(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 May 2021 (27.05.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/102042 A1

(51) International Patent Classification:
C12N 15/11 (2006.01) *C07K 19/00* (2006.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2020/061114

(22) International Filing Date:
18 November 2020 (18.11.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/937,697	19 November 2019 (19.11.2019)	US
62/952,896	23 December 2019 (23.12.2019)	US
63/067,859	19 August 2020 (19.08.2020)	US
63/081,857	22 September 2020 (22.09.2020)	US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **SEGEL, Michael**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **LADHA, Alim**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(74) Agent: **XIN, Xiaoban** et al.; Johnson, Marcou, Isaacs & Nix, LLC, Customer No. 95280, P.O. Box 691, Hoschton, Georgia 30548 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: RETROTRANSPOSONS AND USE THEREOF

(57) **Abstract:** Systems and methods for targeted gene modification, targeted insertion, perturbation of gene transcripts, and nucleic acid editing. Novel nucleic acid targeting systems comprise components of CRISPR systems and non-LTR retrotransposon elements.

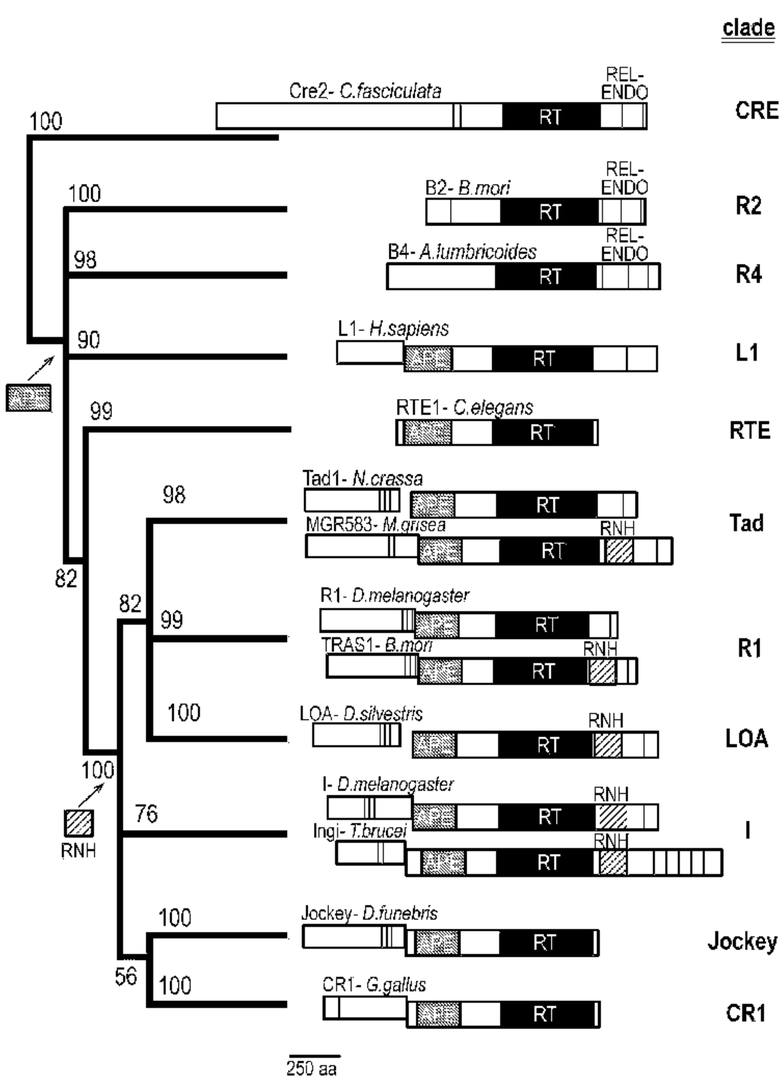


FIG. 1

[Continued on next page]

WO 2021/102042 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

RETROTRANSPOSONS AND USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/937,697, filed November 19, 2019, U.S. Provisional Application No. 62/952,896, filed December 23, 2019, U.S. Provisional Application No. 63/067,859, filed August 19, 2020, and U.S. Provisional Application No. 63/081,857, filed September 22, 2020. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The contents of the electronic sequence listing ("BROD-5025WP_ST25.txt"; Size is 216,022 bytes and it was created on November 18, 2020) is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0003] The subject matter disclosed herein is generally directed to systems, methods and compositions used for targeted gene modification, targeted insertion, perturbation of gene transcripts, nucleic acid editing. Novel nucleic acid targeting systems comprise components of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and retrotransposon elements.

BACKGROUND

[0004] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that employ novel strategies and molecular mechanisms and are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome. This would provide a major resource for new applications in genome engineering and biotechnology.

[0005] Citation or identification of any document in this application is not an admission that such document is available as prior art to the present invention.

SUMMARY

[0006] In one aspect, the present disclosure provides an engineered or non-naturally occurring composition comprising: a. a site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof; b. a non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof; c. a guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide; and d. a polynucleotide encoding a retrotransposon RNA, wherein the retrotransposon RNA comprises or encodes a donor polynucleotide.

[0007] In another aspect, the present disclosure provides an engineered or non-naturally occurring composition comprising: a. two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof; b. two non-LTR retrotransposon polypeptides, each connected to or otherwise capable of forming a complex with one of the two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof; c. two guide molecules, each capable of forming a complex with one of the site-specific nuclease polypeptides and directing site-specific binding to a target sequence of a target polynucleotide; and d. a polynucleotide encoding a retrotransposon RNA comprising or encoding a donor polynucleotide.

[0008] In some embodiments, the retrotransposon RNA capable of forming a complex with the non-LTR retrotransposon polypeptide. In some embodiments, the retrotransposon RNA comprises an binding element capable of binding to the non-LTR retrotransposon polypeptide. In some embodiments, the binding element comprises a hairpin structure. In some embodiments, the donor polynucleotide is for insertion at, or adjacent to, the target sequence. In some embodiments, the site-specific nuclease is a nickase. In some embodiments, the site-specific nuclease lacks nuclease activity. In some embodiments, the non-LTR retrotransposon polypeptide is a dimer, wherein the dimer subunits are connected or form a tandem fusion. In some embodiments, the non-LTR retrotransposon polypeptide comprises a first retrotransposon polypeptide and a second retrotransposon polypeptide, wherein the second retrotransposon polypeptide comprises nuclease or nickase activity. In some embodiments, the site-specific nuclease is connected to the second retrotransposon polypeptide. In some embodiments, the nuclease domain(s) of the retrotransposon polypeptide is inactivated. In

some embodiments, the non-LTR retrotransposon polypeptide is R2. In some embodiments, the R2 is from *Bombyx mori*, *Clonorchis sinensis*, or *Zonotrichia albicollis*. In some embodiments, the non-LTR retrotransposon polypeptide is L1. In some embodiments, the site-specific nuclease polypeptide comprises a nuclear localization signal sequence. In some embodiments, the non-LTR retrotransposon polypeptide comprises a nuclear localization signal. In some embodiments, the polynucleotide encoding a retrotransposon RNA comprises a poly-A tail. In some embodiments, the polynucleotide encoding a retrotransposon RNA.

[0009] In another aspect, the present disclosure provides an engineered composition for non-native, targeted transposition of donor sequence into targeted nucleic acids; a fusion protein comprising a site-specific nuclease fused to the N-terminus of a non-LTR retrotransposon polypeptide, or a polynucleotide comprising a coding sequence thereof; and a donor construct comprising a donor polynucleotide sequence located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide.

[0010] In some embodiments, the donor polynucleotide further comprises a polymerase processing element to facilitate 3' end processing of the donor polynucleotide sequence. In some embodiments, the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both. In some embodiments, the homology region is between 8 and 25 base pairs. In some embodiments, the homology region is on the 3' end of the donor polynucleotide only. In some embodiments, the donor polynucleotide sequence is between 5bp and 50 kb in length. In some embodiments, the non-LTR retrotransposon polypeptide is a wild-type non-LTR retrotransposon polypeptide. In some embodiments, the non-LTR retrotransposon polypeptide comprises one or more modification or truncations. In some embodiments, the one or more mutations or one or more truncations are in an endonuclease domain or reverse transcriptase domain. In some embodiments, the one or more modifications are truncations are in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain. In some embodiments, the fusion protein comprises a nuclear localization signal. In some embodiments, the site-specific nuclease is a Cas protein. In some embodiments, the composition further comprises a guide molecule capable of forming a CRISPR-Cas complex with the Cas polypeptide and directing site-specific binding to a target sequence of a target polynucleotide. In some embodiments, the guide directs the fusion protein to a target sequence 5' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site. In some embodiments, the guide directs the

fusion protein to a target sequence 3' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

[0011] In some embodiments, the Cas polypeptide is a Class 2, Type II Cas or a Type V Cas. In some embodiments, the Cas polypeptide is a Class 2, Type II Cas. In some embodiments, the Type II Cas is a Cas9. In some embodiments, the Cas9 has an HNH domain that is inactivated. In some embodiments, the Cas9 is Cas9D10A or Cas9H840A. In some embodiments, the Cas polypeptide is a Class 2, Type V Cas. In some embodiments, the Type V Cas is Cas12a or Cas12b. In some embodiments, the site specific nuclease is a IscB or a TnpB. In some embodiments, the polynucleotide encoding a retrotransposon RNA comprises a pol2 promoter, a pol3 promoter, or a T7 promoter. In some embodiments, a 3' end of the retrotransposon RNA is complementary to the target sequence, specifically to a portion of a nicked target sequence. In some embodiments, the composition further comprises an RNaseH. In some embodiments, the two site-specific nuclease polypeptides bind to two target sites on the target polynucleotide, and the donor polynucleotide is inserted to a position between the two target sites. In some embodiments, the retrotransposon RNA comprises a region capable of hybridizing with an overhang of the target polynucleotide. In some embodiments, the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA. In some embodiments, the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA. In some embodiments, the mRNA comprises a poly-A tail. In some embodiments, the polynucleotide comprising the coding sequence of non-LTR retrotransposon polypeptide is an mRNA. In some embodiments, the mRNA comprises a poly-A tail. In some embodiments, the donor polynucleotide comprises a homology sequence of the target sequence.

[0012] In some embodiments, the homology sequence is of a region on a strand of the target sequence that contains a PAM of the site-specific nuclease polypeptide. In some embodiments, the region comprises the PAM sequence. In some embodiments, the region is at 3' side of a cleavage site of the site-specific nuclease polypeptide. In some embodiments, the homology sequence comprises from 1 to 30, from 4 to 10, or from 10 to 25 nucleotides in length. In some embodiments, the homology sequence is of a region on a strand that binds to the guide. In some embodiments, the region comprises at least a portion of a guide-binding sequence. In some embodiments, the region comprises a sequence at 3' side of the guide-binding sequence. In some embodiments, the guide molecule forms a RNA-DNA duplex with the target sequence, and the region comprises a sequence of 5 to 15 nucleotides from 3' of the RNA-DNA duplex. In some embodiments, the region comprises a sequence of 10 nucleotides

from 3' side of the RNA-DNA duplex. In some embodiments, the donor polynucleotide is an RNA comprising a poly-A tail.

