the plasmid comprises the nucleotide sequence of SEQ ID NO: 826 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

43. The method of claim 38, wherein the plasmid is substantially of the form as show in **FIG. 9**, or a functional equivalent thereof.

44. The method of claim 38, wherein the plasmid comprises a AAV CAP under the control of an adeno-associated viral (AAV) system encoding AAV2 CAP comprising: a tetO-containing P40 promoter with a modified AAV2 P40 TATA element and modified AAV2 P40 intron, wherein the AAV2 P40 intron comprises the nucleotide sequence of SEQ ID NO: 819 or a functional variant or fragment thereof, an hCMV-derived enhancer element, a Zeocin resistance gene, a WPRE regulatory element, and a cassette flanked with Sleeping beauty ITRs.

45. The method of claim 38, wherein the plasmid comprises the nucleotide sequence of SEQ ID NO: 841 or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

46. The method of claim 38, wherein the plasmid is substantially of the form as show in **FIG. 7**, or a functional equivalent thereof.

47. The method of claim 38, wherein the plasmid comprises a tetO-containing P5 promoter, a tetO-containing SV40 early promoter, a modified TATA box element, an enhancer element derived from a HSV-1 ICP4 promoter, a hygromycin B resistance gene and a cassette flanked with inverted terminal repeats (ITRs) derived from an engineered mobile element enzyme.

48. The method of claim 38, wherein the plasmid comprises the nucleotide sequence of SEQ ID NO: 842, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

49. The method of claim 38, wherein the plasmid is substantially of the form as show in **FIG. 5**, or a functional equivalent thereof.

50. The method of claim 38, wherein the plasmid comprises comprises a tetO-containing P5 promoter, a P40 promoter, a tetO-containing SV40 containing promoter, a modified TATA box element, a HSV-2 ICP4 promoter, a hygromycin B resistance gene, and a cassette flanked with inverted terminal repeats (ITRs) derived from an engineered mobile element enzyme.

51. The method of claim 38, the plasmid comprises the nucleotide sequence of SEQ ID NO: 843, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations, or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

52. The method of claim 38, wherein the plasmid is substantially of the form as show in **FIG. 6**, or a functional equivalent thereof.

53. The method of claim 38, wherein the plasmid comprises a tetO-containing P5 promoter, a tetO-containing P40 promoter with AAV9 P40 intron, a tetO-containing SV40 containing promoter, a modified TATA box element, a HSV-1 ICP4 promoter, a hygromycin B resistance gene, and a cassette flanked with inverted terminal repeats (ITRs) derived from an engineered mobile element enzyme.

54. The method of any one of claims 1-53, wherein one or more of the first, second, and third nucleic acids is or comprises RNA, optionally mRNA, optionally synthetic mRNA or modified mRNA.

55. The method of any one of claims 1-54, wherein one or more of the first, second, and third nucleic acids is DNA, optionally plasmid DNA.

56. The method of any one of claims 1-55, wherein one or more of the first, second, and third nucleic acids is an expression vector, wherein the expression vector is optionally a plasmid.

57. The method of any one of claims 1-56, wherein the transgene is flanked by AAV inverted terminal repeats (ITRs).

58. The method of claim 1 or 57, wherein the terminal ends or ITRs comprise the nucleotide sequence of SEQ ID NO: 831 and/or SEQ ID NO: 832, or a functional variant or fragment thereof, or sequence having at least about 70%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97% identity thereto, or at least about 98% identity thereto, or at least about 99% identity thereto, or at least about 10 mutations, or at least about 9 mutations, or at least about 8 mutations, or at least about 7 mutations, or at least about 6 mutations,

or at least about 5 mutations, or at least about 4 mutations, or at least about 3 mutations, or at least about 2 mutations, or at least about 1 mutation.

59. The method of any one of claims 1-58, wherein the transgene has a size of about 200,000 bases or less.

60. The method of any one of claims 1-59, further comprising culturing the transfected cell in a medium that expands a population of the transfected cells to create a stably transfected packaging and producer cell line.

61. The method of claim 60, wherein the stably transfected producer cell line is capable of producing replication-deficient viral particles in association with the transgene.

62. The method of any one of claims 1-61, wherein the transfection comprises electroporation, nucleofection, lipofection, or calcium phosphate transfection.

63. The method of any one of claims 1-62, wherein the method is helper virus-free.

64. The method of any one of claims 1-63, wherein the enzyme capable of performing targeted genomic integration is a recombinase.

65. The method of claim 64, wherein the recombinase is an integrase or a mobile element enzyme.

66. The method of claim 64 or 65, wherein the enzyme capable of performing targeted genomic integration is a mobile element enzyme.

67. The method of any one of claims 1-66, wherein the enzyme is derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

68. The method of any one of claims 1-67, wherein the enzyme is an engineered version, including but not limited to hyperactive forms, of an enzyme derived from *Bombyx mori*, *Xenopus tropicalis*, *Trichoplusia ni*, *Myotis lucifugus*, *Rhinolophus ferrumequinum*, *Rousettus aegyptiacus*, *Phyllostomus discolor*, *Myotis myotis*, *Pteropus vampyrus*, *Pipistrellus kuhlii*, *Molossus molossus*, *Pan troglodytes*, or *Homo sapiens*.

69. The method of any one of claims 1-68, wherein the mobile element enzyme is from one or more of the Sleeping beauty, Tn1, Tn2, Tn3, Tn5, Tn7, Tn9, Tn10, Tn552, Tn903, Tn1000/Gamma-delta, Tn/O, tnsA, tnsB, tnsC, tniQ, iS10, ISS, IS911, Minos, piggyBac, Tol2, Mos1, Himar1, Hermes, Tol2, Minos, Tel, P-element, MuA, Ty1, Chapaev, transib, Tc1/mariner, or Tc3 donor DNA system, or biologically active fragments variants thereof, inclusive of hyperactive variants.

70. The method of any one of claims 1-69, wherein the mobile element enzyme has the amino acid sequence of SEQ ID NO: 1, or an amino acid sequence having at least about 80%, or an amino acid sequence having at least about

90%, or at least about 93%, or at least about 95%, or at least about 97%, or at least about 98%, or at least about 99% identity thereto.

71. The method of any one of claims 1-70, wherein the mobile element enzyme comprises an amino acid other than serine at the position corresponding to position 2 of SEQ ID NO: 1.

72. The method of claim 71, wherein the amino acid is a non-polar aliphatic amino acid, optionally a non-polar aliphatic amino acid optionally selected from G, A, V, L, I and P, optionally A.

73. The method of any one of claims 70-72, wherein the mobile element enzyme does not have additional residues at the C terminus relative to SEQ ID NO: 1.

74. The method of any one of claims 1-73, wherein the enzyme has one or more mutations which confer hyperactivity.

75. The method of any one of claims 70-74, wherein the enzyme has one or more amino acid substitutions selected from S8X₁ and/or C13X₂, or positions corresponding thereto relative to SEQ ID NO: 1.

76. The method of claim 75, wherein the enzyme has S8X₁ and/or C13X₂ substitutions, at positions corresponding thereto relative to SEQ ID NO: 1.

77. The method of claim 76, wherein the enzyme has S8X₁ and C13X₂ substitutions, at positions corresponding thereto relative to SEQ ID NO: 1.

78. The method of claim 77, wherein the enzyme has S8X₁ substitution, at position corresponding thereto relative to SEQ ID NO: 1.

79. The method of claim 78, wherein the enzyme has C13X₂ substitution, at positions corresponding thereto relative to SEQ ID NO: 1.

80. The method of any one of claims 75-79, wherein X₁ is selected from G, A, V, L, I, and P and X₂ is selected from K, R, and H.

81. The method of claim 80, wherein: X₁ is P and X₂ is R.

82. The method of any one of claims 66-81, wherein the enzyme comprises an amino acid sequence of SEQ ID NO: 11.

83. The method of any one of claims 66-82, wherein the enzyme comprises at least one, at least five, at least seven, at least nine, or ten substitutions at positions corresponding to: 5, 8, 9, 10, 11, 14, 22, 36, 37, 54, 130, 239, 281, 282, 283, 284, 285, 294, 300, 310, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 345, 375, 416, 427, 475, 481, 491, 520, and/or 561 of SEQ ID NO: 11.

84. The method of any one of claims 66-83, wherein the enzyme comprises one, two, three, four, five, six, seven, eight, nine, or ten substitutions at positions corresponding to: 5, 8, 9, 10, 11, 14, 22, 36, 37, 54, 130, 239, 281, 282, 283, 284, 285, 294, 300, 310, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 345, 375, 416, 427, 475, 481, 491, 520, and/or 561 of SEQ ID NO: 11.

85. The method of any one of claims 66-84, wherein the enzyme comprises at least one, at least five, at least seven, at least nine, or ten substitutions selected from S5P, S8F, D9G, D10G, E11G, A14V, T22C, S36G, T37C, S54N, K130T, G239R, Y281A, C282A, G283A, E284A, G285A, T294A, T300A, N310A, G330A, T331A, I332A, R333A, K334A, N335A, R336A, G337A, I338A, P339A, I345V, T375G, D416A, R427H, D475G, M481V, P491Q, A520T, and A561T, wherein the positions are corresponding to positions of SEQ ID NO: 11.

86. The method of any one of claims 66-85, wherein the enzyme comprises one, two, three, four, five, six, seven, eight, nine, or ten substitutions selected from S5P, S8F, D9G, D10G, E11G, A14V, T22C, S36G, T37C, S54N, K130T, G239R, Y281A, C282A, G283A, E284A, G285A, T294A, T300A, N310A, G330A, T331A, I332A, R333A, K334A, N335A, R336A, G337A, I338A, P339A, I345V, T375G, D416A, R427H, D475G, M481V, P491Q, A520T, and A561T, wherein the positions are corresponding to positions of SEQ ID NO: 11.

87. The method of any one of claims 66-86, wherein the mobile element enzyme is an engineered mammalian mobile element enzyme.

88. The method of any one of claims 66-87, wherein the mobile element enzyme is a mammal-derived, helper RNA mobile element enzyme.

89. The method of any one of claims 66-88, wherein the mobile element enzyme is a mammal-derived, helper DNA mobile element enzyme.

90. The method of any one of claims 64-89, wherein the enzyme is capable of inserting a donor DNA at a TA dinucleotide site.

91. The method of any one of claims 64-90, wherein the enzyme is capable of inserting a donor DNA at a TTAA (SEQ ID NO: 440) tetranucleotide site.

92. The method of any one of claims 64-91, wherein the mobile element enzyme has gene cleavage activity (Exo+) and/or gene integration activity (Int+), and the mobile element enzyme having at least about 90% identity to the amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 430, or a nucleotide sequence encoding the same.

93. The method of claim 92, wherein the mobile element enzyme has at least about 95%, or at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to the amino acid sequence of SEQ ID NO: 1, SEQ

ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 430.

94. The method of claim 92 or 93, wherein the mobile element enzyme has one or more mutations which confer hyperactivity.

95. The method of any one of claims 92-94, wherein the mobile element enzyme has an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 10, or SEQ ID NO: 11 or a functional equivalent thereof.

96. The method of one of claims 92-94, wherein the mobile element enzyme has the nucleotide sequence having at least about 90% identity to SEQ ID NO: 5 or a codon-optimized form thereof.

97. The method of claim 93, wherein the mobile element enzyme has an amino acid sequence having S8P and G17R mutations relative to the amino acid sequence of SEQ ID NO: 3 or SEQ ID NO: 4, or a functional equivalent thereof.

98. The method of claim 93, wherein the mobile element enzyme has an amino acid sequence having I83P and/or V118R mutation relative to the amino acid sequence of SEQ ID NO: 6 or a functional equivalent thereof.

99. The method of claim 93, wherein the mobile element enzyme has an amino acid sequence having S20P and/or A29R mutation relative to the amino acid sequence of SEQ ID NO: 7 or a functional equivalent thereof.

100. The method of claim 93, wherein the mobile element enzyme has an amino acid sequence having A12P and/or I28R mutation and/or R152K mutation relative to the amino acid sequence of SEQ ID NO: 9 or a functional equivalent thereof.

101. The method of claim 93, wherein the mobile element enzyme has an amino acid sequence having T4P and/or L13R mutation relative to the amino acid sequence of SEQ ID NO: 8 or a functional equivalent thereof.

102. The method of any one of claims 66-101, wherein the donor DNA is included in a vector comprising left and right end sequences recognized by the mobile element enzyme.

103. The method of claim 102, wherein the end sequences are selected from MER, MER75A, MER75B, and MER85.

104. The method of claim 103, wherein the end sequences are selected from nucleotide sequences of SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 441, and SEQ ID NO: 22, or a nucleotide sequence having at least about 90% identity thereto.

105. The method of any one of claims 102-104, wherein one or more of the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

106. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 12, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 12 is positioned at the 5' end of the donor DNA.

107. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 17, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 17 is positioned at the 3' end of the donor DNA.

108. The method of claim 106 or 107, wherein the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

109. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 13, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 13 is positioned at the 5' end of the donor DNA.

110. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 18, and wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 18 is positioned at the 3' end of the donor DNA.

111. The method of claim 109 or 110, wherein the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

112. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 14, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 14 is positioned at the 5' end of the donor DNA.

113. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 19, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 19 is positioned at the 3' end of the donor DNA.

114. The method of claim 112 or 113, wherein the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.

115. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 15, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 15 is positioned at the 5' end of the donor DNA.
116. The method of claim 104, wherein end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 20, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 20 is positioned at the 3' end of the donor DNA.
117. The method of claim 115 or 116, wherein the end sequences are optionally flanked by a TTAA (SEQ ID NO: 440) sequence.
118. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 16, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 16 is positioned at the 5' end of the donor DNA.
119. The method of claim 104, wherein the end sequences include at least one repeat from a nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441, wherein the at least one repeat from the nucleotide sequence having at least about 90% identity to the nucleotide sequence of SEQ ID NO: 21 or SEQ ID NO: 441 is positioned at the 3' end of the donor DNA.
120. The method of any one of claims 104-119, wherein the mobile element enzyme is an engineered form of a mobile element enzyme reconstructed from *Homo sapiens* or a predecessor thereof.
121. The method of any one of claims 1-120, wherein the enzyme is in a monomeric or dimeric form.
122. The method of any one of claims 1-121, wherein the enzyme is in a multimeric form.
123. The method of any one of claims 1-122, wherein the enzyme comprises:
- (a) a targeting element, and
 - (b) an enzyme that is capable of inserting a donor DNA (e.g. a mobile element) comprising a transgene, optionally at a TA dinucleotide site or a TTAA (SEQ ID NO: 440) tetranucleotide site in a genomic safe harbor site (GSHS).
124. The method of any one of claims 1-123, wherein the donor DNA comprises a transgene encoding a complete polypeptide.

125. The method of any one of claims 1-124, wherein the donor DNA comprises a transgene which is defective or substantially absent in a disease state.
126. The method of any one of claims 1-125, wherein the enzyme has one or more mutations which confer hyperactivity.
127. The method of any one of claims 1-126, wherein the enzyme has gene cleavage activity (Exc+) and/or gene integration activity (Int+).
128. The method of any one of claims 1-127, wherein the enzyme has gene cleavage activity (Exc+) and/or a lack of gene integration activity (Int-).
129. The method of any one of claims 1-128, wherein the mobile element enzyme is a chimeric mobile element enzyme.
130. The method of any one of claims 123-129, wherein the targeting element comprises one or more of a gRNA, optionally associated with a Cas enzyme, which is optionally catalytically inactive, transcription activator-like effector (TALE), catalytically inactive Zinc finger, catalytically inactive transcription factor, nickase, a transcriptional activator, a transcriptional repressor, a recombinase, a DNA methyltransferase, a histone methyltransferase, a paternally expressed gene 10 (PEG10), and a TnsD.
131. The method of any one of claims 123-130, wherein the targeting element comprises a transcription activator-like effector (TALE) DNA binding domain (DBD).
132. The method of claim 131, wherein the TALE DBD comprises one or more repeat sequences.
133. The method of claim 132, wherein the TALE DBD comprises about 14, or about 15, or about 16, or about 17, or about 18, or about 18.5 repeat sequences.
134. The method of claim 132 or claim 133, wherein the TALE DBD repeat sequences comprise 33 or 34 amino acids.
135. The method of claim 131, wherein the one or more of the TALE DBD repeat sequences comprise a repeat variable di-residue (RVD) at residue 12 or 13 of the 33 or 34 amino acids.
136. The method of claim 135, wherein the RVD recognizes one base pair in the nucleic acid molecule.
137. The method of claim 135, wherein the RVD recognizes a C residue in the nucleic acid molecule and is selected from HD, N(gap), HA, ND, and HI.
138. The method of claim 135, wherein the RVD recognizes a G residue in the nucleic acid molecule and is selected from NN, NH, NK, HN, and NA.

139. The method of claim 135, wherein the RVD recognizes an A residue in the nucleic acid molecule and is selected from NI and NS.
140. The method of claim 135, wherein the RVD recognizes a T residue in the nucleic acid molecule and is selected from NG, HG, H(gap), and IG.
141. The method of any one of claims 123-140, wherein the GSHS is in an open chromatin location in a chromosome.
142. The method of any one of claims 123-140, wherein the GSHS is selected from adeno-associated virus site 1 (AAVS1), chemokine (C-C motif) receptor 5 (CCR5) gene, HIV-1 coreceptor, and human Rosa26 locus.
143. The method of any one of claims 123-140, wherein the GSHS is located on human chromosome 2, 4, 6, 10, 11, 17, 22, or X.
144. The method of any one of claims 123-140, wherein the GSHS is selected from TALC1, TALC2, TALC3, TALC4, TALC5, TALC7, TALC8, AVS1, AVS2, AVS3, ROSA1, ROSA2, TALER1, TALER2, TALER3, TALER4, TALER5, SHCHR2-1, SHCHR2-2, SHCHR2-3, SHCHR2-4, SHCHR4-1, SHCHR4-2, SHCHR4-3, SHCHR6-1, SHCHR6-2, SHCHR6-3, SHCHR6-4, SHCHR10-1, SHCHR10-2, SHCHR10-3, SHCHR10-4, SHCHR10-5, SHCHR11-1, SHCHR11-2, SHCHR11-3, SHCHR17-1, SHCHR17-2, SHCHR17-3, and SHCHR17-4.
145. The method of any one of claims 123-130, wherein the targeting element comprises a Cas9 enzyme guide RNA complex.
146. The method of claim 145, wherein the Cas9 enzyme guide RNA complex comprises a nuclease-deficient dCas9 guide RNA complex.
147. The method of any one of claims 123-130, wherein the targeting element comprises a Cas12 enzyme guide RNA complex or wherein the targeting element comprises a nuclease-deficient dCas12 guide RNA complex, optionally dCas12j guide RNA complex or dCas12a guide RNA complex.
148. The method of any one of claims 123-147, wherein the targeting element comprises: a gRNA of or comprising a sequence of **TABLE 3A-3F**, or a variant thereof; or a TALE DBD of or comprising a sequence of **TABLE 4A-4F**, or a variant thereof; or a ZNF of or comprising a sequence of **TABLE 5A-5E**, or a variant thereof.
149. The method of any one of claims 123-130, wherein the targeting element is or comprises a nucleic acid binding component of the gene-editing system.
150. The method of any one of claims 123-149, wherein the enzyme and the targeting element are connected.

151. The method of any one of claims 123-149, wherein the enzyme and the targeting element are fused to one another or linked via a linker to one another.
152. The method of claim 151, wherein the linker is a flexible linker.
153. The method of claim 152, wherein the flexible linker is substantially comprised of glycine and serine residues, optionally wherein the flexible linker comprises (Gly₄Ser)_n, where n is from about 1 to about 12.
154. The method of claim 152, wherein the flexible linker is of about 20, or about 30, or about 40, or about 50, or about 60 amino acid residues.
155. The method of any one of claims 1-154, wherein the nucleic acid comprises a gene encoding a complete polypeptide.
156. The method of any one of claims 1-155, wherein the nucleic acid comprises a gene which is defective or substantially absent in a disease state.
157. The method of any one of claims 1-156, wherein the transgene is flanked by one or more inverted terminal ends.
158. The method of any one of claims 1-157, wherein at least one of the first nucleic acid and the second nucleic acid is in the form of a lipid nanoparticle (LNP).
159. The method of any one of claims 1-158, wherein the first nucleic acid encoding the enzyme and the second nucleic acid are in the form of the same LNP, optionally in a co-formulation.
160. The method of claim 158 or claim 159, wherein the LNP comprises one or more lipids selected from 1,2-dioleoyl-3-trimethylammonium propane (DOTAP), a cationic cholesterol derivative mixed with dimethylaminoethane-carbamoyl (DC-Chol), phosphatidylcholine (PC), triolein (glyceryl trioleate), and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[carboxy(polyethylene glycol)-2000] (DSPE-PEG), 1,2-dimyristoyl-rac-glycero-3-methoxypolyethyleneglycol – 2000 (DMG-PEG 2K), and 1,2 distearol -sn-glycerol-3phosphocholine (DSPC) and/or comprising of one or more molecules selected from polyethylenimine (PEI) and poly(lactic-co-glycolic acid) (PLGA), and N-Acetylgalactosamine (GalNAc).
161. The method of any one of claims 1-160, wherein the enzyme is encoded by a recombinant or synthetic nucleic acid.
162. The method of claim 161, wherein the nucleic acid is mRNA or a helper RNA.
163. The method of claim 162, wherein the nucleic acid is RNA that has a 5'-m7G cap (cap0, cap1, or cap2) with pseudouridinesubstitution, and a poly-A tail of about 30, or about 50, or about 100, of about 150 nucleotides in length.

164. The method of claim 162, wherein the enzyme is incorporated into a vector or a vector-like particle.
165. The method of claim 164, wherein the vector is a non-viral vector.
166. The method of any of claims 1-165, wherein the enzyme and the one or more elements (a)-(g) are included in the same vector.
167. The method of any of claims 1-166, wherein the enzyme and the one or more elements (a)-(g) are included in different vectors.
168. The method of any one of claims 1-167, wherein the enzyme and the one or more elements (a)-(g) are included in a single pharmaceutical compositions.
169. The method of any one of claims 1-168, wherein the enzyme and the one or more elements (a)-(g) are included in different pharmaceutical compositions.
170. The method of any one of claims 1-169, wherein the enzyme and the one or more elements (a)-(g) are co-administered.
171. The method of any one of claims 1-170, wherein the enzyme and the one or more elements (a)-(g) are administered separately.
172. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line provides a stable cell with substantially reduced or ablated CAP expression.
173. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line provides a stable cell with REP expression or substantially enhanced REP expression.
174. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line provides a stable cell with substantially reduced or ablated CAP expression and REP expression or substantially enhanced REP expression.
175. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line provides a stable cell with REP expression or substantially enhanced VP1 expression.
176. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line provides a stable cell with substantially reduced or ablated CAP expression.
177. The method of any one of claims 1-171, wherein the viral particle packaging and producer cell line is suitable for providing substantially reduced empty or cargo-free capsid.
178. A method of producing an AAV bearing a gene of interest, comprising employing a method of any one of claims 1-177, to produce the AAV bearing the gene of interest.

179. A cell for gene therapy, generated by a method of any one of claims 1-178.
180. A method of delivering a cell therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of any one of claims 1-179.
181. A method of treating a disease or condition using a cell therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of any one of claims 1-180.
182. A method of treating a disease or condition using a biologic, e.g., antibody, therapy, comprising administering to a patient in need thereof the transfected cell generated by a method of any one of claims 1-181.