[0013] In another aspect, the present disclosure provides a vector system comprising one or more vectors, the one or more vectors comprising one or more polynucleotides encoding the polypeptides and/or polynucleotides herein, or a combination thereof. In some embodiments, the one or more polynucleotides comprise one or more regulatory elements operably configures to express the polypeptide(s) and/or the nucleic acid component(s), optionally wherein the one or more regulatory elements comprise inducible promoters. In some embodiments, the polynucleotide molecule encoding the Cas polypeptide is codon optimized for expression in a eukaryotic cell.

[0014] In another aspect, the present disclosure provides a cell or progeny thereof transiently or non-transiently transfected with the vector system herein. In another aspect, the present disclosure provides an organism comprising the cell of herein.

[0015] In another aspect, the present disclosure provides a method of inserting a donor polynucleotide sequence into a target polynucleotide comprising: introducing the engineered or non-naturally occurring composition herein to a cell or population of cells, wherein the complex of the site-specific nuclease polypeptide and the guide directs the non-LTR retrotransposon polypeptide to the target sequence, and wherein the non-LTR retrotransposon polypeptide inserts the donor polynucleotide encoded by the retrotransposon RNA at or adjacent to the target sequence.

[0016] In some embodiments, the donor polynucleotide: a. introduces one or more mutations to the target polynucleotide, b. inserts a functional gene or gene fragment at the target polynucleotide, c. corrects or introduces a premature stop codon in the target polynucleotide, d. disrupts or restores a splice cite in the target polynucleotide, e. causes a shift in the open reading frame of the target polynucleotide, or f. a combination thereof.

[0017] In some embodiments, the one or more mutations include substitutions, deletions, and insertions. In some embodiments, the donor polynucleotide is between 100 bases and 30 kb in length. In some embodiments, the polypeptide and/or nucleic acid components are provided via one or more polynucleotide encoding the polypeptides and/or nucleic acid component(s), and wherein the one or more polynucleotides are operably configured to express the polypeptides and/or nucleic acid component(s). In some embodiments, the composition is delivered via liposomes, nanoparticles, exosomes, microvesicles, microinjection, a gene-gun, or one or more viral vectors. In some embodiments, the donor polynucleotide is inserted to a region on the target sequence that is 3' of a PAM-containing strand. In some embodiments, the

donor polynucleotide is inserted to a region on the target sequence that is 3' of a sequence complementary to the guide molecule.

[0018] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0020] **FIG. 1** shows exemplary non-LTR retrotransposons.

[0021] **FIG. 2** shows an exemplary mechanism for insertion of DNA by non-LTR retrotransposons.

[0022] **FIG. 3** shows details on an exemplary mechanism for insertion of DNA by non-LTR retrotransposons.

[0023] **FIG. 4** shows the mechanism of an exemplary system herein.

[0024] **FIG. 5** shows the 5' and 3' RNA for the R2 retrotransposon.

[0025] **FIG. 6** shows an exemplary mechanism for insertion of DNA by L1.

[0026] **FIG. 7** shows exemplary systems comprising two Cas polypeptides and two retrotransposon polypeptides.

[0027] **FIG. 8** shows R2bm transposition is protein dependent, with measured insertion frequency of wild-type and RT-R2.

[0028] **FIG. 9** shows R2bm can move in trans, with measured insertion frequency %.

[0029] **FIG. 10** depicts R2bm-Cas9 fusions, with Cas9 5' and 3' to R2.

[0030] **FIGs. 11A-11B** shows N-term Cas9 (11A) and C-term Cas0 (11B) fusions.

[0031] **FIG. 12** shows approaches to purify R2bm protein, including purification rounds V1: Overnight Induction at 22C, 4L prep, V2: 6hr Induction at 37C, 50mL prep, sonication lysis, 10 min purification, cleavage off beads, V3: Overnight Induction at 16C, 4L prep, microfluidizer lysis, 2hr purification, biotin elution, V4: 6hr Induction at 18C, 50mL prep, lysozyme lysis, 10 min purification, biotin elution.

[0032] **FIG. 13** shows detection of R2bm protein has reverse transcriptase activity.

[0033] **FIG. 14** depicts R2bm insertion into natural plasmid confirmed in HEK cells and NGS confirmation of R2 target integration.

[0034] **FIG. 15** shows gel images depicting correct insertion size, on left, 1. R2bm mRNA + Target Plasmid lysate, 2. R2bm mRNA lysate + Target lysate, and 3. R2bm mRNA lysate; on right, 1. R2bm DNA + Target Plasmid lysate, 2. R2bm ORF (no 3' UTR) + Target lysate, and 3. R2bm DNA.

[0035] **FIG. 16** shows R2bm protein shows *in vitro* transposition.

[0036] **FIG. 17** shows R2bm TPRT can be reprogramed with a nick; other targets tested with no activity: no cut, double nick, completely cut.

[0037] **FIG. 18** shows R2bm can resolve 5' end in a homology dependent manner.

[0038] **FIG. 19** shows gel exploring R2 substrate, where nicked target site with correct insertion size shown, indicative of R2 preference for a nicked target site, in line with proposed transposition mechanism.

[0039] **FIGs. 20A-20C** show investigation of homology dependence. FIG. 20A gel shows 10 bp homology is preferred to longer homology lengths, but reason unclear; FIG. 20B sequencing of product reveals polyA not incorporated into insertion product, 1: 5' homology: None, 3' Homology: 10bp; 2: 5' homology: 25bp, 3' Homology: 10bp; 3: 5' homology: None, 3' Homology: 10bp + 40bp polyA; 4: 5' homology: 25bp, 3' Homology: 10bp + 40bp polyA; FIG. 20C sequencing shows that only up to 9bp of the 25bp are incorporated into the transposition product, with most have 0/25bp inserted, indicating R2 must either start reverse transcribing at its 3' end or process its RNA at the 3' end upon complexing.

[0040] **FIG. 21** shows assay investigating whether Cas9 can work with R2.

[0041] **FIG. 22** includes images showing R2bm expression may limit efficiency.

[0042] **FIG. 23** includes graphs of donors with insertion frequency for several fusions; results indicate R2bm only depends on the UTRs, no internal sequence; GFP tagging (increasing protein expression), increases insertion frequency significantly; N-terminal Cas9 produces superior R2 activity.

[0043] **FIG. 24** shows R2tg is functional and 2-fold better than R2bm.

[0044] **FIG. 25** includes an orthogonal readout of retrotransposition.

[0045] **FIGs. 26A-26B** include evaluation of how much 28S sequence is required for homing' chart includes insertion frequency pSR70, pSR65, pMAX GFP (FIG. 26A) and pSR70 – helper (FIG. 26B).

[0046] **FIG. 27** includes results of luciferase assay evaluating mutants for pSR106 and pSR107.

[0047] **FIG. 28** includes results evaluating whether R2tg retrotransposon activity in assay for pSR125 and pSR126.

[0048] **FIG. 29** shows sequencing showing insertions seen with R2bm and WT Cas9 at most target sites, lesser with R2tg. Helpers are Cas9-R2, Cas0-D10A-R2, H840Cas9-R2 and R2bm and R2tg. Donors are URT-luc reporter – UTR-10bp of homology to target site either upstream or downstream, 10 donors (5 Cas9 targets).

[0049] **FIG. 30** shows gel evaluating whether insertions are TPRT dependent.

[0050] **FIG. 31** shows plasmid pcdna-r2bm-orf-n-hspcas9.

[0051] **FIG. 32** shows plasmid pcdna-r2bm-utrs-luciferase-28s-homology.

[0052] **FIG. 33** shows detection of insertion products by amplifying junction between 3' UTR of donor and target site.

[0053] **FIG. 34** shows exemplary mRNA constructs transfected to HEK293 cells.

[0054] **FIG. 35** shows insertion frequency of the donor constructs with various homology sequences, and with or without poly-A tails.

[0055] **FIG. 36** shows exemplary mRNA constructs designed to insert at 3' side of target sequences.

[0056] **FIG. 37** shows insertion of constructs in FIG. 36 at 3' side of target sequences.

[0057] **FIGs. 38A-38F** show sequence validation of insertions shown in FIG. 37.

[0058] **FIG. 39** shows insertion of constructs in FIG. 36 at 5' side of target sequences.

[0059] **FIGs. 40A-40B** show sequence validation of insertions shown in FIG. 39.

[0060] **FIG. 41** shows insertion by R2 orthologs.

[0061] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0062] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX,

published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011)

.

[0063] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0064] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0065] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0066] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0067] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

[0068] The term "exemplary" is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as "exemplary" is not necessarily to be

construed as preferred or advantageous over other aspects or designs. Rather, use of the word exemplary is intended to present concepts in a concrete fashion.

[0069] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

[0070] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0071] The term “about” in relation to a reference numerical value and its grammatical equivalents as used herein can include the numerical value itself and a range of values plus or minus 10% from that numerical value. For example, the amount “about 10” includes 10 and any amounts from 9 to 11. For example, the term “about” in relation to a reference numerical value can also include a range of values plus or minus 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% from that value.

[0072] Whereas the terms "one or more" or "at least one" or "X or more", where X is a number and understand to mean X or increases one by one of X, such as one or more or at least one member(s) or "X or more" of a group of members, is clear per se, by means of further exemplification, the term encompasses inter alia a reference to any one of said members, or to any two or more of said members, such as, e.g., any >3, >4, >5, >6 or >7 etc. of said members, and up to all said members.

[0073] The term “functional variant or functional fragment” means that the amino-acid sequence of the polypeptide may not be strictly limited to the sequence observed in nature, but may contain additional amino-acids. The term “functional fragment” means that the sequence of the polypeptide may include less amino-acid than the original sequence but still enough amino-acids to confer the enzymatic activity of the original sequence of reference. It is well known in the art that a polypeptide can be modified by substitution, insertion, deletion and/or addition of one or more amino-acids while retaining its enzymatic activity. For example, substitutions of one amino-acid at a given position by chemically equivalent amino-acids that do not affect the functional properties of a protein are common.

[0074] A protein or nucleic acid derived from a species means that the protein or nucleic acid has a sequence identical to an endogenous protein or nucleic acid or a portion thereof in the species. The protein or nucleic acid derived from the species may be directly obtained from an organism of the species (e.g., by isolation), or may be produced, e.g., by recombination production or chemical synthesis.

[0075] All references cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings of all references herein specifically referred to are incorporated by reference.

OVERVIEW

[0076] The present disclosure provides engineered or non-naturally compositions, systems and methods for modify a target polynucleotide, for example, by inserting a donor polynucleotide at desired position in the target polynucleotide. In general, the systems comprise one or more components of a CRISPR-Cas system and one or more components of a retrotransposon. The components of the CRISPR-Cas system may direct the retrotransposon to a target nucleic acid sequence. In one aspect, the present disclosure provides a system comprising a Cas polypeptide, a non-Long Terminal Repeat (LTR) retrotransposon associated with the Cas polypeptide, a guide molecule, and a retrotransposon RNA, and/or polynucleotide(s) encoding these components. In some examples, the Cas polypeptide is a nickase or lacks nuclease activity. In some examples, the non-LTR retrotransposon may be R2 or L1. The present disclosure also includes polynucleotides encoding components of the nucleic acid targeting systems, and vector systems comprising one or more polynucleotide of or encoding components of the CRISPR-Cas system and/or the retrotransposons. Further provided herein also includes cells, tissues, organs, and organisms comprising the systems or generated using the systems.