FIG. 1A

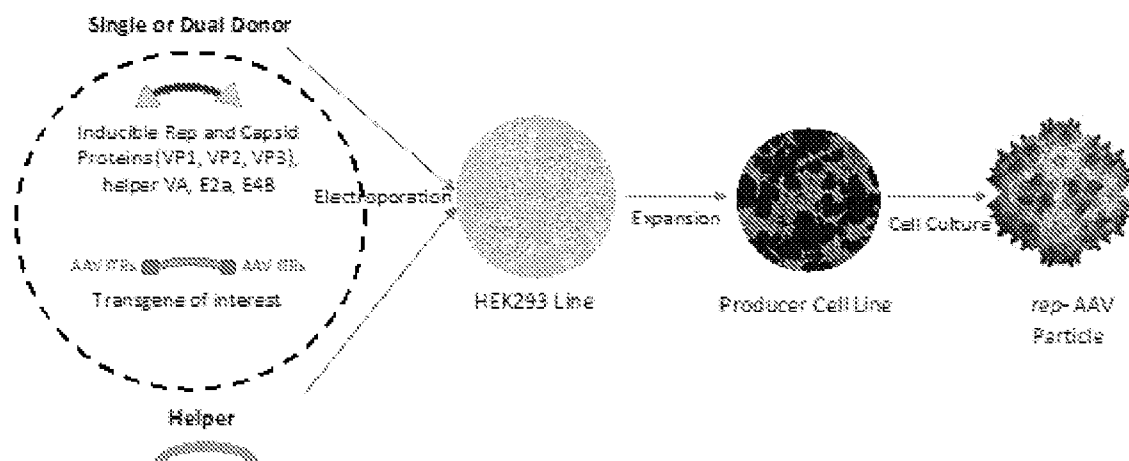


FIG. 1B

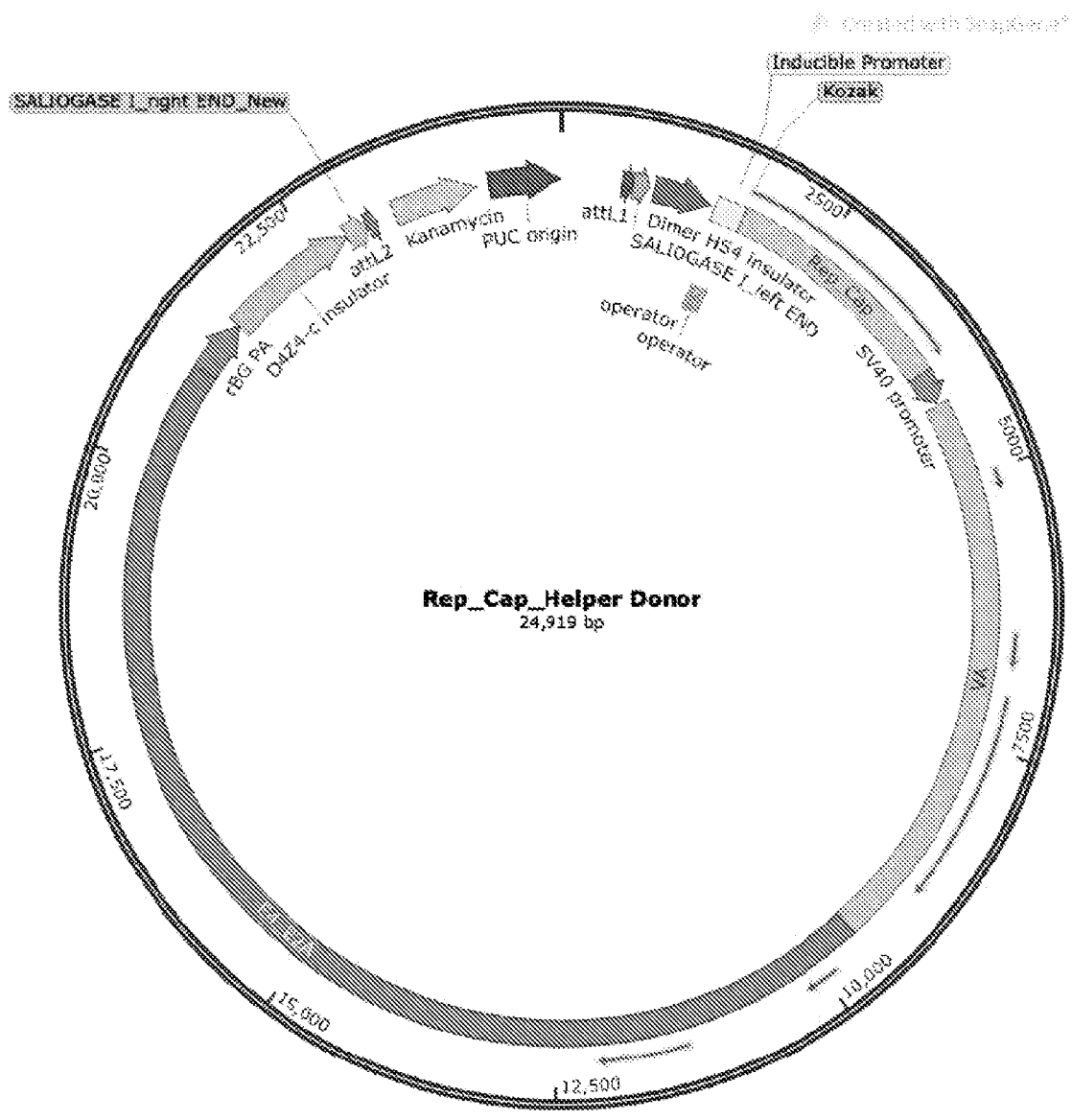


FIG. 1C

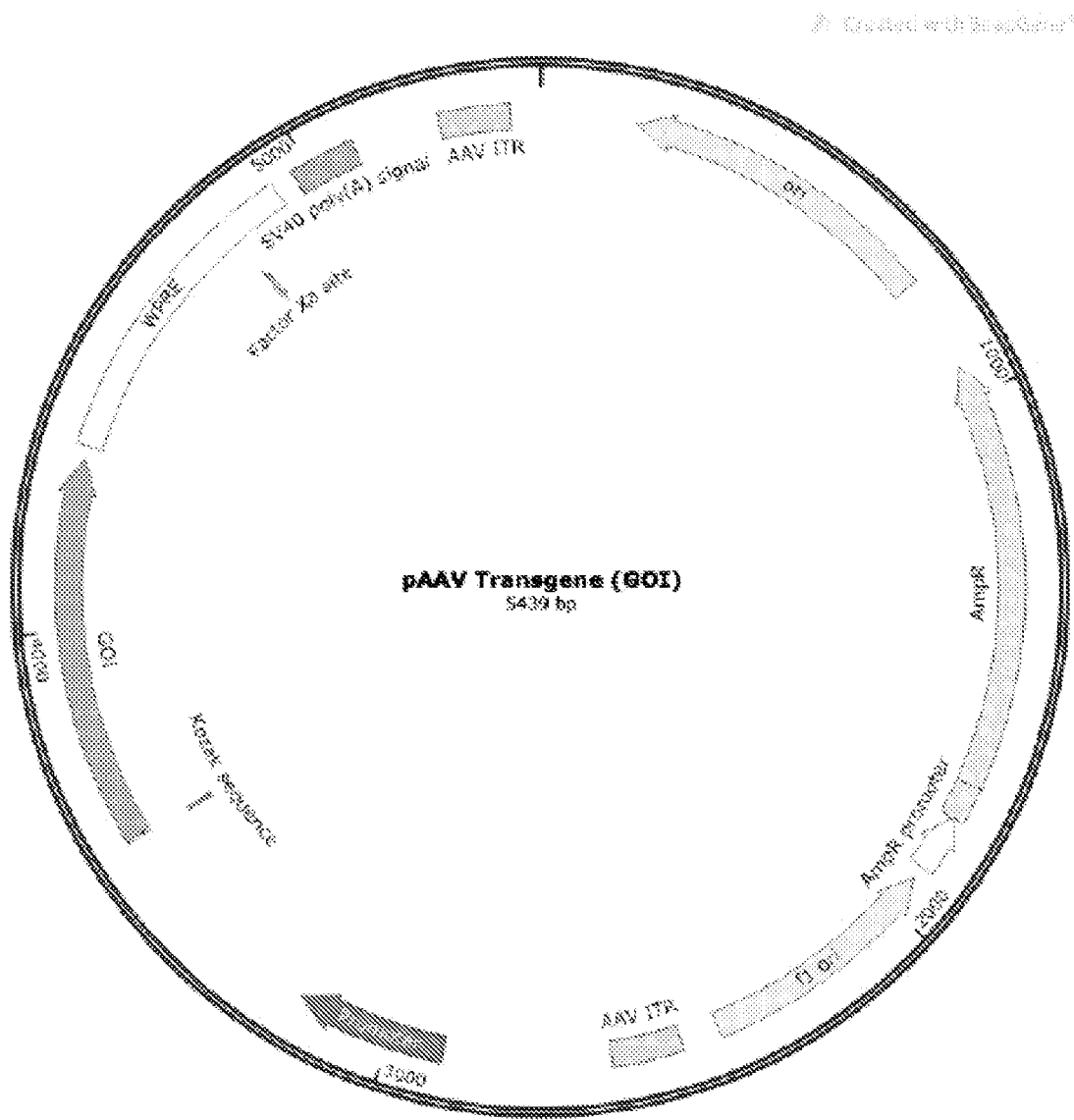


FIG. 2A

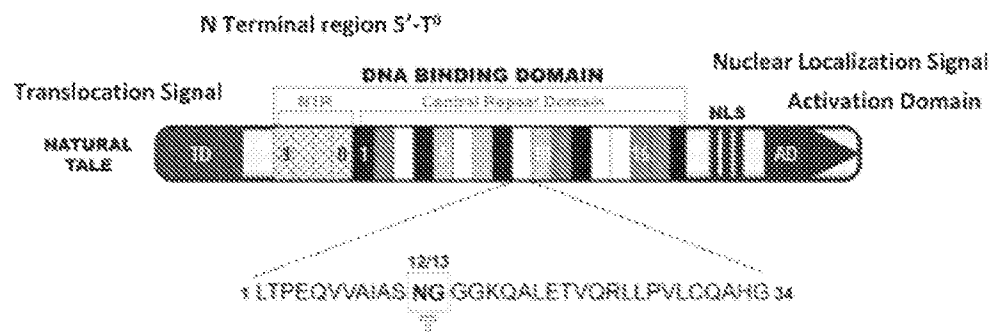


FIG. 2B

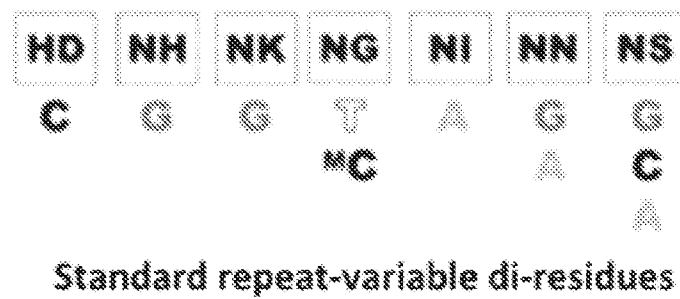


FIG. 2C

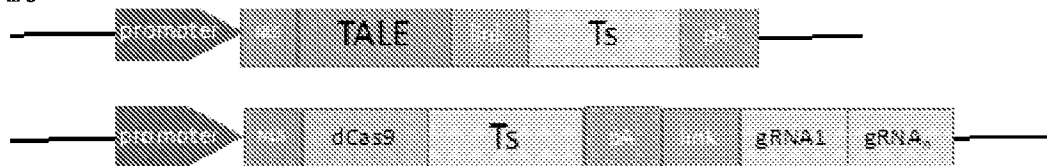


FIG. 2D

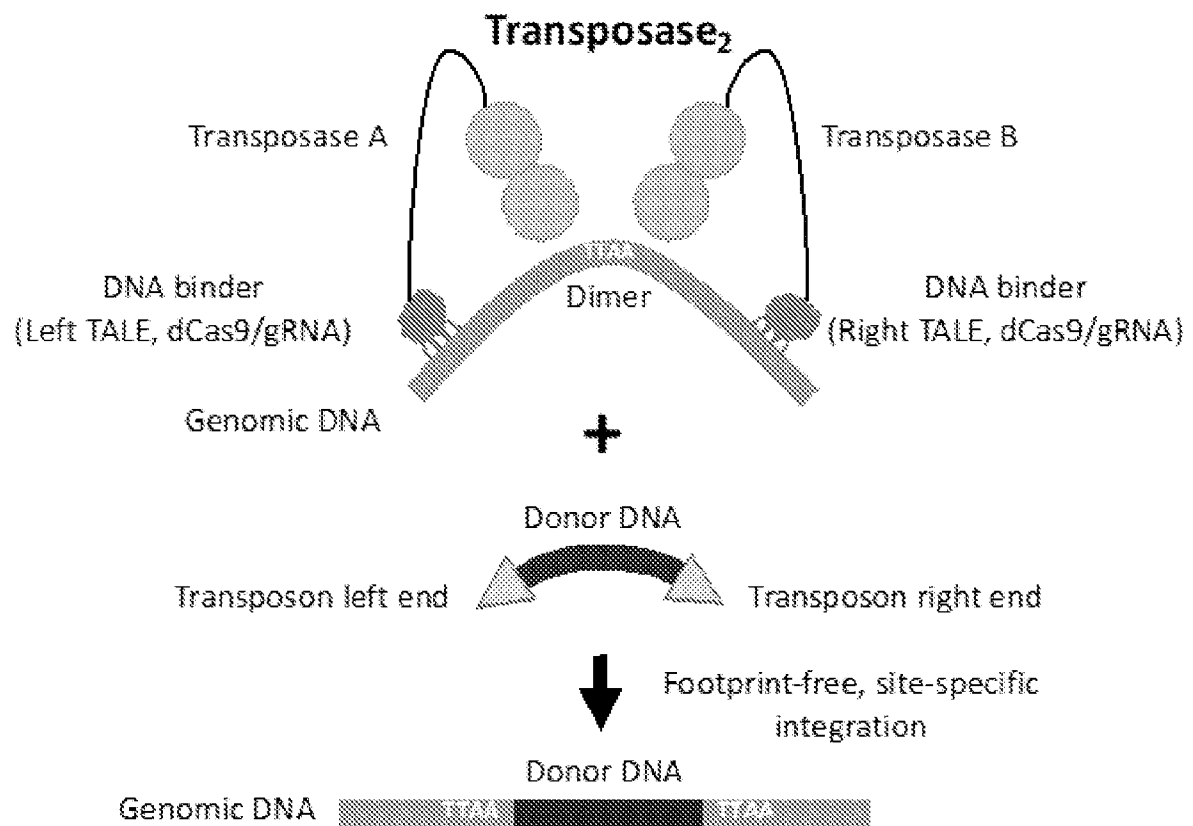


FIG. 2E

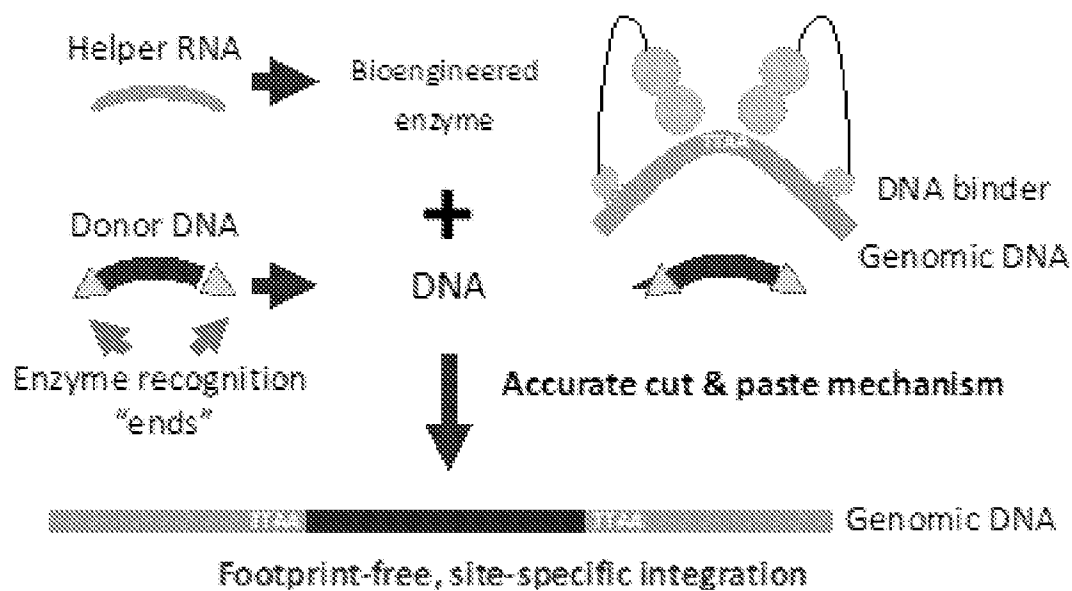


FIG. 3A. Plasmid construct template that transcribes helper RNA that is later processed with a 5'-m7G cap