SYSTEMS AND COMPOSITIONS

[0077] In one aspect, the present disclosure includes systems that comprise one or more components of a retrotransposon and one or more components of a site-specific nuclease. In some embodiments, the retrotransposon may be a non-LTR retrotransposon. For example, the present disclosure provides an engineered or non-naturally occurring composition comprising; a site-specific nuclease; a non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease, and a donor construct comprising a donor polynucleotide sequence located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide. The site-specific nuclease may be programmed to guide the non-LTR retrotransposon-donor construct complex to a targeted insertion site in a target polynucleotide, such as double-stranded DNA. The site-specific nuclease may either create a double-strand break or a single-strand nick at the target site. The non-LTR retrotransposon polypeptide may then facilitate target-primed reverse transcription of the donor polynucleotide sequence and insertion of the donor polynucleotide sequence into the target polynucleotide.

[0078] The term "fusion protein" is used herein to refer to protein construction comprising the site-specific nuclease connected to the non-LTR retrotransposon polypeptide for example by a polypeptide linker or other suitable linker. It should be understood that the term "fusion protein" includes embodiments where the composition comprises a site-specific nuclease and a non-LTR retrotransposon already connected to one another, or embodiments where the site-specific and non-LTR retrotransposon comprise two separate components that may come together to form a single complex, for example, through the use of engineered domains on each polypeptide that functions as binding partners to bring the site-specific and non-LTR retrotransposon together.

[0079] In certain example embodiments, the site specific nuclease may comprise a paired nickase in which each site-specific nuclease in the pair is fused with a non-LTR retrotransposon protein and creates a nick on opposing strands of a targeted insertion site and whereby the corresponding non-LTR retrotransposons facilitate insertion of the donor polynucleotide from the donor construct.

[0080] In certain example embodiments, the site-specific nuclease is a Cas polypeptide and the composition further comprises a guide molecule capable of forming a complex with the Cas polypeptide and directing the Cas polypeptide-non-LTR retrotransposon polypeptide to a target site adjacent to the targeted insertion site.

[0081] In some examples, the guide directs the polypeptides (e.g., a complex or fusion protein of the Cas and non-LTR retrotransposon polypeptide) to a target sequence 5' or 3' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

Retrotransposons

[0082] The systems and compositions herein may comprise one or more components of a retrotransposon, e.g., a non-LTR retrotransposon. Native or wild-type non-LTR retrotransposons encode the protein machinery necessary for their self-mobilization. The non-LTR retrotransposon element comprises a DNA element integrated into a host genome. This DNA element may encode one or two open reading frames (ORFs). For example, the R2 element of *Bombyx mori* encodes a single ORF containing reverse transcriptase (RT) activity and a restriction enzyme-like (REL) domain. L1 elements encode two ORFs, ORF1 and ORF2. ORF1 contains a leucine zipper domain involved in protein-protein interactions and a C-terminal nucleic acid binding domain. ORF2 has a N-terminal apurinic/apyrimidinic endonuclease (APE), a central RT domain, and a C-terminal cysteine histidine rich domain. An example replicative cycle of a non-LTR retrotransposon may comprise transcription of the full-

length retrotransposon element to generate an mRNA active element (retrotransposon RNA). The active element mRNA is translated to generate the encoded retrotransposon proteins or polypeptides. A ribonucleoprotein complex comprising the active element and retrotransposon protein or polypeptide is formed and this RNP facilitates integration of the active element into the genome. FIG. 2 shows an exemplary mechanism for insertion of DNA by non-LTR retrotransposons. The RNA-transposase complex nicks the genome. The 3' end of the nicked DNA serves as a primer to allow the reverse transcription of the transposon RNA into cDNA. Fourth, the transposase proteins integrate the cDNA into the genome. More details on the insertion mechanism is shown in FIG. 3.

[0083] Elements of these systems may be engineered to work within the context of the invention. For example a non-LTR retrotransposon polypeptide may be fused to a site-specific nuclease. The binding elements that allow a non-LTR retrotransposon polypeptide to bind to the native retrotransposon DNA element, may be engineered into a donor construct to facilitate entry of a donor polynucleotide sequence into a target polypeptide.

[0084] In the present invention the protein component of the non-LTR retrotransposon may be connected to or otherwise engineered to form a complex with a site-specific nuclease. The retrotransposon RNA may be engineered to encode a donor polynucleotide sequence. Thus, in certain example embodiments, the Cas polypeptide, via formation of a CRISPR-Cas complex with a guide sequence, directs the retrotransposon complex (i.e. the retrotransposon polypeptide(s) and retrotransposon RNA to a target sequence in a target polynucleotide, where the retrotransposon RNP complex facilitates integration of the donor polynucleotide sequence into the target polynucleotide. Accordingly, the one or more non-LTR retrotransposon components may comprise retrotransposon polypeptides, or function domains thereof, that facilitate binding of the retrotransposon RNA, reverse transcription of the retrotransposon RNA into cDNA, and/or integration of the donor polynucleotide into the target polynucleotide, as well as retrotransposon RNA elements modified to encode the donor polynucleotide sequence.

[0085] Examples non-LTR retrotransposons include CRE, R2, R4, L1, RTE, Tad, R1, LOA, I, Jockey, CR1 (see FIG. 1). In one example, the non-LTR retrotransposon is R2. In another example, the non-LTR retrotransposon is L1. Examples of non-LTR retrotransposons may include those described in Christensen SM et al., RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site, Proc Natl Acad Sci U S A. 2006 Nov 21;103(47):17602-7; Eickbush TH et al, Integration, Regulation, and Long-Term Stability of R2 Retrotransposons, Microbiol Spectr. 2015 Apr;3(2):MDNA3-0011-2014. doi: 10.1128/microbiolspec.MDNA3-0011-2014; Han JS, Non-long terminal

repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions, Mob DNA. 2010 May 12;1(1):15. doi: 10.1186/1759-8753-1-15; Malik HS et al., The age and evolution of non-LTR retrotransposable elements, Mol Biol Evol. 1999 Jun;16(6):793-805, which are incorporated by reference herein in their entireties.

[0086] Examples of the non-LTR retrotransposon polypeptides also include R2 from *Clonorchis sinensis*, or *Zonotrichia albicollis*. Example non-LTR retrotransposon polypeptides and binding components (5' and 3' UTRs) that may be used in the context of the invention are listed in Table 1 along with codon optimized variants of the non-LTR retrotransposons for expression in eukaryotic cells.