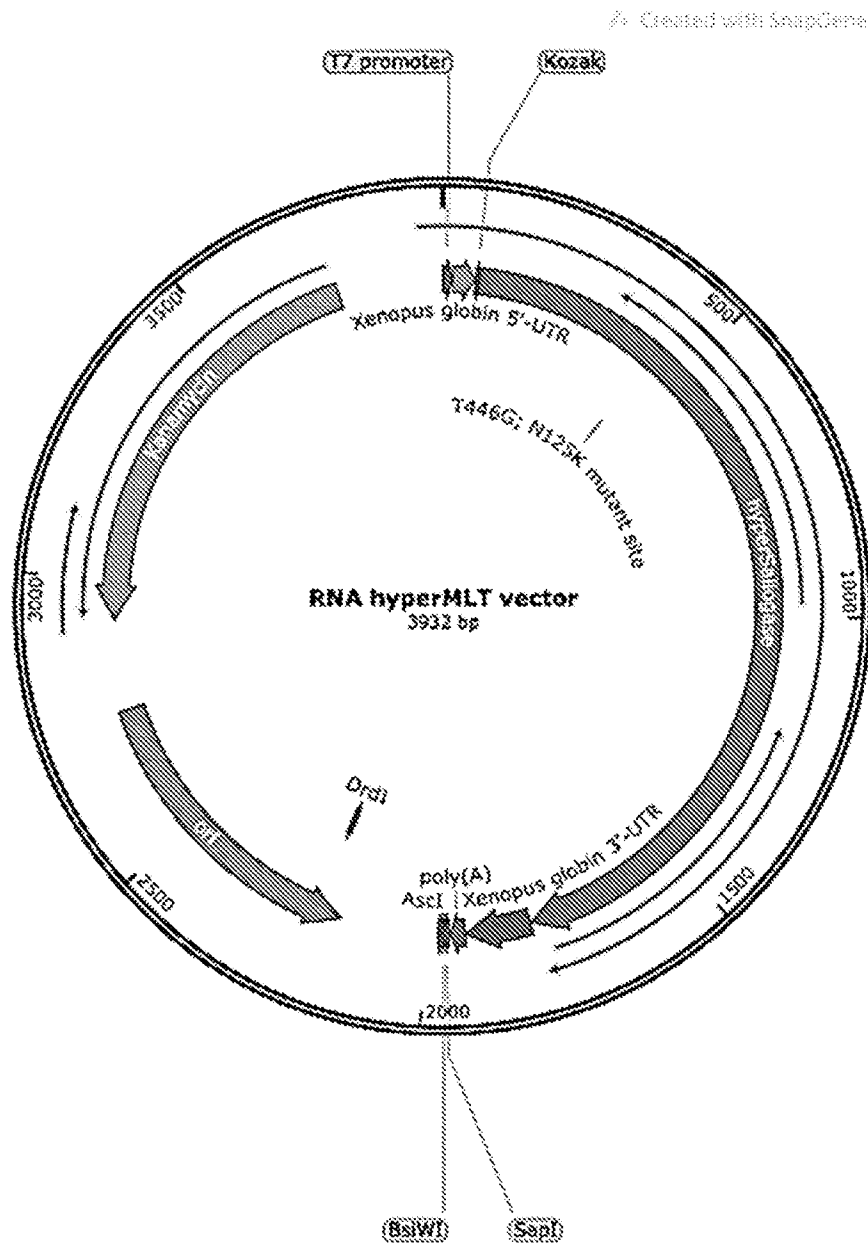


FIG. 3B. Donor DNA construct template with the transgene VLDLR

File Created with SoapShare®

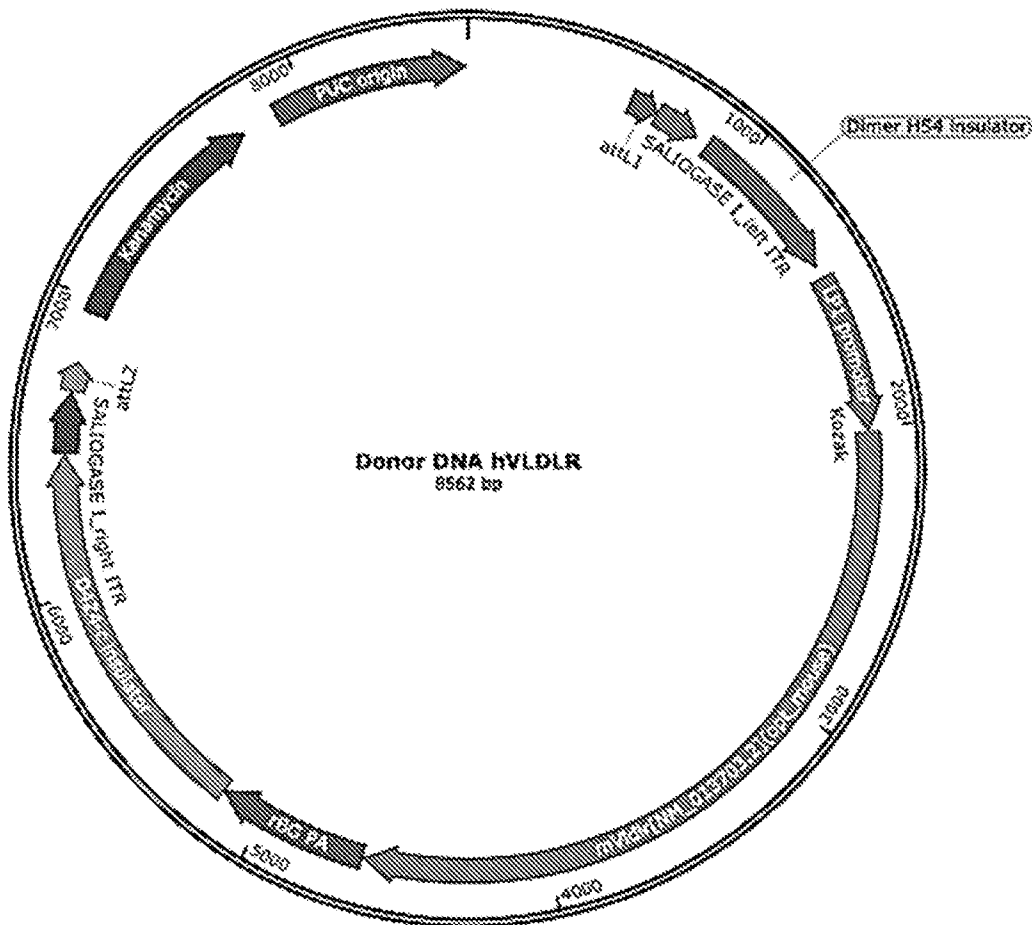


FIG. 3C. Chimeric mobile element enzyme construct template with a TALE DNA binder

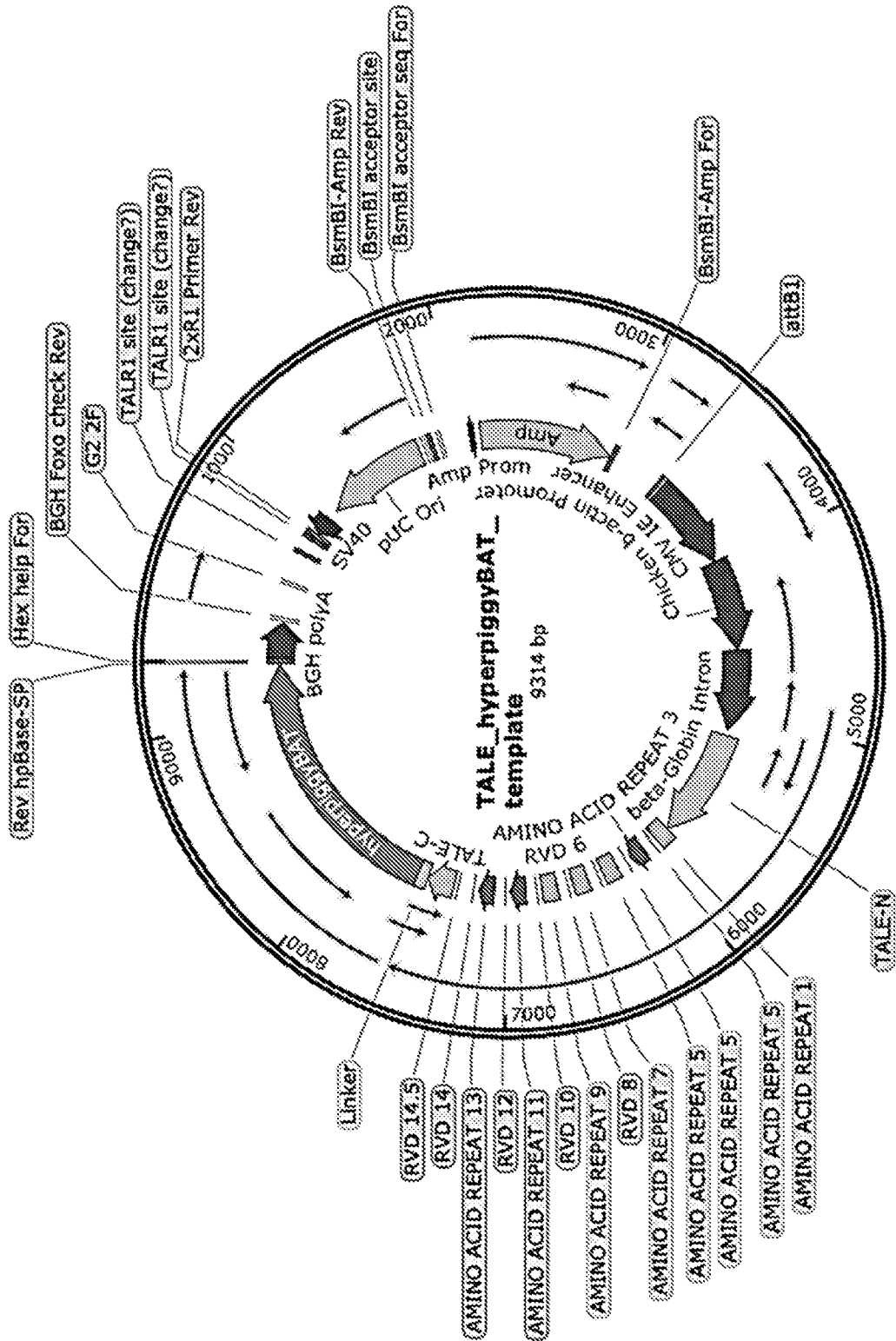


FIG. 3E. A mammalian mobile element enzyme system comprising donor DNA and helper RNA