TABLE 1

Name	Sequences	
AB09 7126	**5' UTR**	AATCCCCCCTACCCAATCCCCCCGTCGTGACCTCCAGGCCAGGAATCACGAGCGTACGACAGTGGCCATCCGGCAATGACAATAG CGTGACTAACGACAATGAGTCAGATCCATGACCCTTGGAGTGGGTTAACCTCCGCCTCTTTAAAAAC (SEQ ID NO: 1)
	CDS	ATGGAAAGTACAGCAAAAGGAAAGTCATACTGGATGGCCCGTCGCCCAGTAGAAGGTGCCACGGAGGGATCTTTGGGTCGGGT CCCTTTCGTAACGCGAGATCCTAAGCGCAAACCAGAGGCTAAACGAACACTTACGCATGGCTTAGGACTACGAGAATGCTCGGT TGTCTTGACACGCCTCATCGAGGGGCGTCGAGGTCGCGATCACACACCATCAGGATGGAACGCACAGCGCGGCATGCCAAACG ACGAAAGCTCGGTCGAGGAGCCCAATGGGCCGATACCATCTAACCCCATACCAACGGGCACCCAAGCCCTGCCTGAACCTATGG CGGACGGGGAGCAGGGGGAGCACCCGGGAGTGGTGGTGACCCTGCCGCTCAGGGACTTAAACTGCCCCCTATGTGGCGGGTC GGCGAGCACCGCGGTGAAAGTGCAAAGACACTTGGCATTTCGCCACGGAACAGTGCCGGTTAGATTCAGCTGTGAATCATGTG GAAAAACTTCTCCGGGTTGCCATTCCGTCCTCTGTCACATTCCGAAATGTCGCGGACCGACAGGCGAGCCGCCTGAGAAAGTGG TTAAGTGCGAGGGATGCAGTAGGACGTTTGGCACAAGGAGAGCGTGTAGTATACATGAGATGCACGTTCACTCAGAAATCCGC AATAGGAAAAGAATTGCTCAAGACAGGCAAGAAAAAGGGACCTCGACAGATGGAGAGGGGAGAGCTGGAGTCGAAAGGGCT GACGCTGGGGAAGGTCCCTCTGGGGAAGGGATCCCCCCTAAACGTCCCAGACGTGCGAGAACGCCCAGAGAACCGTCTGAGCC CCCCGCGAATCCGCCGATTCTCTCGCCACAACCCGATCTGCCCCCAGGAGGCCTCCGGGACCTACTCCGGGAGGTGGCCAGTGG GTGGGTAAGGGCAGCGAGAGACGGAGGTACGGTGATTGACAGCGTGCTCGCAGCATGGTTGGATGGCAACGATCGGCTCCCT GAGCTGGTTGACGCGGCGACGCAAAGGACACTGCAGGGCTTACCTGCAGGGAGGTTGGCCCGAAGACCCGCAACTTTTGTTGC GCCTAACCGGAGGAGAGGCAGGTGGGGGCGCCGGCTCAAACTGCTCGCTAAGCGCCGCGCCTACCACGATTGCCAAATTCGGT TCCGAAAAGACCCAGCCCGCCTAGCCGCGAACATCCTAGACGGCAAAAGCGAAACAAGTTGCCCAATCAATGAGCAAGCGATTC ATGAGCACTTTCGAAACAAATGGGCAAATCCAAGTCCATTTGGTGGGCTGGGACGATTTGGGACGGAAAACAGGGCCAACAAC GCCCACCTCCTCGGGCCAATCTCCAAAAGCGAGGTCCAAACTAGCCTCCGAAATGCATCGAACGCCTCCACACCAGGCCCAGACG GCGTTGGGAAAAGGGACATTTCCAACTGGGATCCTGAGTGTGAGACCCTCACTCAGCTGTTTAACATGTGGTGGTTCACAGGTG TCATCCCCTCTCGCTTGAAGAAAAGTCGTACGGTGCTTCTGCCCAAGTCCTCAGACCCAGGAGCGGAGATGGAGATCGGCAACT GGAGACCAATCACCATCGGGTCGATGGTCTTGCGGCTTTTCACAAGGGTGATCAATACGAGATTAACGGAAGCCTGTCCGTTGC ACCCAAGACAGAGAGGGTTTCGACGAAGCCCCGGGTGTTCGGAGAACCTTGAAGTACTCGAATGTCTCCTCCGACACTCCAAAG AAAAGCGCAGCCAACTGGCAGTGGTATTCGTCGATTTTGCACAAGCGTTTGACACCGTCTCTCATGAACACATGCTGTCAGTCCT TGAGCAGATGAACGTGGATCCCCACATGGTAAATCTGATCCGGGAGATTTACACAAACAGCTGCACAAGTGTCGAGCTAGGCCG GAAAGAGGGACCAGACATCCCAGTGAGGGTTGGTGTTAAGCAAGGGGATCCTCTGTCCCCGCTGCTTTTCAACCTGGCTTTGGA TCCTCTCATCCAAAGTCTCGAACGCACAGGCAAAGGGTGTGAGGCCGAAGGTCACAAAGTGACAGCTTTAGCGTTCGCGGATGA CCTGGCACTGGTTGCGGGCTCGTGGGAGGGAATGGCACACAACCTTGCGCTTGTAGACGAATTCTGCCTAACCACCGGCCTCAC AGTCCAACCCAAAAAGTGCCACAGTTTCATGGTCAGGCCCTGCAGAGGTGCCTTCACAGTGAACGACTGCCCCCCATGGGTTCTG GGGGGCAAGGCCCTGCAGCTAACAAACATCGAAAACTCCATCAAATATCTGGGAGTAAAAGTCAATCCTTGGGCGGGGATTGA AAAGCCTGACCTTACAGTGGCACTAGACCGATGGTGCAAGCGCATTGGGAAGTCACTGCTCAAACCCTCACAGAAGGTATACAT TCTCAATCAGTTTGCCATCCCGCGACTCTTCTACCTGGCTGATCACGGTGGGGCCGGCGACGTCATGCTCCAGAACCTGGATGGG ACAATCAGGAAGGCGGTGAAGAAATGGCTGCATCTTCCACCGTCAACCTGCAACGGGCTGTTGTATGCCAGGAACTGTAATGGT GGCCTCGGTATATGCAAGCTCACTCGGCACATCCCATCAATGCAGGCGAGACGAATGTTCCGCTTGGCCAACTCATCGGACCCGT TGATGAAGGCCATGATGCGCGGCTCCCGAGTCGAACAGAAATTCAAAAAGGCCTGGATGCGGGCCGGGGGAGAGGAGAGTGC GCTCCCACGGGTGTTCGGGGCGAATCAGTACCAGGAAGGGGAGGAGGTCGCTAACGATCTGGTACCTCGCTGCCCAATGCCGA GCGATTGGAGACTGGAAGAATTCCAACACTGGATGGGCCTGCCGATCCAGGGTGTGGGTATAGCCGGCTTCTTCAGAAACAGG GTGGCTAACGGATGGCTCAGGAAGCCGGCAGGGTTCAAAGAGCGGCACTACATCGCCGCTCTACAACTGCGAGCATGTGTATA CCCCACCCTCGAATTCCAGCAAAGGGGCAGGAGCAAAGCGGGTGCGGCCTGCAGGCGGTGCTCATCCCGGTTGGAATCCAGCT CTCACATCCTCGGCAAATGTCCGGCGGTGCAGGGAGCCAGAATCAGGCGTCATAACAAAATATGCGACCTCCTGAAGGCCGAAG CCGAAACCCGGGGTTGGGAGGTACGCCGGGAATGGGCCTTCAGAACTCCGGCTGGGGAACTGAGAAGGCTCGACCTGGTACTC ATCCTCGGGGATGAGGCATTGGTCATTGACGTCACAGTAAGGTACGAGTTCGCTCCGGATACCCTCCAGAATGCCGGAAAGGAC AAGGTCAGCTACTACGGCCCGCACAAAGAAGCGATCGCTCGGGAGCTGGGCGTAAGAAGGGTCGACATACATGGGTTTCCGTT GGGTGCACGCGGACTTTGGCTCGCCAGCAACTCCAAAGTGCTGGAACTGATGGGATTGAGCAGGGAAAGAGTGAAGGTCTTCT CCAGACTCTTGAGTCGGAGAGTGCTCCTGTACTCTATCGACATCATGAGGACATTTTACGCAACCCTGCAATGA (SEQ ID NO: 2)
	3' UTR	AAATCCCAGCGGGATACAGCAAGAAGGTATCGGATCTAATAAGGTTGAGCGAGGAGAGGGTGGAGATCCTTTGGGGGGGGTC GGGCTAAGTTCCCCTCTCGGGTCCTCCCACGGTGACGCTCTACCCCTCCCTCCTCGCTCGTAGAACCCAACGGTGAACACGGTTG GCAGGATGAAGTGACGTGAGGGGTAAGACATGCGTACGTGAGCGCGCATTTTTGCTGTTCTCTGGACTGGGTTTCGTCCCCCTC

		ACAACCATCACTTACACTATAGGGGCACAGCGGCTCCTACCTCCCTCCCTATGACCCCCCCTTCCCATACCGATCCATGGCTGTTCT AGTCTGGACCGAGGGTCGGACGGGGCATTTGAAGGTAGCTGGAATCCTCCGCTGCTGCGAGCCTGAGGTCGATGGTTAGAGGT GAAATACTTGGGAGGAGACACAGCCTCCGGAGAGCCCCTCCCGGGTGGTCATCATGGCAACCGGGTGAAACCTTACGGTTTCAC TTACGAAACAGCACCATAACAGCGCCGTAATAGCGCACCGGTGTGACTACTGTCCAGTGCTGATATTCTCATCTGGAGAATACAA CACGGGTAATGGCAGAGTATTCAAAACCCAAATGTTTACGATCGACCAACGGAGTCGTTCCCTTGCATCTAGGCCGGACCCGAA ACTGCCGTAATTGCCCGTCCCCAAGGTAGCCTCTTAGAAAACCGAAGCCCGGTCGGGGCGGTGGTTGCGGCGGCGCTGCGGGG GCCTGCTGCTCGGGCGGCGTCGGTGTGCCGCGGTGGTTGCGGTGGTGCGGCGGGGATCTCGGTCCTTGCGGTGCCGCTGTGCC GCCGCGGTCGCGTCGGTGGCGCTGGGGTGGTGGCCCGAGTGGCGTCGGCGTGCCACTGCCCATAGTCGCCCGCGGGGGCGAC CGATCTGGAGGGGCGAGGGGGCTCGCGGGACTTTAACGAGAAACGGAACGCAACTTCTCGCATCGCTCCCGGGACTTTCCCCCC TCGTTCAGCCGAGGGATGCCAAAAGGCATGAAAGGTAAGTACCATACCGGTCCGCAAAACTCTCTTCTGACTCGGTTCTCTGTTG GTTTTCTAGAGTAACAACGAGGTGGAGGAGAGGGACATGGCAGGGACTCCCATTCGTGCCAGCGGGTGGGGACAGATCGAAG GAACGGTTCGAGGGCGTAACAGACGAGAGGGAATCCGGTCACACATTGATGCCATGCCTAAATAGGCGAGGTTTGTATTTCTAC TTTGTGGGTTCAGTATAGTCGGAGCATATGGTCGGTTGTCCCGTTGTTTTCACGGCGGGCAAGCGACTATCATGATAAAGTAGAA TGGGAGACGGGCTCCCTGACAAACCCGGAAAGGCGCCCCCCCGTGGTTCGTAGCAGCTGACGGATCACGCTCGAAGAAAAATG AGTGAGAGGGGACGCCGCAACCAC (SEQ ID NO: 3)
	Codon optimized	ATGGAAAGCACCGCCAAGGGCAAGAGCTACTGGATGGCCAGAAGGCCTGTTGAGGGCGCCACAGAAGGATCTCTGGGCAGAG TGCCTTTCGTGACACGGGACCCCAAGAGAAAGCCCGAGGCCAAGAGAACACTGACACACGGACTGGGCCTGCGCGAGTGTTCT GTGGTTCTGACCAGACTGATCGAGGGCAGAAGAGGCAGAGATCACACACCCTCTGGCTGGAACGCCCAGAGAGGAATGCCTAA CGACGAGAGCAGCGTGGAAGAACCTAACGGCCCCATTCCTAGCAACCCCATTCCAACCGGAACACAGGCCCTGCCTGAACCTAT GGCTGATGGCGAACAGGGCGAACATCCTGGCGTGGTGGTTACACTGCCTCTGCGGGATCTGAACTGCCCTCTGTGTGGCGGATC TGCCTCTACAGCCGTGAAGGTGCAGAGACACCTGGCCTTCAGACACGGCACAGTGCCTGTGCGGTTTAGCTGCGAGAGCTGCGG CAAGACATCTCCTGGCTGTCACAGCGTGCTGTGTCACATCCCTAAGTGCAGAGGCCCTACAGGCGAGCCTCCTGAGAAGGTTGT GAAGTGCGAGGGCTGCAGCAGAACCTTCGGAACAAGAAGGGCCTGCAGCATCCACGAGATGCATGTGCACAGCGAGATCCGG AACCGGAAGAGAATCGCCCAGGACAGACAAGAGAAGGGCACCAGCACAGACGGCGAAGGCAGAGCTGGTGTTGAAAGAGCT GACGCCGGCGAAGGACCTTCTGGCGAGGGAATCCCTCCTAAGAGGCCCAGAAGGGCCAGAACACCTAGAGAGCCTAGCGAGCC ACCAGCCAATCCTCCAATCCTGTCTCCTCAGCCTGATCTGCCACCTGGCGGACTGAGAGATCTGCTGAGAGAAGTGGCCAGCGGC TGGGTTCGAGCTGCTAGAGATGGCGGAACCGTGATCGATAGCGTGCTGGCTGCTTGGCTGGACGGCAACGATAGACTGCCTGA GCTGGTGGATGCCGCCACTCAGAGAACTCTGCAAGGACTGCCTGCTGGCAGACTGGCTAGAAGGCCAGCCACATTCGTGGCCCC TAATAGGCGGAGAGGAAGATGGGGCCGCAGACTGAAACTGCTGGCCAAGCGGAGAGCCTACCACGACTGCCAGATCCGGTTCA GAAAGGACCCTGCTAGACTGGCCGCCAATATCCTGGATGGCAAGAGCGAGACAAGCTGCCCCATCAACGAGCAGGCCATTCAC GAGCACTTCCGGAACAAGTGGGCCAATCCATCTCCTTTCGGCGGCCTGGGCAGATTCGGCACAGAGAACAGAGCCAACAACGCC CATCTGCTGGGCCCCATCAGCAAGTCTGAGGTGCAGACCAGCCTGCGGAACGCCTCTAATGCTAGCACCCCTGGACCTGATGGC GTGGGCAAGAGAGACATCAGCAACTGGGACCCCGAGTGCGAGACACTGACCCAGCTGTTCAACATGTGGTGGTTCACCGGCGT GATCCCCAGCAGGCTGAAGAAAAGCAGAACCGTGCTGCTGCCCAAGAGCAGCGATCCTGGCGCCGAGATGGAAATCGGCAATT GGAGGCCTATCACCATCGGCAGCATGGTGCTGCGGCTGTTCACCAGAGTGATCAACACCCGGCTGACCGAGGCCTGTCCTCTGC ATCCTAGACAGCGGGGCTTCAGAAGAAGCCCTGGCTGTAGCGAGAACCTGGAAGTGCTGGAATGTCTGCTGCGGCACAGCAAA GAGAAGAGATCCCAGCTGGCCGTGGTGTTCGTGGATTTCGCCCAGGCCTTCGACACCGTGTCACACGAGCACATGCTGTCCGTG CTGGAACAGATGAACGTGGACCCTCACATGGTCAACCTGATCAGAGAGATCTACACCAACAGCTGCACCTCCGTGGAACTGGGC AGAAAAGAGGGCCCTGACATCCCTGTTAGAGTGGGCGTTAAGCAGGGCGACCCTCTGAGCCCTCTGCTGTTCAATCTGGCCCTG GATCCTCTGATCCAGAGCCTGGAAAGAACAGGCAAGGGATGCGAGGCCGAGGGCCACAAAGTTACAGCTCTGGCCTTTGCTGA CGACCTGGCTCTGGTTGCCGGAAGCTGGGAAGGCATGGCCCATAATCTGGCACTGGTGGACGAGTTCTGTCTGACCACCGGACT GACCGTGCAGCCCAAGAAATGCCACAGCTTCATGGTCCGACCTTGCAGGGGCGCCTTCACCGTGAATGATTGTCCTCCATGGGT GCTCGGCGGAAAGGCTCTGCAGCTGACCAACATCGAGAACAGCATCAAGTACCTGGGCGTGAAAGTGAACCCCTGGGCCGGAA TCGAGAAGCCCGATCTGACAGTGGCACTGGACCGGTGGTGCAAGAGGATCGGCAAGTCTCTGCTGAAGCCCAGCCAGAAAGTG TACATCCTGAACCAGTTCGCCATTCCTCGGCTGTTCTACCTGGCCGATCATGGCGGAGCTGGGGATGTCATGCTGCAGAATCTGG ACGGAACCATCAGAAAGGCCGTGAAGAAGTGGCTGCATCTGCCTCCAAGCACCTGTAACGGCCTGCTGTACGCCAGAAACTGCA ACGGCGGACTCGGCATCTGCAAGCTGACAAGACACATCCCCTCCATGCAGGCCAGACGGATGTTCAGACTGGCCAACAGCAGCG ACCCACTGATGAAGGCCATGATGAGAGGCAGCAGAGTGGAACAGAAGTTCAAGAAAGCCTGGATGAGAGCCGGCGGAGAAGA GTCTGCTCTGCCTAGAGTGTTCGGCGCCAACCAGTACCAAGAGGGCGAAGAAGTCGCCAACGACCTGGTGCCTAGATGCCCTAT GCCTAGCGATTGGCGGCTGGAAGAGTTTCAGCACTGGATGGGCCTGCCTATCCAAGGCGTGGGAATCGCCGGCTTCTTCAGAAA CAGAGTGGCCAATGGCTGGCTGAGGAAGCCTGCCGGCTTTAAAGAGCGGCACTATATCGCTGCACTGCAGCTGCGCGCCTGCGT GTACCCTACACTGGAATTTCAGCAGCGGGGCAGAAGCAAAGCCGGCGCTGCTTGTAGAAGATGCAGCTCCAGACTGGAAAGCA GCAGCCACATCCTGGGCAAGTGTCCTGCAGTTCAGGGCGCCAGAATCCGGCGGCACAACAAGATCTGTGACCTGCTGAAGGCC GAGGCCGAAACCAGAGGATGGGAAGTGCGAAGAGAGTGGGCCTTTAGAACACCAGCCGGCGAGCTGAGAAGGCTGGATCTGG TTCTGATTCTGGGCGACGAGGCCCTGGTCATCGATGTGACAGTCAGATACGAGTTCGCCCCTGACACACTGCAGAATGCCGGCA AGGACAAGGTGTCCTACTACGGCCCTCACAAAGAGGCCATTGCCAGAGAGCTGGGCGTCAGAAGAGTGGACATCCACGGATTC CCACTGGGAGCCAGAGGACTGTGGCTCGCCAGCAATAGCAAGGTGCTCGAGCTGATGGGACTGAGCCGCGAGCGAGTGAAGG TGTTCAGCAGACTGCTGTCCAGACGGGTGCTGCTGTACTCCATCGACATCATGCGGACCTTCTACGCTACCCTGCAGTGA (SEQ ID NO:4)

PERE RE-9	**5' UTR**	ATCTCACGTTTTAATTTATTTTTGAACTACTGCAGTCTGAGTGCTTCTAACGACCCGAAGGCTCAGAAACTACCCACTTCTTGAACT GCTACTTTTTGCTGTTTATCCACAACAACAGTTGTGATTCTATTCTCCANATATTCCTTGTGCTTTTGTCAACATTATTCTATACCAA CTGTACCACCTACTTCTTCATCTCACGTTTTAATTCTGGTCTTATTTTCTCATCATTAGTCACGGAGAGGGCCTATGAACGGTCCGT GACGCGAAATTCTATCCGCGATTTCGACCTCTCCTGCTAGTGGTCCCCGAAGTACGGTTCCTCTGGCCTGTCAGTTGTGTTAAAAC TATATAATAACG (SEQ ID NO: 5)
	CDS	ATGCCGGTCTCAACCGGCGCAGAAACTGACATAACCTCTTCTTTGCCTATTCCTGCATCCTCAATCGTCTCGCCAAACTACACACTC CCTGATTCCTCTTCAACCTGCCTTATATGTTTCGCTATCTTCCCCACCCACAACATACTCCTCTCCCATGCCACTGCAATCCACCATA TTTCTTGTCCTCCTACTCCAGTGCAAGACGGTTCTCAGCAGATGTCTTGTGTTCTTTGCGCCGCCGCTTTTTCATCTAACAGGGGAC TAACACAACACATTCGCCACCGGCACATCTCCGAATATAACGAACTAATCAGACAACGAATTGCAGTGCAGCCGACGTCTCGCAT ATGGTCACCATTCGATGATGCTTCTTTACTATCCATCGCTAACCATGAAGCCCATAGATTCCCCACGAAGAATGACCTATGCCAAC