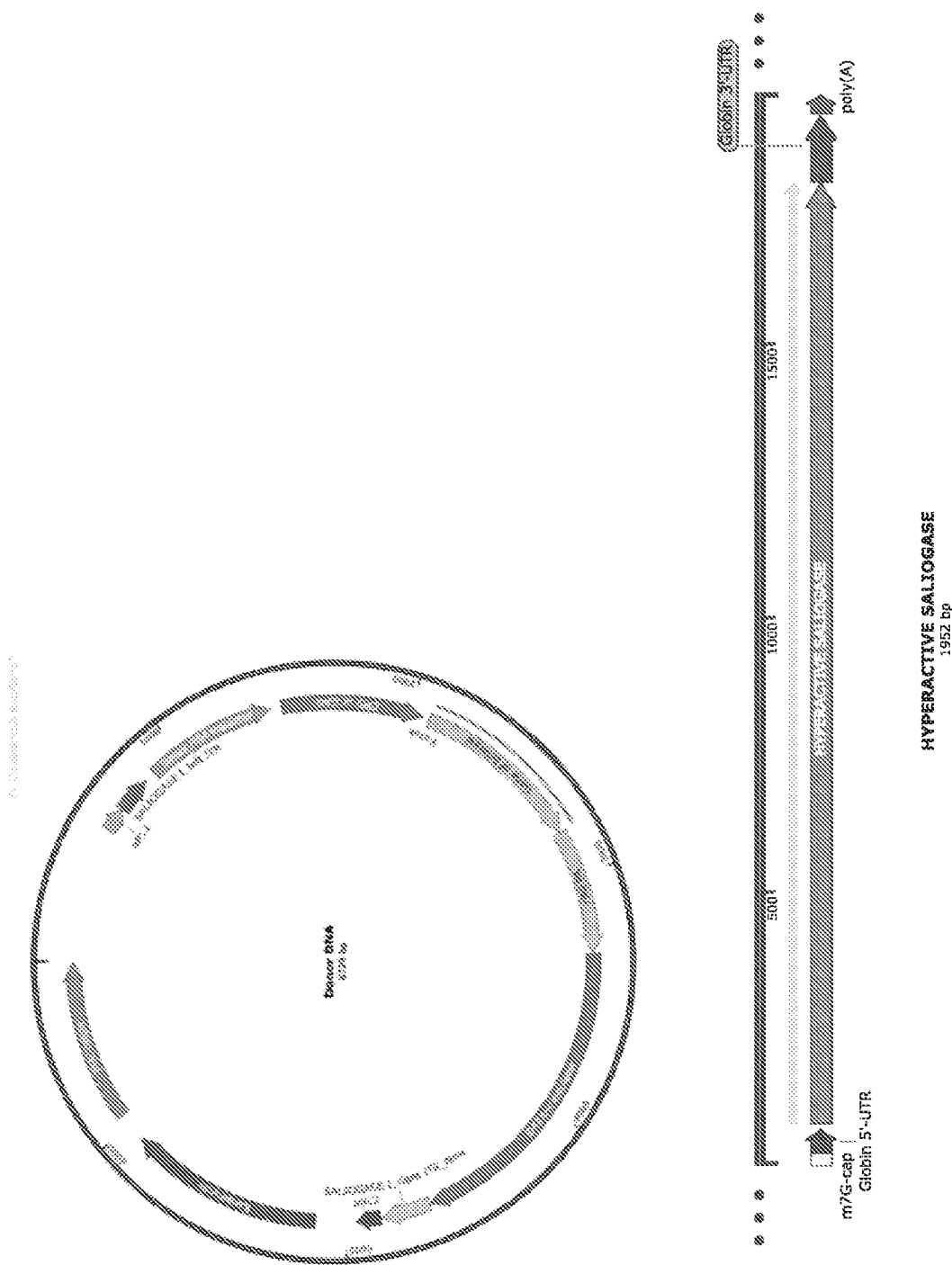


FIG. 4

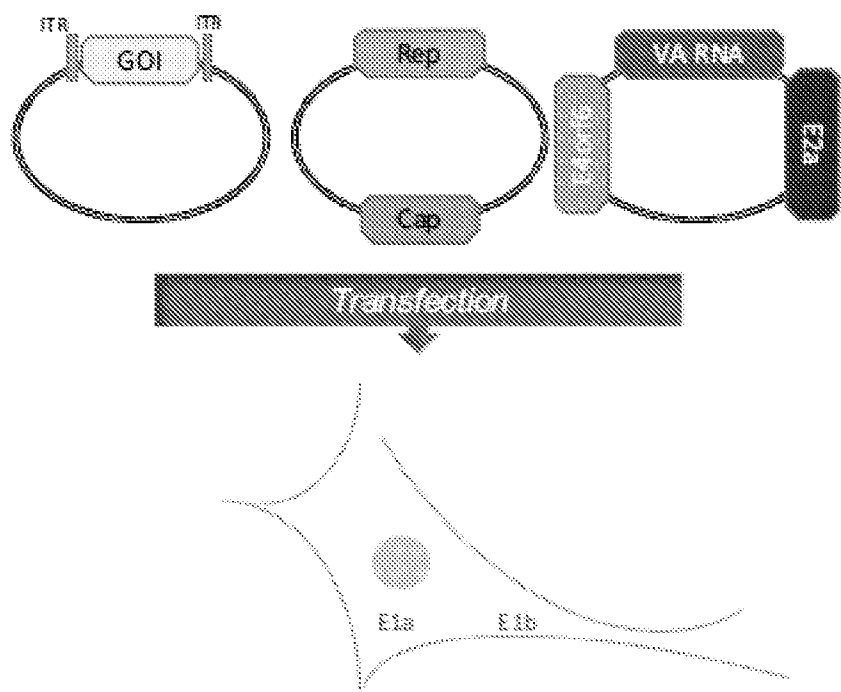


FIG. 5

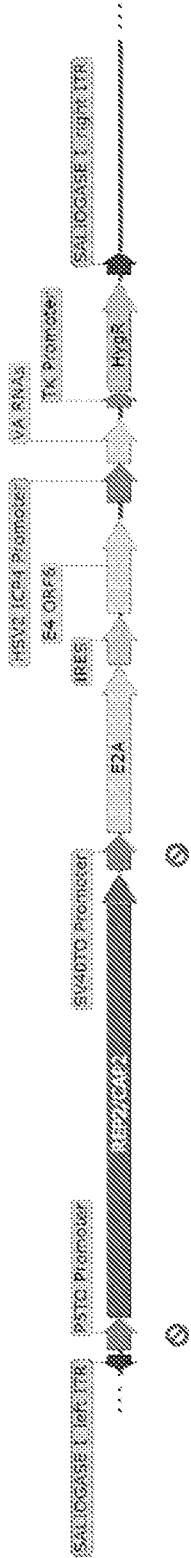


FIG. 6

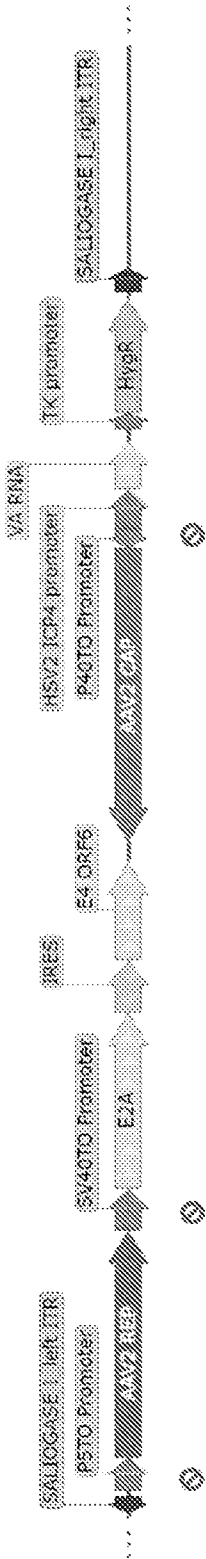


FIG. 7

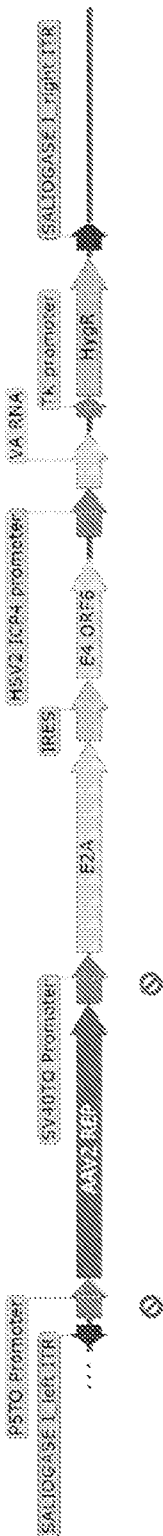


FIG. 8

TF 2: pSFA2-EP40CAP2WZ1-SB

7816 bp

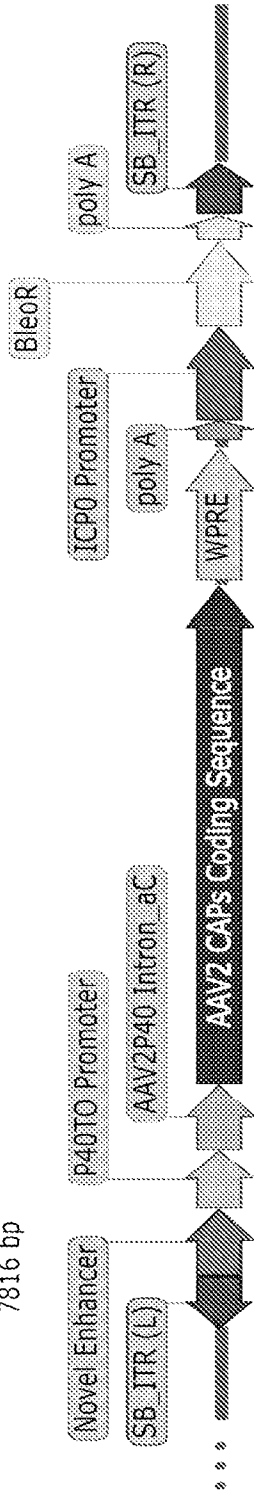


FIG. 9

TF 3: pSFA2-EP40CAP2WZ2-SB

7816 bp

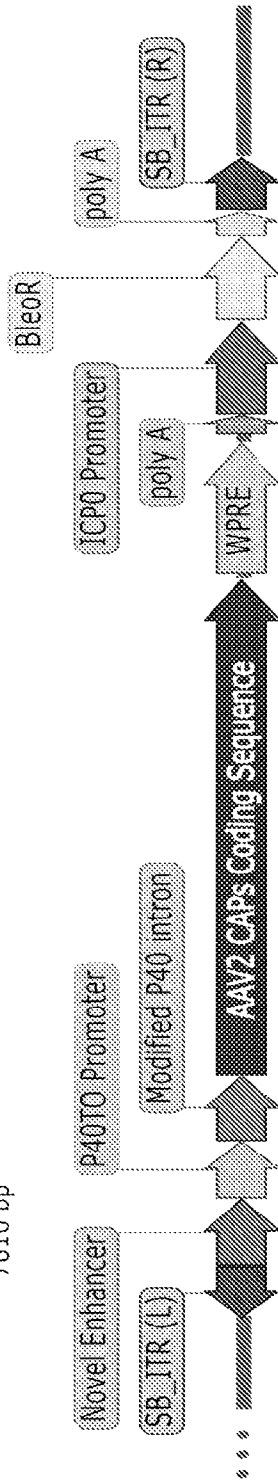
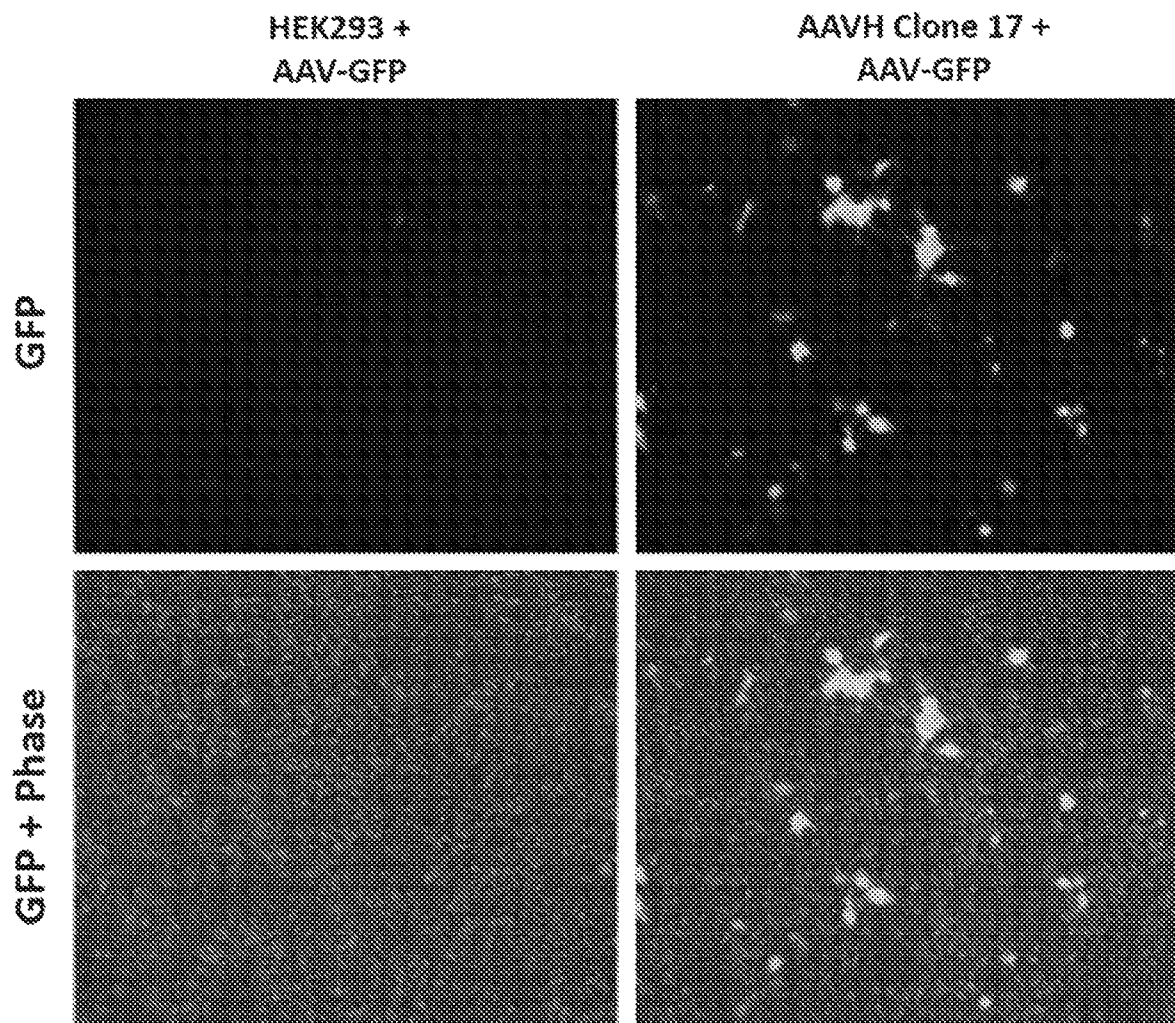