		ATATCAGCACCATACTAACACGCAGGACGGCAGAAGCCGTCAAACGCCGACTCCTCCACCTACAGTGGTCGAGATCACCAACAG CGATTACTACCTCTTCGAATAATCACACAATCACAGACATCCCCAATACCGAGGCCCGATATATTTTTCCGGTAGACCTAGACGAA CATCCACCATTGTCTGATGCCACAACCCCCAACGCATCGACACATCCACTCCCAGAACTCCTTGTCATCTTGACACCGCTTCCATCC CCGACTAGACTACAAAACATATCCGAATCACAGACCTCCCATGAATCTAATAAGAACTCAATGCATACACCGCCAACGTATGCCT GCGATCCGGATGAGACACTAGGGGCTACTCCCTCATCAACTATTCCCTCATGCTTCCACAGTTATCAGGACCCCCTAGCTGAACAA AGAGGCAAACTCCTGAGGGCATCCGCCAGCCTACTACAAAGCAGTTGTACTCGCATACGGTCCTCCAGCCTGCTCGCCTTCCTCC AAAACGAATCCACATTAATGGACGAGGAACACGTGTCCACCTTCCTCAATAGTCATGCAGAATTCGTCTTCCCTAGAACATGGAC CCCATCCCGACCCAAACACCCCTCCCACGCCCCAGCTAATGTTTCTAGGAAGAAAAGGAGGAAAATAGAGTACGCACACATCCA GAGACTCTTCCACCACCGTCCCAAAGATGCCTCCAACACCGTTCTAGACGGTCGGTGGAGAAACCCCTATGTCGCAAACCATTCA ATGATTCCAGACTTCGACTGCTTCTGGACAACAGTCTTTACTAAAACAAATTCCCCAGACAGCCGGGAGATTACTCCAATCATCCC TATGACTCCCTCTCTCATTGACCCGATCCTCCCCTCTGACGTCACATGGGCGCTGAAAGAAATGCATGGCACGGCCGGTGGGATT GATCGTCTGACATCGTACGATCTGATGAGATTCGGGAAGAATGGTCTTGCTGGATATCTCAACATGCTACTCGCTCTTGCATACCT TCCCACTAATCTTTCAACAGCACGGGTAACTTTCGTCCCCAAGTCATCAAGTCCTGTGTCACCTGAGGACTTCCGTCCCATCAGTG TCGCTCCAGTAGCCACTAGGTGCCTGCACAAAATTCTAGCTAAGAGATGGATGCCGCTCTTTCCACAGGAACGACTTCAGTTCGC TTTCCTAAACCGAGATGGATGCTTTGAAGCAGTTAATCTTCTGCACTCGGTCATACGGCACGTCCACACCCGCCATACAGGAGCA TCCTTCGCCCTGCTCGACATATCACGGGCCTTTGACACTGTATCACATGACTCCATCATCAGAGCGGCGAAAAGATATGGGGCAC CTGAACTGTTATGCCGCTACCTCAATAACTATTACCGACGTTCAACCAGCTGCGTCAACCGCACTGAATTGCATCCTACGTGTGGG GTGAAGCAAGGAGACCCCCTGTCGCCACTCCTCTTCATCATGGTTCTCGACGAAGTACTGGAAGGTCTAGATCCAATGACCCACC TAACAGTTGATGGAGAGAGCTTGAACTACATAGCTTATGCTGACGATCTCGTAGTTTTCGCTCCAAATGCGGAACTCCTTCAACG GAAACTCGATCGGATCTCCATACTTCTACACGAGGCTGGATGGTCGGTTAACCCTGAAAAAAGCCGGACCCTGGACCTAATCTCT GGTGGCCATTCCAAAATCACAGCGCTCTCTCAGACAGAATTCACCATCGCGGGGATGCGTATACCACCGCTTTCTGCCGCCGACA CCTTCGACTACCTGGGTATCAAATTCAACTTCAAGGGCCGATGCCCAGTGGCCCATATTGACTTATTGAACAACTACCTCACGGAA ATATCGTGCGCTCCACTTAAGCCGCAGCAGCGCATGAAGATCTTGAAAGATAATCTACTCCCTCGACTCCTATACCCCCTGACTCT AGGAATAGTACACCTGAAAACCCTGAAGTCAATGGACCGAAATATCCACACGGCCATAAGGAAATGGTTGCGGCTACCCTCCGA CACCCCGCTAGCATATTTTCACTCACCCGTCGCTGCCGGAGGCCTAGGGATCCTCCATCTGTCCTCATCGGTTCCATTCCACCGTC GAAAACGTCTAGAAACCCTCCTATCTTCACCGAACCGCCTACTGCACAAGTTGCCAACTTCCCCAACACTAGCTTCTTATTCACACC TTAGTCAACTGCCAGTTCGAATTGGGCACGAGACCGTAACGTCTAGAGAAGAGGCTTCCAACAGCTGGGTGAGACGATTACATT CGTCCTGCGACGGGAAGGGACTACTCCTAGCACCACTAAGCACCGAGTCCCATGCATGGCTGCGCTACCCCCAGTCTATTTTTCC ATCTGTTTACATCAACGCCGTTAAATTACGAGGTGGCTTACTATCCACCAAAGTCAGGAGATCTCGCGGAGGTAGAGTGACGAAT GGCCTGAACTGTCGAGGCGGTTGCGCCCATCATGAAACGATCCACCACATTCTGCAACATTGCGCGCTCACCCACGACATCAGAT GCAAACGCCATAACGAACTATGCAACCTTGTGGCAAAGAAACTGCGTAGGCAAAAAATCCATTTCTTACAGGAGCCCTGCATTCC TCTAGAAAAAACCTACTGCAAACCTGATTTTATAATTATACGTGACTCAATTGCTTATGTTCTAGACGTCACTGTATCGGACGACG GAAACACCCACGCCAGCCGCCTGTTAAAAATATCAAAATACGGCAATGAGCGAACCGTCGCGTCGATCAAGCGTTTCCTCACATC CAGTGGATATATCATTACCAGTGTTCGACAAACACCAGTCGTCCTTACATTCAGAGGTATTCTGGATAGAGCAAGTTCACAATCCC TACGACGCCTATGTTTTTCATCCCGTGACCTCGGTGACCTTTGCCTGAGTGCGATTCAAGGCTCAATTAAAATATATAATACCTAT ATGAGAGGAACCTAA (SEQ ID NO: 6)
	3' UTR	CGGCTGAACGAATAGCCCCCTTCACTCTTAGACATTCCCCCACTGTTGTTGCTTATCTTCATGCTCTTGTGTTAATTGACTGCTCTCT TCTGGGTTGACGTCTGATTGTCTCTCTCTCTTTCCATATTGCTTGCTCTGCCCGCTTACTTCCAATAGTTGTCATATTATGTCTTTGTT TACTTGCCATGTCTAACGACAATTACTTTATCTACCTTAGTTTGTCCTCTTGGTTTCGATTGCCTTCATATGTTCATGGCGGAATCTG ATGTTTATAATGACTATTCCTATTACCACCACTACAACTACTATTATTATTTTCATTACTATTAACATTATTATAAACATTATTACTAT TATTATTATTACTATTATTACTTCTACAATTAATATTATGGCTACTCCTCTCAGCACACCAATAAAATATCAATCAAACATCTCAATT ATATCCACCTATTAAACTCTCTCTATTTCCCCTGAGTTATAAACTTACAATTCAGTCTAACCGAATATCTCTCTTTTACAAATCTTAA GTATGTAATTTTGTGCCAAACCCATTTGGGTCTGTACAATTTGATACTTAAAAATAAATGTTATTAGCC (SEQ ID NO: 7)
	Codon optimized	ATGCCAGTGTCTACAGGCGCCGAGACAGACATCACAAGCAGCCTGCCTATTCCTGCCAGCAGCATCGTGTCCCCAAACTACACCC TGCCTGACAGCAGCAGCACCTGTCTGATCTGCTTCGCCATCTTTCCCACACACAACATCCTGCTGAGCCACGCCACAGCCATCCAC CACATCAGCTGTCCTCCAACACCTGTGCAGGATGGCAGCCAGCAGATGAGCTGTGTGCTGTGTGCCGCCGCTTTCAGCAGCAAC AGAGGACTGACCCAGCACATCCGGCACAGACACATCAGCGAGTACAACGAGCTGATCCGGCAGAGAATCGCCGTGCAGCCCAC CAGCAGAATCTGGTCCCCATTCGATGATGCCAGCCTGCTGTCTATCGCCAACCACGAGGCCCACAGATTCCCCACCAAGAACGAT CTGTGTCAGCACATCAGCACCATCCTGACCAGACGGACAGCCGAGGCCGTGAAAAGAAGGCTGCTGCATCTGCAGTGGTCTAGA AGCCCTACCGCCATCACCACCAGCTCCAACAACCACACCATCACAGACATCCCCAACACAGAGGCCCGGTACATCTTCCCCGTGG ACCTGGATGAACACCCTCCTCTGTCCGATGCCACCACACCTAACGCCTCTACACACCCTCTGCCTGAGCTGCTGGTCATCCTGACA CCTCTGCCTTCTCCAACCAGACTGCAGAACATCTCCGAGAGCCAGACCAGCCACGAGAGCAACAAGAACAGCATGCACACCCCTC CAACCTACGCCTGCGATCCCGATGAGACACTGGGAGCCACACCTAGCAGCACAATCCCCAGCTGCTTCCACAGCTACCAGGATCC TCTGGCCGAGCAGAGAGGCAAACTGCTGAGAGCCTCTGCTAGCCTGCTGCAGAGCAGCTGCACCAGAATCAGAAGCTCAAGCCT GCTGGCCTTCCTGCAGAACGAGAGCACCCTGATGGACGAGGAACACGTGTCCACCTTTCTGAACAGCCACGCCGAGTTCGTGTT CCCCAGAACCTGGACACCCAGCAGACCTAAGCACCCTTCTCATGCCCCTGCCAACGTGTCCAGAAAGAAGCGGCGGAAGATCGA GTACGCCCACATCCAGCGGCTGTTCCACCACAGACCAAAGGACGCCAGCAACACAGTGCTGGACGGCAGATGGCGGAATCCCTA CGTGGCCAACCACAGCATGATCCCCGACTTCGACTGCTTCTGGACCACCGTGTTCACCAAGACAAACAGCCCCGACAGCAGAGA GATCACCCCTATCATCCCCATGACTCCCAGCCTGATCGACCCCATCCTGCCTTCCGATGTGACATGGGCCCTGAAAGAGATGCAC GGCACAGCCGGCGGAATCGACAGACTGACAAGCTACGACCTGATGCGCTTCGGCAAGAATGGCCTGGCCGGCTACCTGAATAT GCTGCTCGCTCTGGCCTACCTGCCTACCAATCTGAGCACCGCCAGAGTGACCTTCGTGCCCAAGTCTAGCAGCCCCGTGTCTCCCG AGGACTTCAGACCTATTTCTGTGGCCCCTGTGGCCACCAGATGCCTGCACAAGATTCTGGCCAAGCGGTGGATGCCTCTGTTCCC TCAAGAGAGACTGCAGTTCGCCTTCCTCAACCGCGACGGCTGTTTCGAAGCCGTGAATCTGCTGCACAGCGTGATCAGGCACGT GCACACAAGACACACCGGCGCCAGTTTTGCCCTGCTGGATATCTCCAGAGCCTTCGACACCGTGTCTCACGACAGCATCATCAGA GCCGCCAAGAGATATGGCGCCCCAGAGCTGCTGTGCAGATACCTGAACAACTACTACCGGCGGAGCACCAGCTGCGTGAACAG AACAGAACTGCACCCTACCTGCGGCGTGAAGCAAGGCGATCCTTTGAGCCCTCTGCTGTTCATCATGGTGCTGGATGAGGTGCT GGAAGGACTGGACCCCATGACACACCTGACAGTGGATGGCGAGAGCCTGAACTATATCGCCTACGCCGACGACCTGGTGGTGTT CGCCCCTAATGCTGAACTGCTGCAGCGGAAGCTGGACAGGATCTCTATCCTGCTGCATGAGGCTGGCTGGAGCGTGAACCCCGA GAAGTCTAGAACCCTGGACCTGATCTCTGGCGGCCACTCCAAGATCACAGCCCTGAGCCAGACAGAGTTCACAATCGCCGGCAT GCGGATCCCTCCACTGTCTGCCGCCGATACCTTTGACTACCTGGGCATCAAGTTCAACTTCAAGGGCAGATGCCCCGTGGCTCAC ATCGACCTGCTGAACAATTACCTGACCGAGATCAGCTGCGCCCCTCTGAAGCCTCAGCAGAGAATGAAGATCCTGAAGGACAAC

		CTGCTGCCTAGACTGCTGTACCCTCTGACACTGGGCATCGTGCACCTGAAAACCCTGAAGTCCATGGATCGGAACATCCACACCG CCATCCGGAAGTGGCTGAGACTGCCTAGCGATACCCCACTGGCCTACTTTCACTCTCCTGTGGCTGCTGGCGGACTGGGAATCCT GCATCTGTCTAGCTCCGTGCCTTTCCACAGACGGAAGCGGCTGGAAACACTGCTGTCAAGCCCCAACAGACTGCTGCATAAGCTG CCTACAAGCCCCACACTGGCCAGCTACTCTCACCTGTCTCAGCTGCCTGTGCGGATCGGACACGAGACAGTGACCTCTAGAGAAG AGGCCAGCAACTCCTGGGTCCGAAGGCTGCACTCTAGCTGTGATGGCAAGGGACTGCTGCTTGCCCCACTGAGCACAGAATCTC ACGCCTGGCTGAGATACCCTCAGAGCATCTTCCCTAGCGTGTACATCAACGCCGTGAAGCTGAGAGGCGGACTGCTGTCCACAA AAGTGCGGAGATCTAGAGGCGGCAGAGTGACAAACGGCCTGAATTGCAGAGGCGGATGCGCCCACCACGAGACAATTCACCAC ATCCTGCAGCACTGCGCCCTGACACACGACATCAGATGCAAGCGGCACAACGAACTGTGCAACCTGGTGGCTAAGAAGCTGCGG AGACAGAAGATCCACTTCCTGCAAGAGCCCTGCATTCCCCTGGAAAAGACCTACTGCAAGCCCGACTTCATCATCATCCGGGACA GCATTGCCTACGTCCTGGACGTGACCGTGTCCGACGATGGAAATACCCACGCCTCCAGGCTGCTGAAGATCTCTAAGTACGGCA ACGAGCGGACCGTGGCCAGCATCAAGAGATTTCTGACCAGCTCCGGCTACATCATCACCAGCGTCAGACAGACCCCTGTGGTGC TGACCTTTAGGGGCATCCTGGATAGAGCCAGCTCTCAGAGCCTGCGGAGGCTGTGCTTTAGCTCCAGAGATCTGGGCGACCTGT GCCTGAGTGCCATCCAGGGCTCCATCAAGATCTACAACACCTACATGCGGGGCACCTGA (SEQ ID NO: 8)
R2-1_GA	**5' UTR**	CATATTGGGGTCTCAGGAGGAGACACAGGGTCTGTTGCGGCTCCGGTAAACGGTACCGGAGTCGGTTAAGCATCGTTTGGGCCC GCCTCCACGTGGTGGTCCGCGGTAACACCAATAGGGTGGCTAAGAGGCCCAGTAATTTCCCCGAATTGTCTTCCCCCCCGCGCGG GGGGGACCCCCCTTTAGTGTCGGAGCGGTCGCGCCTCCGCGTTTGGGGTGTCGCAGGCGTGAGCCTTCGTCCCCTTAAGTTCAG ACGGTCCCGGCTTCTTGCCGGGCCAACCCCCGGTGCAGCGTTCTCCCATGTTGGATCGGCACCCAGCCCCGGGTGCCATGCGAGT TCAGACATTTTGTTTATGTATCGTCTGCGTGGTTGACTTGCTAAGCTCATTTCCTCCTCTCACTGCGTCCCCCCAGGTGCTGATCGG TTGAAGAGGATTCGTCGTTGACCTCGGCGGTGAATTTGGGATTGTATTATACAGGTAGGTATAGAGGGCGTGCGG (SEQ ID NO: 9)
	CDS	ATGTTGCGTGGCGGTGTTGGTACTCCCCCGGCTGGGGGAGCGGGTGCGGTGGGGCCAGGCATGGCCTCGCCGGGTGGTTGCA GTGTCCGGTTCAGTCCCGGAGGGAGGCGACTGCTTGGCCACAGGACTGGAGGGTTGAGTCCCTCCGTGTCCTGGAGGCTCAAG CGACTGTCTGTCTCTCTGAGGCGCTGGAGCGGGCCTGGGCTGCTAGGTGCGGATGGTGCGGGGGGAGGCGCTGCGGTGGCCTC CCCCAGGGGTACGCAGGTCCTGGGAAGTGGGGCCGGGCGTCGGTGGCTTGGGCACGGGTCGCGAGGGTCTTCTCCTTCTGCGG CCCGGGGGCTAAGGCGGCTGACGGTACGGTTGAAGCGACTCAGCGGTGGCCTGTTGTCCCCTAAGGCGTGTCGGGATGCGGAA GAAGGAAGCTCCAGCAGCCCAGGGTTCCGGAATCCAAAAGGTCTCGGGGGAAGGGGGTTGACGCCTCTCGGATCCCGTAGATT TTGTCGGCTGACCGTCTCCCTGAATCGCTGGAGGGGCAGTCTGGTGAAGTTGAACGCTAGTAGCAGGGCCTCCGGCCGGAGGA CCCCTGTGAAACCCGCTTGTGACTCTAGAGCCGGACGGGGCTCGGAGCATGCGGAGGGAGGTGGAGTGAGCGCTGCACCTATG GTGTTGCGCAGTCGGCGTAAGCTCACCTTCTCTGTGGATGGCGACTCTAACTCCGGGGATAGGGCCCGGAGCGGGTCCGTCTCT GCAGCCCGTCCTGGCCACTTGTTGGTGGATGGTGAGAGTGCGTCCTCAAGATCTGGCCCCGCGGGGGATGCCAGGTTGGCGGG GCCTTCTACGCGGAGTAGGAGGAAGGGTTGCCTTCCCCCGGTCGACTTTGAAAACCCGAAGAAGCGCACACGGTTGATGGCTAA GATGACGAATGGTAATCCTACCTCGCACGTCCCTTGCCCTGCCCCGTGCTCAAATGGGCATGAAGGAGGTGGGCGAGTTGCGGT GATCGAGGGGCGGCTGCCGGAGTTAAGCGGTAGTAGGATCTCTGGAATACAGCCAGCCCTGCCTGTTGAAACCAGCTTTGTCGG CCAATCGACTGGCCGGGGCGCGGACGGCGATGCGAATGCGAATAGTAGCCCGCCTTCTCCTAATCTGGGCGGCTCGGTTGGGA TGGTGCCTGCCGTGCGTGATGGTACCCCGCCGCTTGGGCGTCCAGGAGAGGATCACTCGCGGGAGTGTGCAGGGGGAAATACT CCCCTCTGGATGCTGGAGGACAGTTTCCGGTGTGACTACTGTCCTAGGGAATTCGGCACAAGAGCGGGGCGCTCGTTGCACATG CGCAGGGCTCACCTGGCCGAGTACGACGGGGCAGGTTTCTGTTGGGGTGAACGTCTCAGTGAATTCGCCGCTACGCGCCTCTGG TCGACGGAGGAAACCAAAAAGCTGGCCGTGTTTTGTGAGAGGGGTGTGCCCTCACCGTCGGAATGCAGAGCCATTGCAGCCTCT CTGGGCGCAGGAAAAACACATCATCAGGTTAGATCGAAGTGTCGACTGGTGTTCGAGGCCATTCGGCGGCGTGAATTGCTTGAG GTGGCTGCTGCCACGGAGCGTTTGGAGAAAAGCGCTAGGCGGAAGCAGCCCGCCGTACCACCGGCACCCGTACACGGAGTGAG AGGGGTCCTGCGGGGCCTACTAGGGAAGCGGGTGCCGAGAGAGGGTGGTACCACAGGCAGCACCTCAGCAAGGATCGTCAGG AGAGACGACTGCCGTCAGGGGGCAGTTGCGTCGGCTTCTCTCAATCTGATCAGAAGGCTGGGTCGAAAGGCAACGGGCCGCTC CGGCAGGAGACGGGTCCTTGGACGCCCACCCAGGATGGATGTAAGGCGTAGCGTGAGGATGAGGAGGATGCGCAGGTTCCTC TATCGGTTGGCCCGGCTGGGCTGGGCCAAGTTGGCTATGTTTGTCCTGGACGGACAGATGGGGGCGAGCTGCCCCGTTCCACTC GTCGAAGTGTCGGCGGTCTTCCGGGAGAGGTGGAGCATAGTCAGAGCCTTCCTGGGTCTGGGTCAGTTCGGGGGCTTCGGGAC TGCCGACAACGCAGGATTTGGGAAGCTGATCGATCCGGCTGAAGTCAGGGCCCATCTCCAGTCCATCAAGAACCGGTCTTCCCC GGGCCCGGATGGCATCACCAAGGTGGCGCTGTCCAAATGGGACCCCGAAGGGATTAAATTGGCGCACATGTACTCAACATGGTT GGTATCGGCAGGCATCCCTAAGGTCTTCAAGAAGTGCAGGACGACACTTATCCCAAAGACCGGGGACGTTAGTCTACATGGTGA CGTGGGGCAATGGAGGCCCATAACCATTGCGTCCCTGGTCCTGAGACTCTATTCGCGGATCCTGACGGAAAGGATGACAGTGGC CTGTCCTAGCCACCCGCGCCAGAGGGGCTTCATTGCCTCCCCGGGCTGTTCGGAAAACCTCATGCTGTTGGAAGGTTGCATGAGT CTCAGCAAGGCAGGAAATGGCTCCCTCGCGGTTGTGTTCGTCGACTTTGCGAAGGCCTTCGATACCGTCTCCCACGAGCACCTCC TGAGTGTTCTGGTGCAGAAAGGCTTGGACCAACACATGGTGGAGTTGATCAAGGACTCCTACGAGAACAGCGTGACCAAGGTG CACTGTCAGGAGGGTTGTTCCACTGACATCGCCATGAAGGTGGGAGTGAAGCAGGGTGACTCCATGTCCCCTCTCCTCTTTAACC TGGCGCTGGATCCGCTTATCCAGCAACTTGAACGCGAGGGCCGGGGCTTCCCAGTAAATGGGAAGTCCATTACTGCGATGGCAT TTGCGGATGACTTGGCCATAGTGAGTGACTCTTGGGAAGGCATGAGAGCCAACCTTGATATCCTGGTGGACTTCTGCGAGCTTA CTGGAATGCGAACCCAGCCCAGTAAGTGCCACGGGTTCCTGATTGAGAAGAGTGGCAGCAGGTCGTACAAAGTGAACAGGTGC GAACCGTGGCTGCTGAACGACACAGCTCTTCACATGGTCGGGCCTAAGGAATCAATCAAGTACCTGGGCGTCCAGGTGAACCCG TGGACAGGGATCTTCGCTGAGGATACGGTTGCCAAACTACGACAGTGGGTAGTTGCAATCTCCAAGACGCCTCTACGTCCGCTT GACAAGGTGTCCCTGTTGTGCCAGTTTGCCGTACCGAGGGTCATCTTCGTGGCTGATCACTGCATGCTATCTGCGAAGGCCCTGA CAGAAATGGATAGGAGCATAAGACAAGCAGTGAAGAGGTGGTTGCACCTGGCCAGGTGTACCACGAACGGCCTCCTCTACTCA AGGAAATCCAGCGGTGGTCTGGGTATCCCAAAATTGTCGATGATTGTTCCGGCCATGCAGGCCAGGAGACTCCTGGGCCTGTCC CGTTCTAAGGACGAGACGGTCAGGTGGATGTTTCTGGAGACAACTGATCACGTGGCGTTTGAGAGGGCATGGCTGAGGGCTGG AGGGTCGCCAGATGAGGTACCGGAGCTGGGTCCGGATCTGGTGGAGGGCTCCCCTGCGGAGGGGAACGCTGACCCTGTCAGC ACGGTGAGGCCAAGGAAGCGCATAGTCCCGTGTGACTGGCGTCAAGTCGAGTTCGACAGATGGGCCGGTCAATTGGTGCAGGG AAAAGGGATTCGGACGTTCGAAGCGGACAAGATCAGCAACTGCTGGTTGTACGACTACCCGCCAAACAAGCTGAAGCCTGGGG ATTTTACGGCGGCTGTCCAGCTTAGAGCGAACGTTTACCCGACCCGGGAGCTAGCGGGTCGCGGAAGGACCGATACGATAGAT GTCTGTTGTCGACACTGTGGGGAGGCCCCAGAGACTTGCTGGCACATCCTTGCGCTCTGCCCGAAGGTTAAGCGGTGCCGTATTC AGAGGCACCACAAGGTGTGCCAGGTCCTCGTCGCGGAGGCTGAGCGCCATGGATGGGAAGTGGAAAGGGAAAAGCGCTGGAT GCTGCCCTCCGGGGAGTGTGTCGCGCCGGACCTGATCTGCTGGTTGGATGAGCTGGCGCTCATTGTCGATGTGACGGTGAGGTA