FIG. 10A



GFP: 150 ms

FIG. 10B

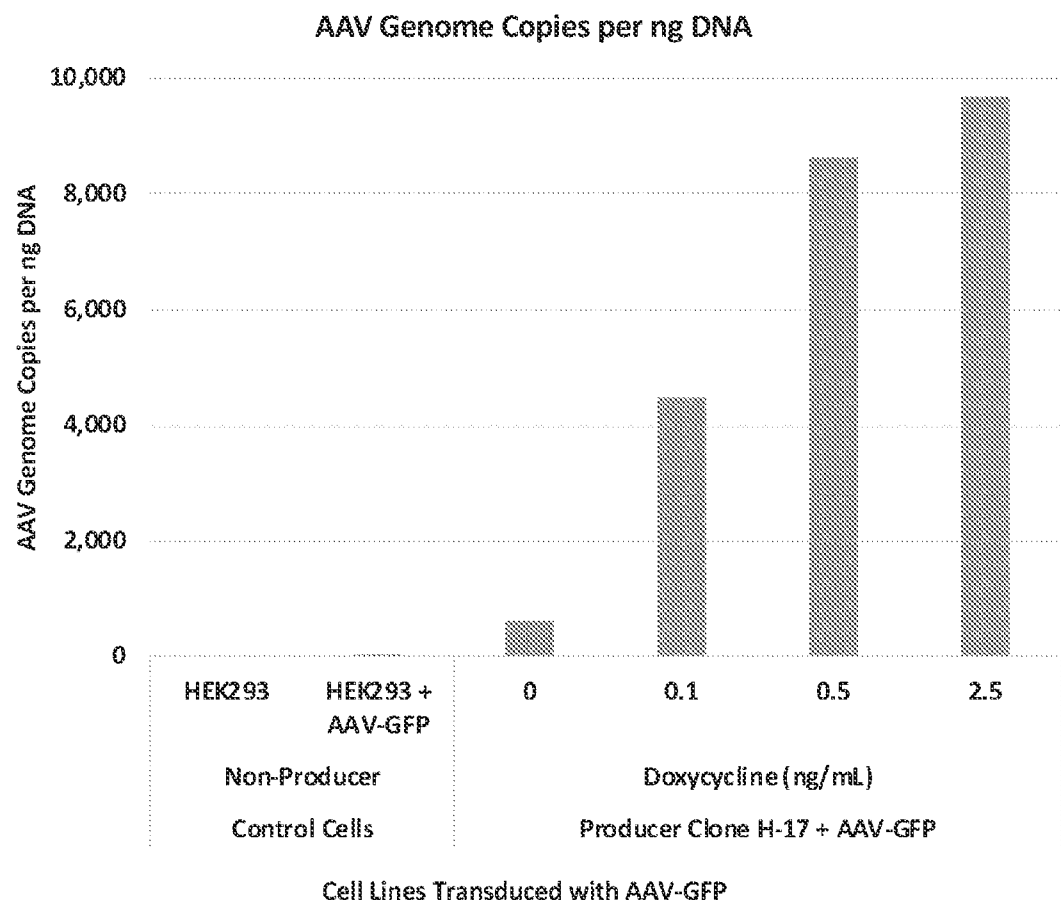


FIG. 11A

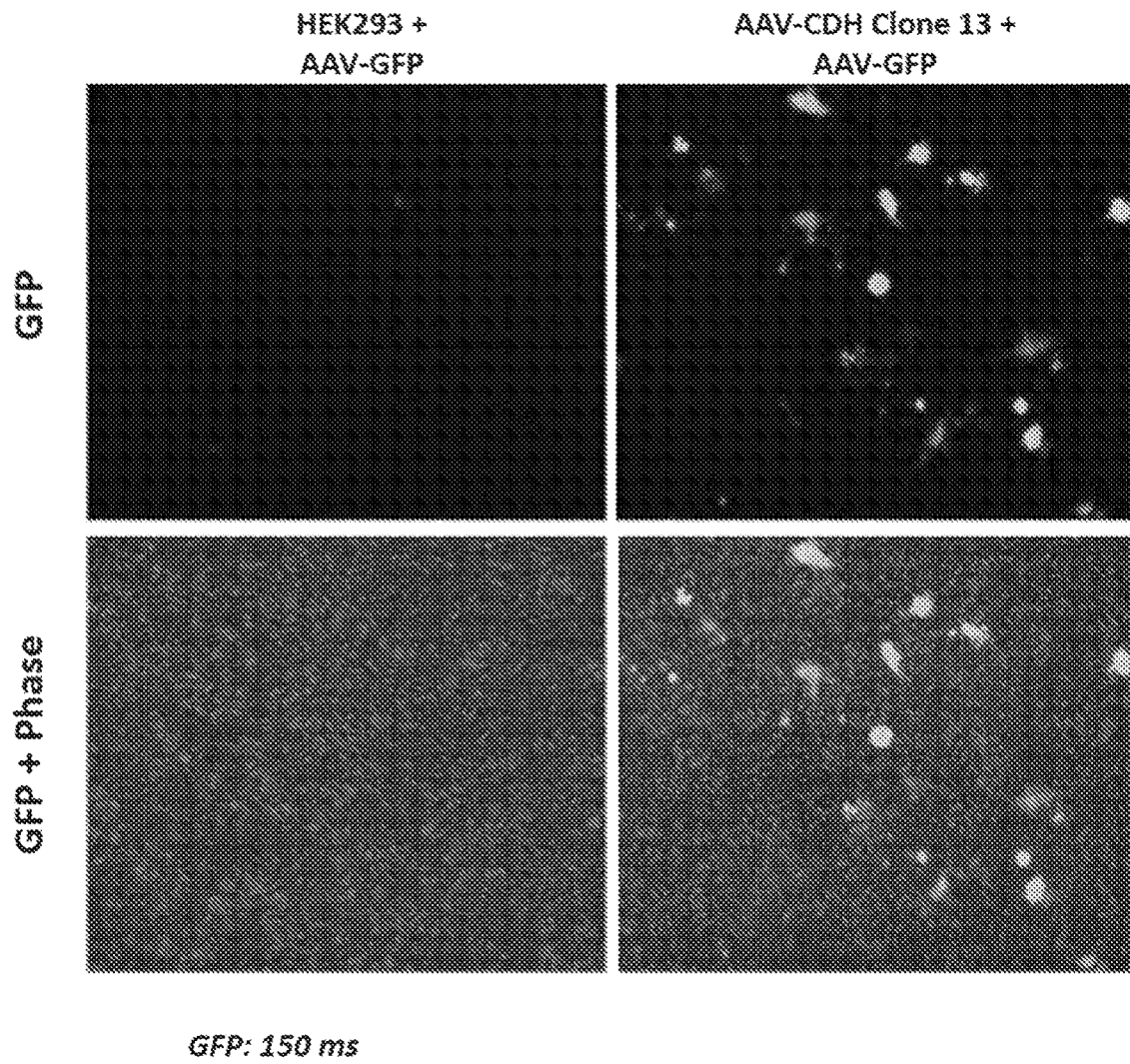


FIG. 11B

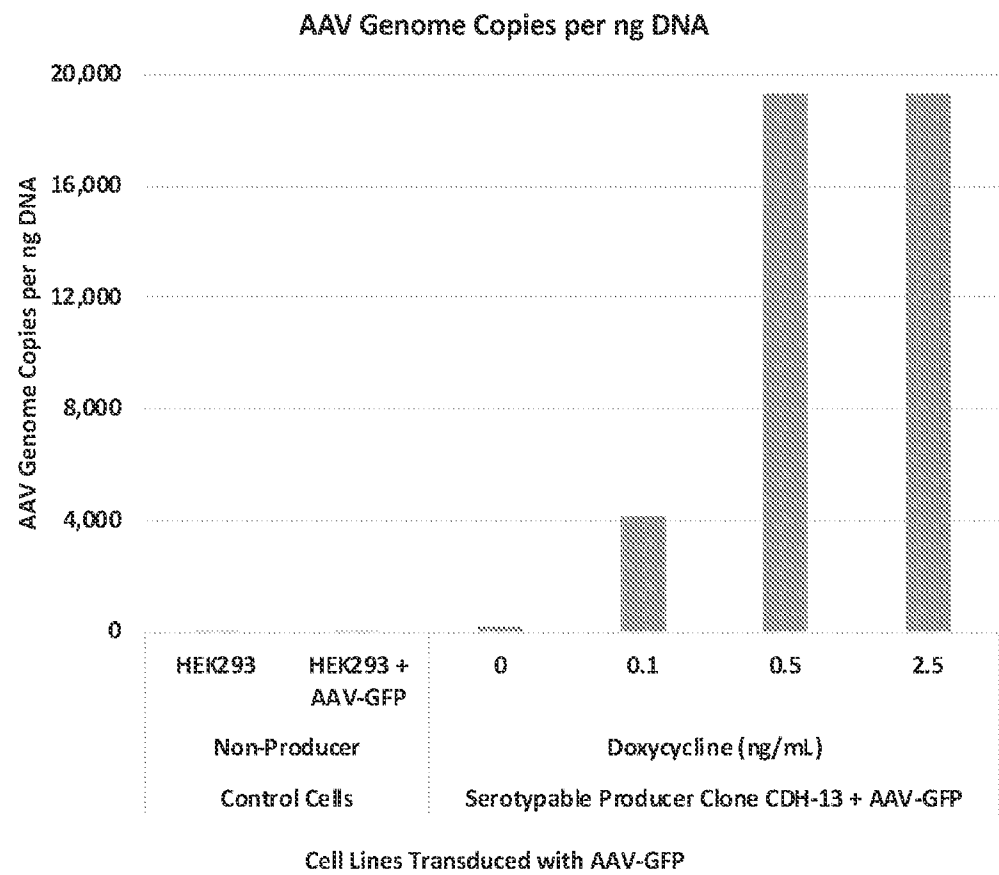


FIG. 12A

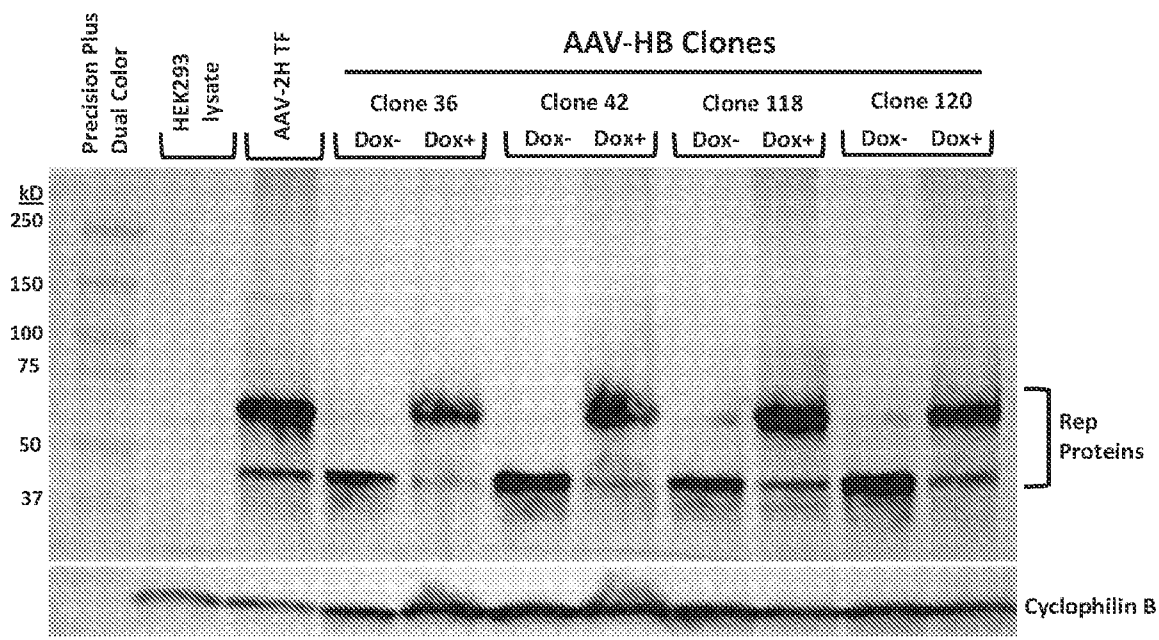


FIG. 12B

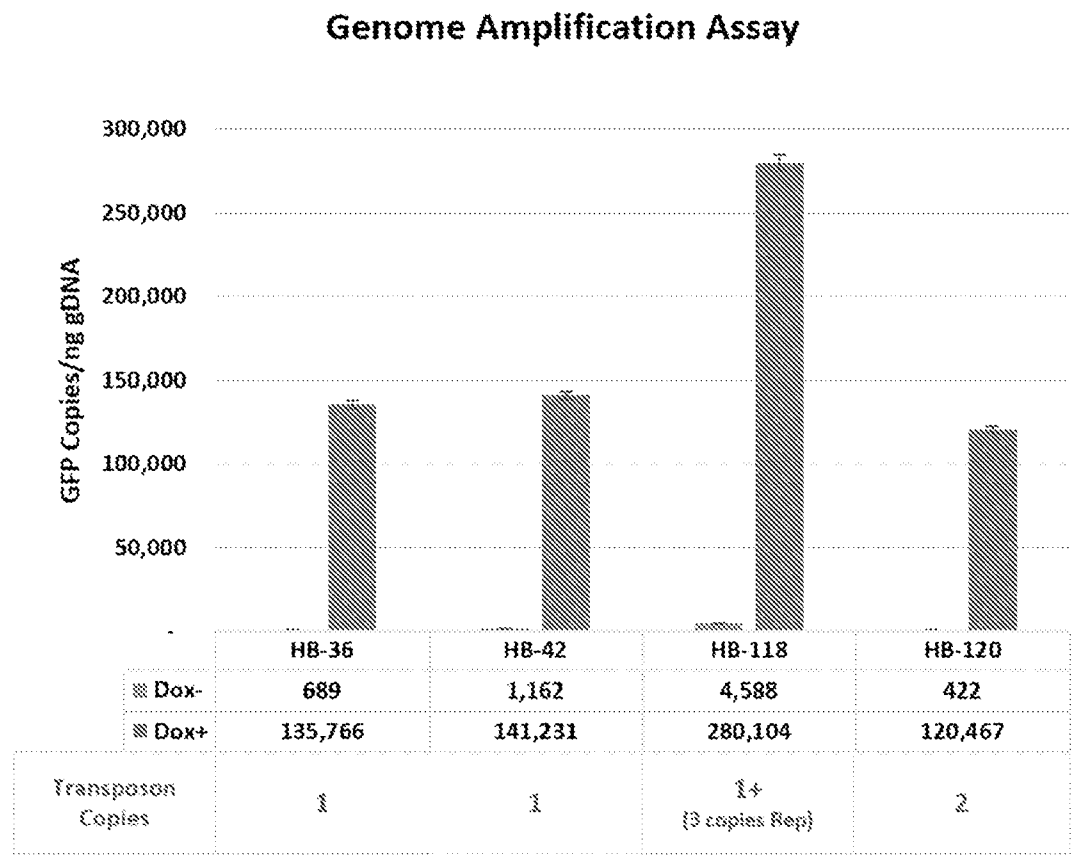
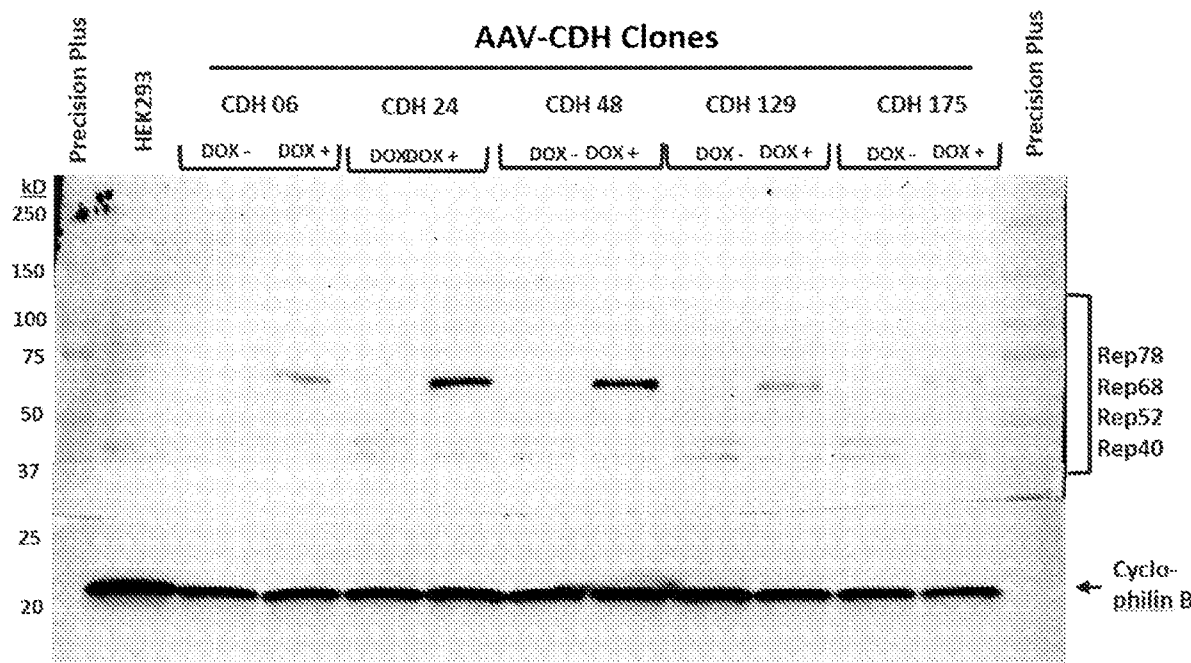


FIG. 13A



AAV-CDH clones 48 and 129 identified as
top AAV Packaging Cell Lines

FIG. 13B

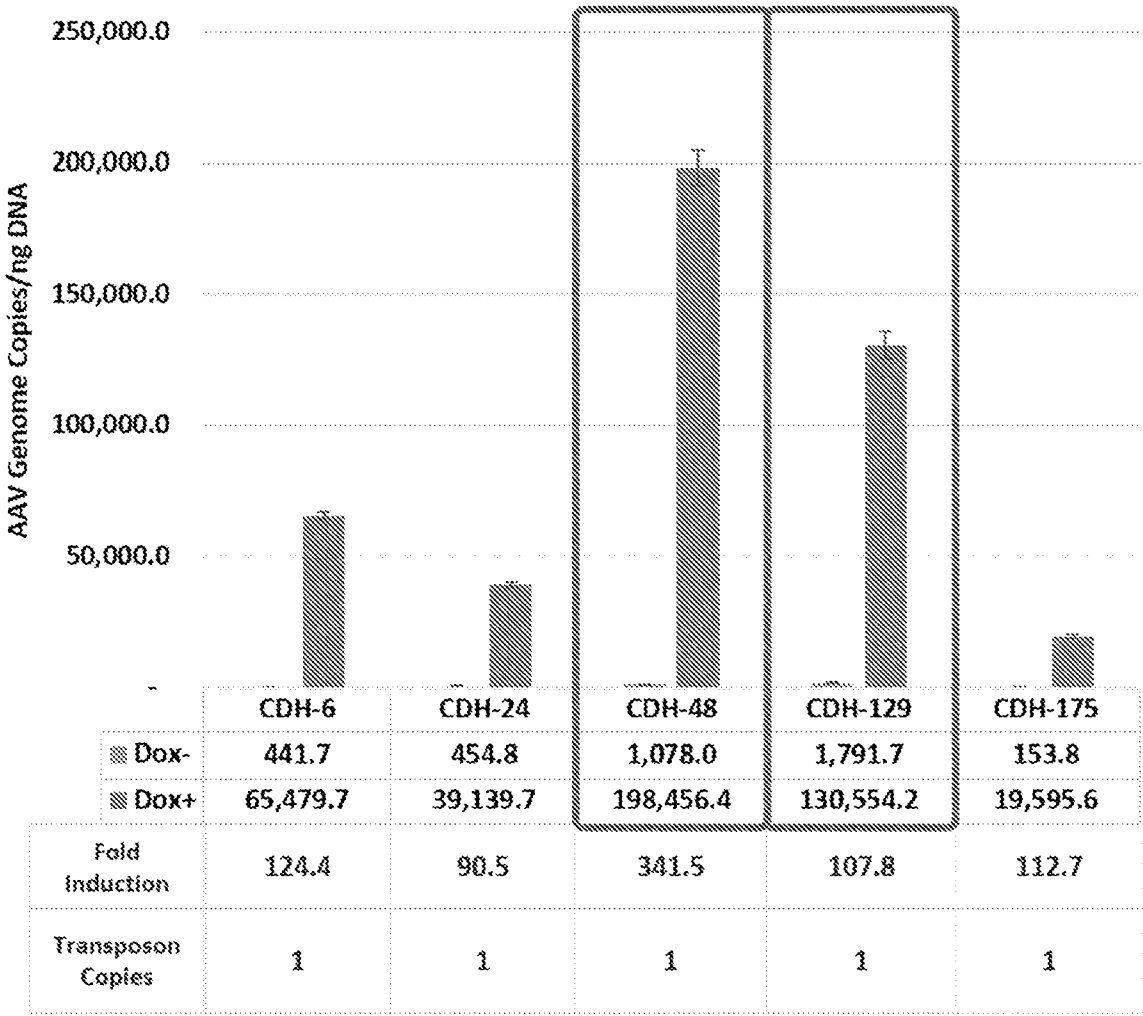


FIG. 14A

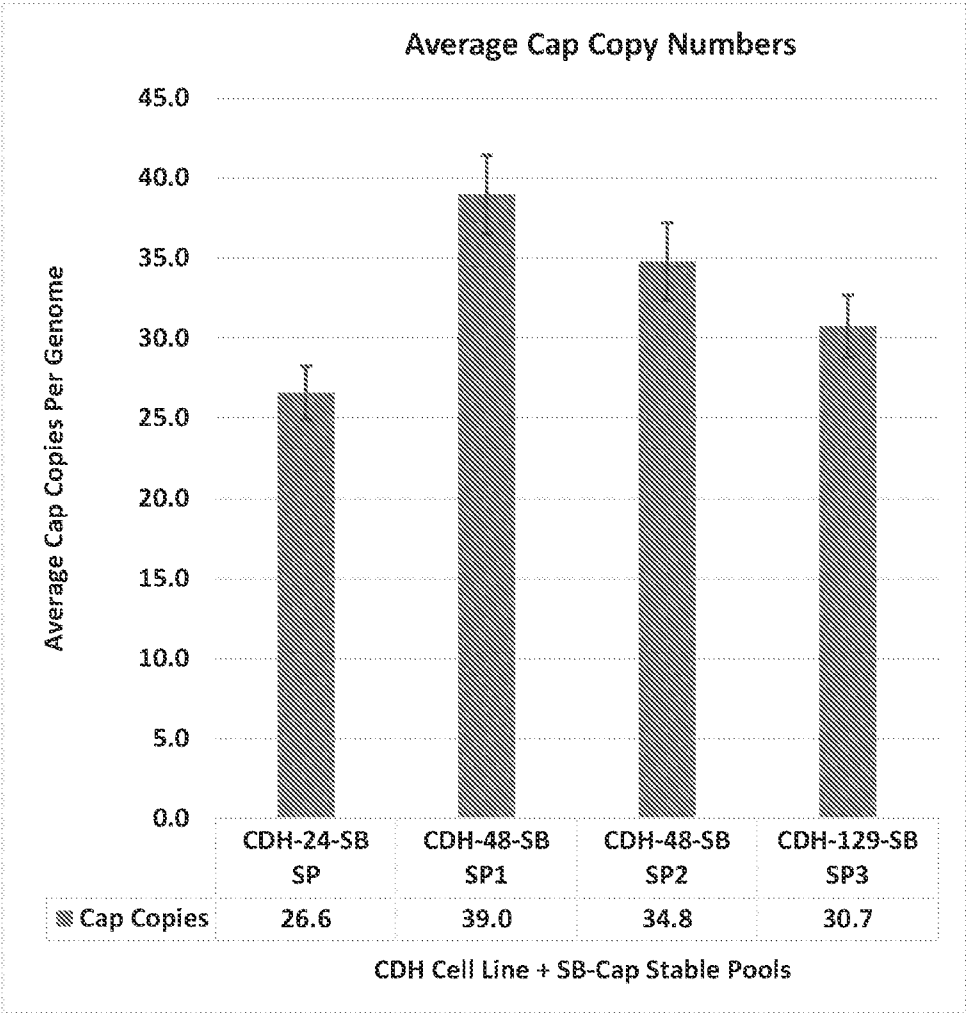


FIG. 14B

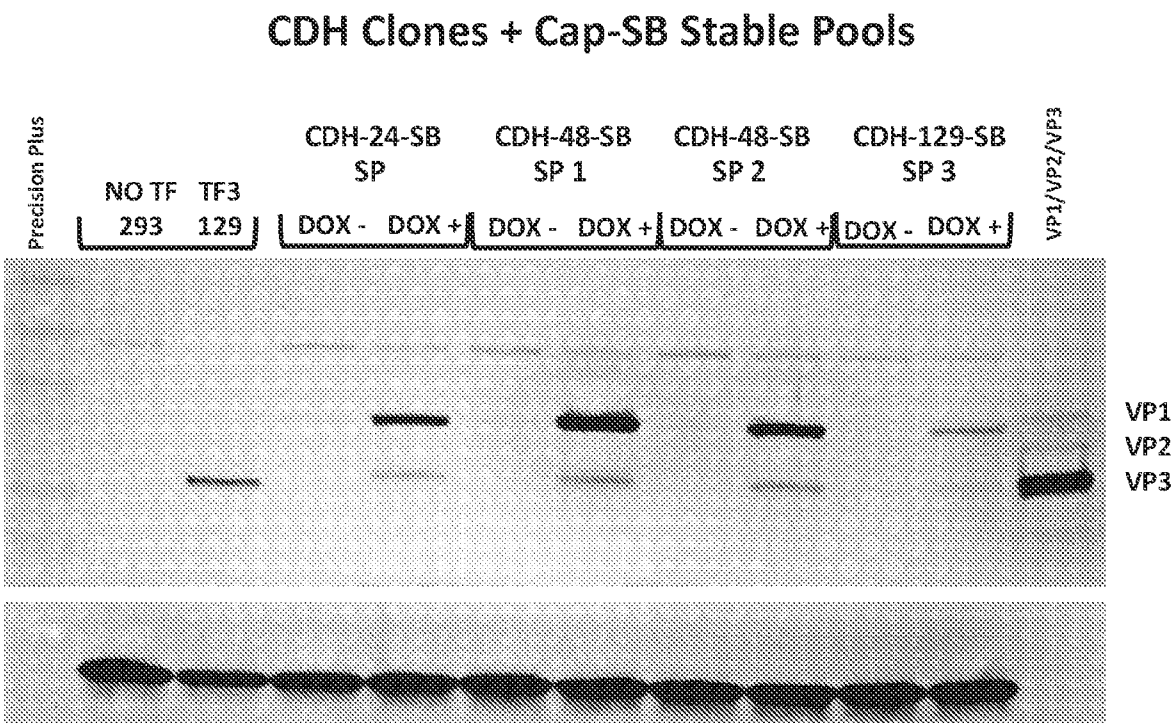


FIG. 15A

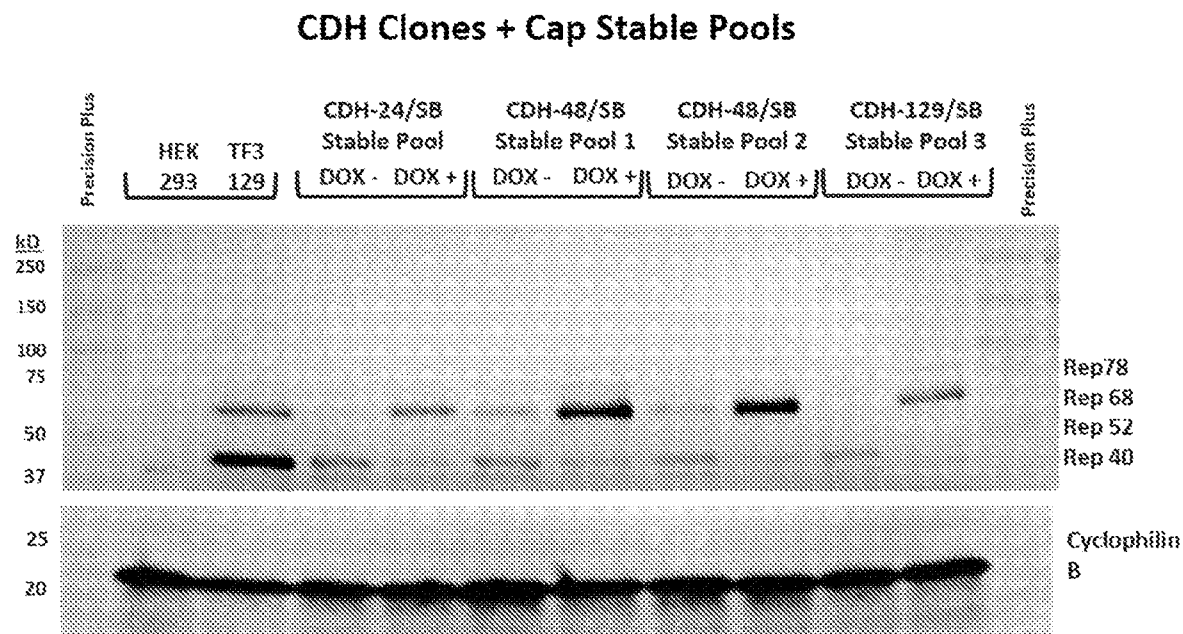


FIG. 15B

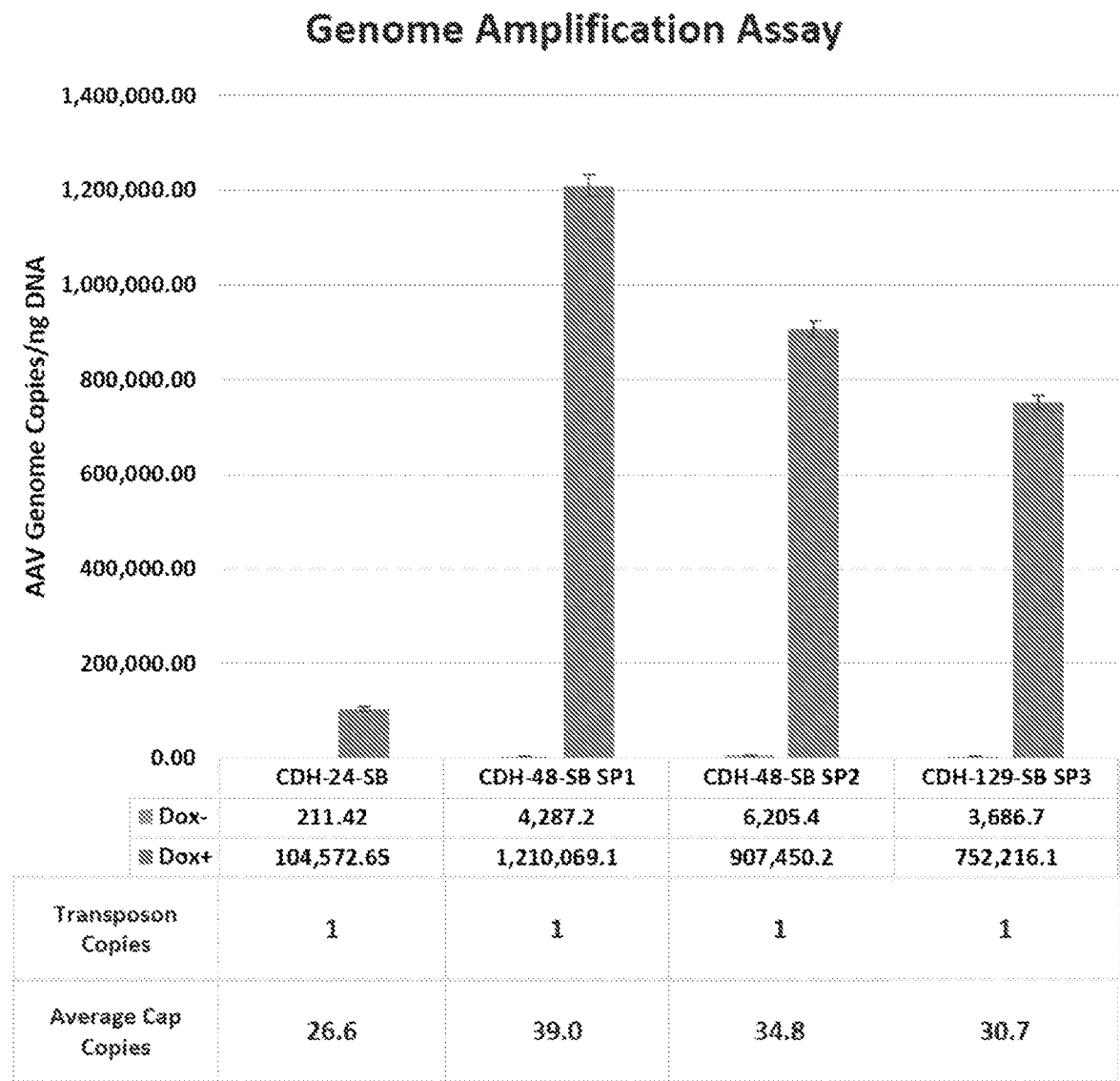


FIG. 16

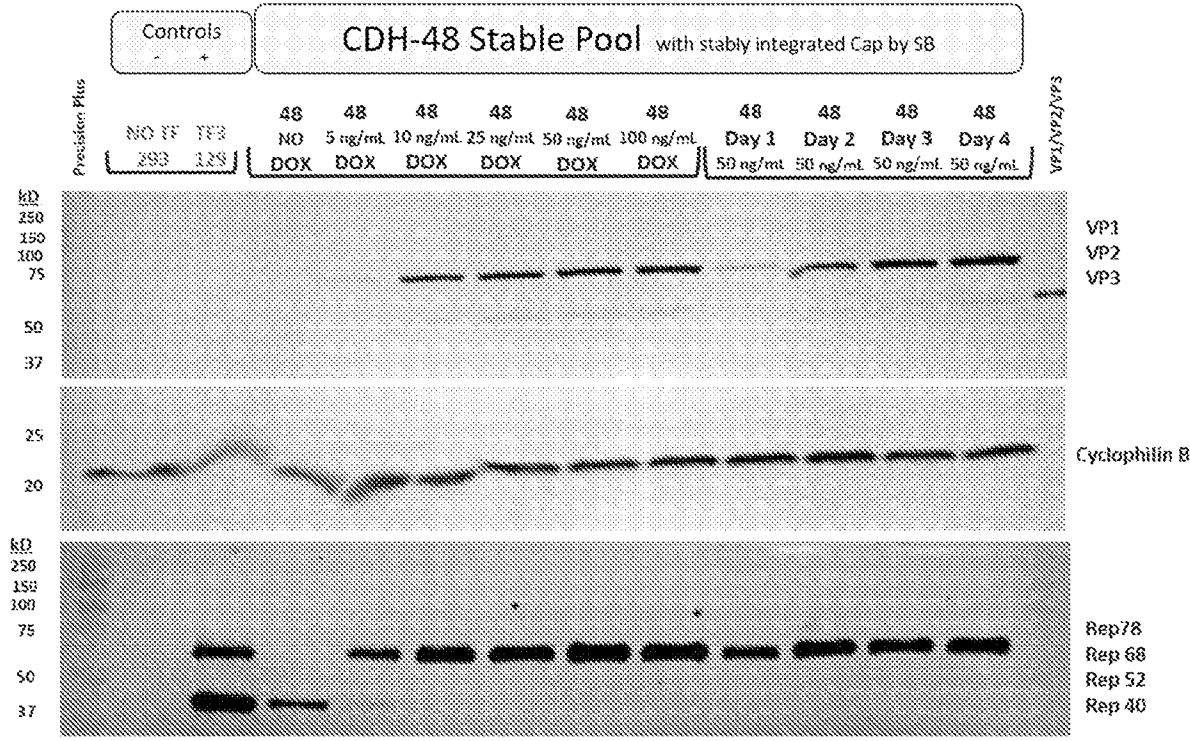


FIG. 17A

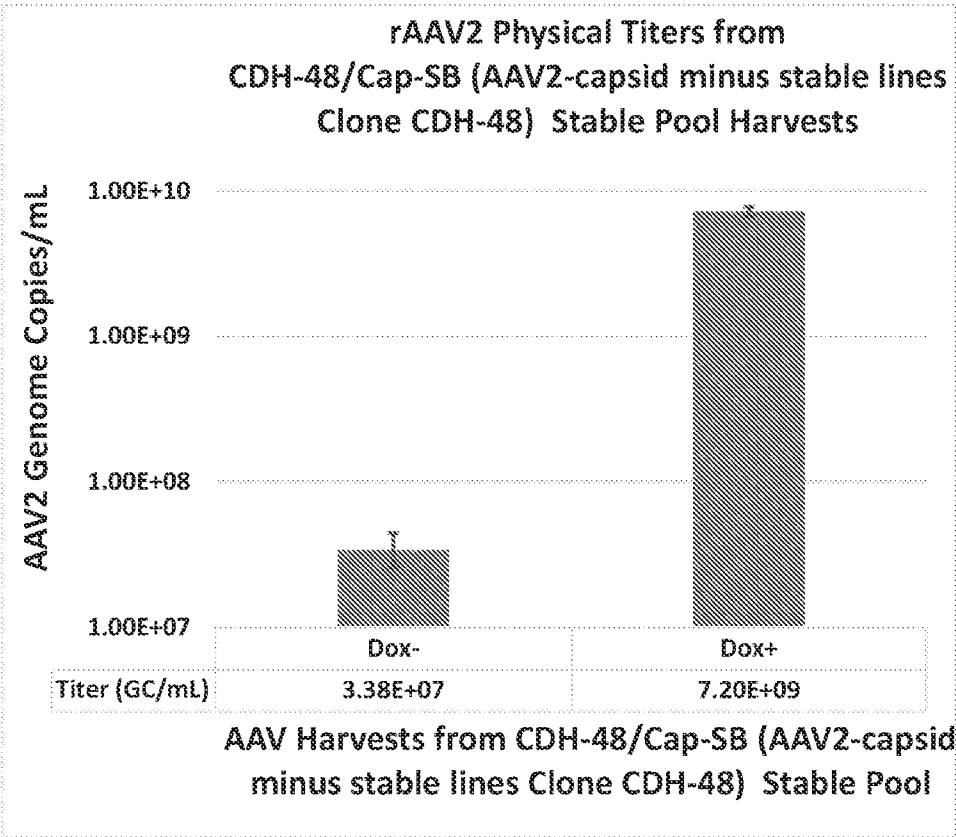


FIG. 17B

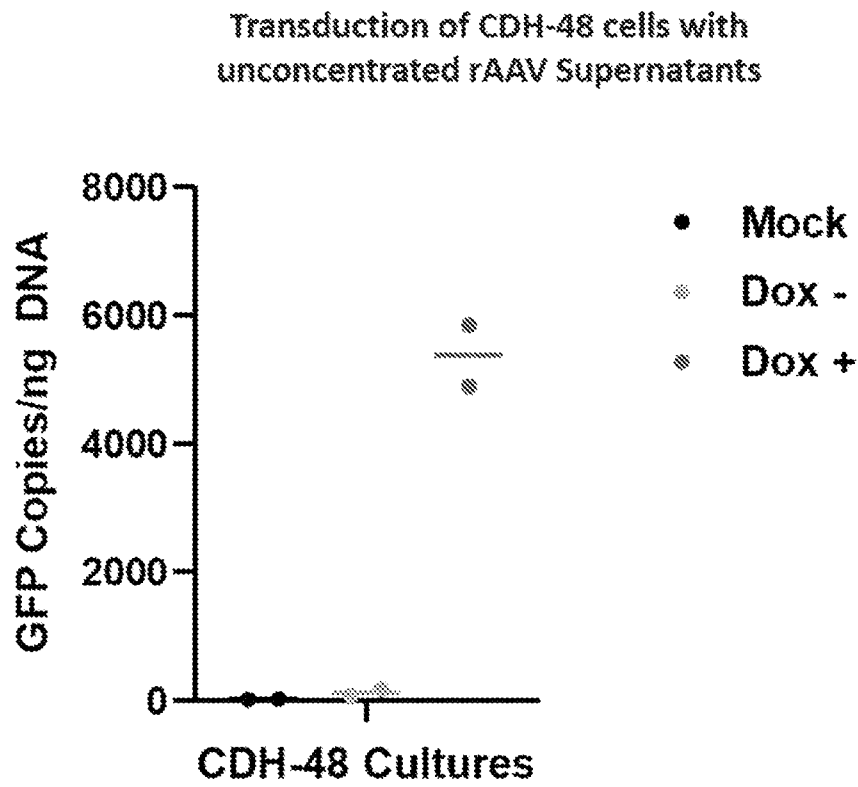


FIG. 18A

CDH-48/SB Clone Cap Copy Numbers

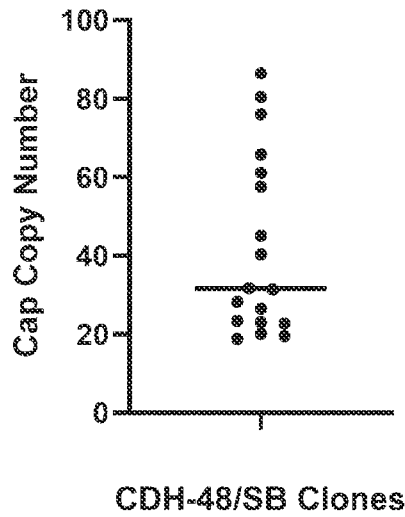


FIG. 18B

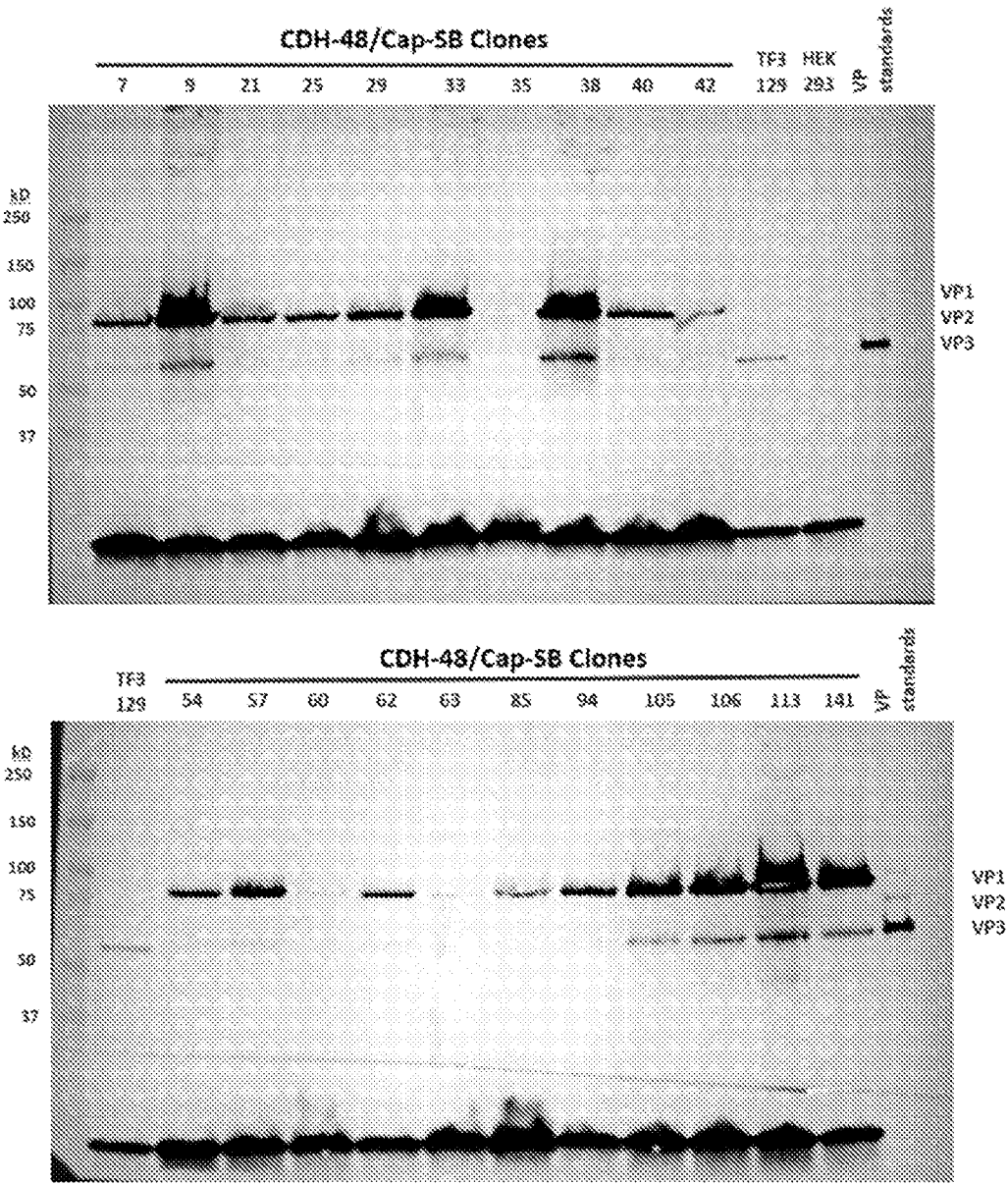


FIG. 19A

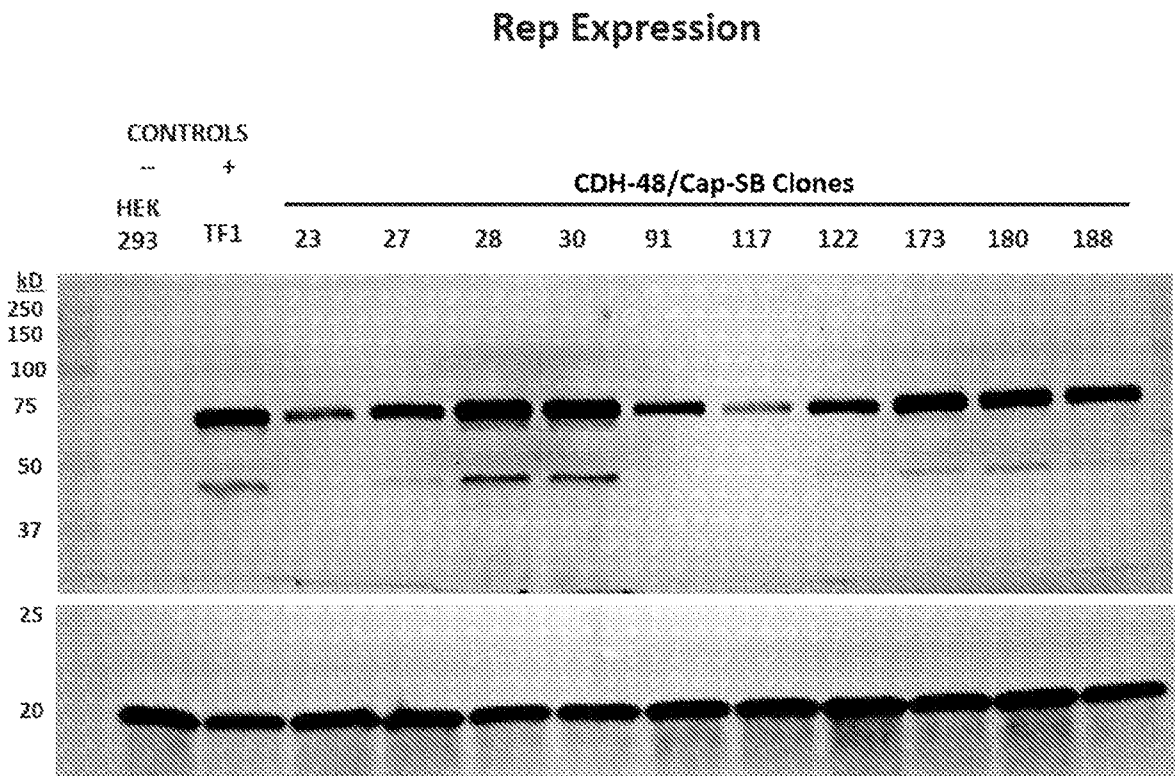


FIG. 19B