		CGAGTTCGATGAGGAGTCGCTAGAACGCGCGCGAATCGAGAAGGAATGCAAGTACCGCCCTCTCATTCCAGTGATCAGGGCGA GCAGAGTTCAGACGAAGAAGGTGACGGTCTATGGCTTCCCTCTGGGAGCCAGGGGAAAGTGGCCTGCTAAGAACGAGCTGCTG CTCGCCGACCTCGGCCTGAGCAAGGCTCGGACTCGGAGTTTTGCTAAACTCCTGAGCCGCAGAGTTCTCTTACATTCTCTGGATG TTATGAGGACGTTTATGCGTTAA (SEQ ID NO: 10)
	3' UTR	GGAGGGGAGTAGGTCTCTACTCTGACCCGAAGGGCCCCCCCGTTTCAGACCTGATTCTAGGCTACCTGTGCCTAATTGGGGGGG TCCCAAAGAGATGTTGTCTGTTGTAGAAGGGTTTGCGCCACTGACTGCACGGAAGGGTGGGCCTCGACAGGTAGGGGTTACAT GACTCCGTGCTGCTCAGCAGACCCGCGCCTCTGAGACCGGGTAGGGCTACTTGAACAAGCGACGCCCTGGTGTATGTCCGTATC CTAACCTGGTTTGGGAAAGCCGATACCGGCAATGCCCGCCACAGGTGTCGCGCACCCCACGGGATGACGTATGGGCCCCGGGG GACCTCATGGATACTCCACTGGACTTGCACAATCCTGGTGTACTGGATGCAGCGACGTTGGTGACATAAGCAATCGCTAAGTCG GGGTAGGGGAGGTGGGGACCTCGGCACGGCTGTAGGAACGGGTGTATGGGCTCCGGCAGCCGTCGTCACTCCCATACAACACA GGGGCTGCATCCTGGTGGCCGGTGCTAGTTGGTTCTGGAAGCCCGCCCGGGCTGGTTCGCAGAAGCAGGGTGCGCCCAGGGTA GGTTTGGTATATCTGGGTCCGGTGCGATACCTATCGATGGGCAGCGAGGGCCGCCTCGTGACGCGCTGTGTGGAGCTGGAGCC GGCCTGGGTATGAACAGTTCTTGCGGATGTGGCGTAGCTAGATAGTACCCGTGGTTGTGGGCGTGGTGTCGACCAAATGTTGTC CTGTGTGCACATAGGCCAAGGGTTACGTGGGTGGCAGTCAGAAGCACCCGCACCTGGAAGTGATTGCCCCGGGATCCCGGCTCT CTGTGAAGAGCTACCTTGAGGAAAGGTGTTCCGCTGGAACTCAAGACCCTACAGTAGGGGATATCAACTGGCTTTGAGGTGCTG TGATTCCGGAACCAGGGCGAGGGCGAGTACTTAGAGCATGTCCAAAAGCCCGGGGAACGTTCCGGGGGCCTGCTTGGGTCGTT GGACCCACATCCGTAAAACGATGGATCTCGCGTCGGCGCTCGGGAGAACTTCCCGCATGAACGCTGATTGCATGTGAGAACGCC CCCACGGCGGCGGGGCAGGCGCTCCCCCTGGGTGTAAGGCTCGGGGGGGTCACGGCTCCGCTCTAAAAG (SEQ ID NO: 11)
	Codon optimized	ATGCTGAGAGGCGGAGTGGGAACACCTCCTGCTGGCGGAGCTGGTGCTGTTGGACCTGGAATGGCTTCTCCTGGCGGATGCAG CGTGCGATTTTCTCCAGGTGGAAGAAGGCTGCTGGGCCACAGAACAGGTGGACTGAGCCCTTCCGTGTCTTGGCGGCTGAAGA GACTGTCCGTGTCTCTGAGAAGGTGGAGCGGACCTGGACTGCTTGGAGCTGATGGTGCTGGTGGCGGAGCAGCTGTTGCTTCTC CTAGAGGAACACAGGTGCTCGGATCTGGCGCCGGAAGAAGATGGCTTGGCCACGGCTCTAGAGGCAGCTCTCCATCTGCTGCTA GAGGCCTGAGAAGGCTGACCGTGCGCCTGAAAAGACTGAGCGGAGGACTGCTGAGCCCTAAGGCCTGTAGAGATGCCGAGGA AGGCAGCAGCTCTAGCCCCGGCTTCAGAAACCCTAAAGGCCTCGGAGGCAGAGGACTGACACCTCTGGGCAGCAGAAGATTCT GCCGGCTGACAGTGTCCCTGAACAGATGGCGGGGCTCTCTGGTCAAGCTGAACGCCTCTTCTAGAGCCAGCGGCAGACGGACAC CTGTGAAGCCTGCCTGTGATAGCAGAGCCGGAAGAGGATCTGAACACGCCGAAGGCGGCGGAGTTTCTGCTGCTCCTATGGTG CTGCGGAGCAGACGGAAGCTGACCTTTTCCGTGGACGGCGACAGCAACTCTGGCGACAGAGCTAGAAGCGGCTCTGTGTCTGCT GCCAGACCTGGACATCTGCTGGTGGATGGCGAAAGCGCCTCCTCTAGATCTGGACCTGCTGGGGATGCTAGACTGGCCGGACCT AGCACCAGAAGCAGAAGAAAGGGCTGCCTGCCTCCTGTGGACTTCGAGAACCCCAAGAAACGGACCCGGCTGATGGCCAAGAT GACCAACGGCAACCCTACCAGCCACGTGCCATGTCCTGCTCCATGTAGCAATGGCCACGAAGGTGGCGGAAGAGTGGCCGTGAT TGAAGGCAGACTGCCTGAGCTGAGCGGCTCCAGAATCTCTGGAATCCAGCCTGCTCTGCCCGTGGAAACCTCTTTTGTGGGCCA GAGCACTGGCAGAGGCGCTGATGGGGATGCCAATGCCAATAGCAGCCCTCCTTCTCCTAATCTCGGCGGCAGCGTTGGAATGGT GCCTGCCGTTAGAGATGGCACCCCTCCACTTGGTAGACCCGGCGAGGATCACAGCAGAGAATGTGCCGGCGGAAACACCCCTCT GTGGATGCTGGAAGATAGCTTCAGATGCGACTACTGCCCCAGAGAGTTCGGCACCAGAGCTGGCAGATCCCTGCATATGAGAA GGGCCCACCTGGCCGAGTATGATGGCGCTGGATTTTGCTGGGGCGAACGCCTGTCTGAGTTCGCCGCTACAAGACTGTGGTCCA CCGAGGAAACAAAGAAACTGGCCGTGTTCTGCGAGCGGGGAGTGCCTTCTCCAAGCGAGTGTAGAGCCATTGCCGCTTCTCTCG GAGCCGGAAAGACACACCATCAAGTGCGGAGCAAGTGCCGGCTGGTGTTCGAGGCCATTCGGAGAAGAGAACTGCTGGAAGT GGCCGCAGCCACCGAGAGACTGGAAAAGAGCGCTAGAAGAAAGCAGCCCGCCGTGCCTCCTGCTCCTGTTCATGGTGTTAGAG GCGTGCTGAGGGGCCTGCTGGGAAAGAGAGTTCCTAGAGAAGGCGGCACCACCGGCAGCACATCTGCCAGAATCGTGCGGAG AGATGACTGCAGACAGGGCGCTGTGGCTAGCGCCTCTCTGAACCTGATTCGGAGACTGGGCAGAAAGGCCACAGGCAGATCTG GAAGGCGGAGAGTTCTTGGCCGGCCTCCTAGAATGGATGTGCGGAGAAGCGTCCGGATGCGGCGGATGAGAAGATTCCTGTAC AGACTGGCCAGACTCGGCTGGGCCAAGCTGGCTATGTTTGTGCTGGATGGCCAGATGGGCGCCAGCTGTCCTGTTCCTCTGGTG GAAGTGTCCGCCGTGTTTCGCGAGAGATGGTCTATCGTGCGGGCCTTTCTTGGCCTGGGCCAGTTTGGCGGATTTGGCACAGCC GATAATGCCGGCTTCGGCAAGCTGATCGATCCTGCTGAAGTGCGGGCCCATCTGCAGAGCATCAAGAACAGAAGCAGCCCCGG ACCTGACGGCATCACAAAAGTGGCCCTGAGCAAGTGGGACCCCGAGGGCATTAAGCTGGCCCATATGTACAGCACCTGGCTGGT GTCTGCCGGCATTCCCAAGGTGTTCAAGAAGTGCCGGACCACACTGATCCCCAAGACAGGGGATGTTTCCCTGCACGGCGACGT TGGACAATGGCGGCCTATCACAATCGCTAGCCTGGTGCTGAGACTGTACAGCCGGATCCTGACCGAGAGAATGACCGTGGCTTG CCCATCTCACCCCAGACAGAGAGGCTTTATCGCCTCTCCTGGCTGCAGCGAGAACCTGATGCTGCTCGAGGGCTGTATGAGCCTG TCCAAGGCCGGCAATGGATCTCTGGCCGTGGTGTTCGTGGATTTCGCCAAGGCCTTCGACACCGTGTCTCACGAGCATCTGCTGA GCGTGCTGGTGCAGAAGGGACTCGATCAGCACATGGTGGAACTGATCAAGGACAGCTACGAGAACAGCGTGACCAAGGTGCAC TGCCAAGAGGGCTGCAGCACCGATATCGCCATGAAAGTGGGAGTGAAGCAGGGCGATAGCATGAGCCCTCTGCTGTTCAACCT GGCTCTGGACCCTCTGATCCAGCAGCTGGAAAGAGAAGGCAGAGGCTTCCCCGTGAACGGCAAGAGCATTACCGCCATGGCCTT TGCCGATGACCTGGCCATCGTGTCCGATAGCTGGGAGGGCATGAGAGCCAACCTGGATATCCTGGTCGACTTTTGCGAGCTGAC CGGCATGAGAACCCAGCCTTCTAAGTGCCACGGCTTTCTGATCGAGAAGTCCGGCAGCCGGTCCTACAAAGTGAATAGATGCGA GCCCTGGCTGCTGAACGACACAGCCCTCCATATGGTCGGACCCAAAGAGTCCATCAAGTACCTGGGCGTGCAAGTGAACCCCTG GACCGGAATCTTTGCCGAGGACACCGTGGCTAAGCTGAGACAGTGGGTCGTCGCCATCAGCAAGACACCACTGAGGCCCCTGG ATAAGGTGTCCCTGCTGTGCCAGTTTGCCGTGCCTAGAGTGATCTTTGTGGCCGACCACTGCATGCTGAGCGCCAAGGCTCTGAC CGAAATGGACAGATCCATCCGGCAGGCCGTGAAGCGTTGGCTGCATCTGGCTAGGTGTACCACCAACGGCCTGCTGTACTCCAG AAAGTCTAGCGGCGGACTGGGCATCCCAAAGCTGAGCATGATTGTGCCCGCCATGCAGGCTCGTAGACTGCTGGGACTGAGCA GATCCAAGGACGAGACAGTGCGGTGGATGTTCCTGGAAACCACCGACCACGTGGCCTTCGAAAGAGCCTGGCTTAGAGCCGGC GGATCCCCTGATGAAGTGCCTGAACTGGGCCCTGATCTGGTTGAGGGATCTCCTGCCGAGGGCAATGCCGATCCTGTGTCTACC GTCAGACCCCGGAAGCGGATCGTGCCTTGTGATTGGAGACAGGTGGAATTCGACCGCTGGGCCGGACAACTGGTTCAAGGCAA GGGCATCAGAACCTTCGAGGCCGACAAGATCTCCAACTGCTGGCTGTACGACTACCCTCCTAACAAGCTGAAGCCCGGCGATTTC ACAGCCGCTGTGCAGCTGAGAGCTAACGTGTACCCCACAAGAGAGCTGGCCGGCAGAGGCAGAACCGACACAATCGATGTGTG CTGCAGACACTGTGGCGAGGCCCCAGAAACCTGCTGGCATATTCTGGCCCTGTGTCCTAAAGTGAAGCGGTGCCGGATCCAGAG ACACCACAAAGTGTGCCAGGTTCTGGTGGCCGAGGCTGAAAGACACGGCTGGGAAGTCGAGCGCGAGAAGAGATGGATGCTG CCTAGCGGAGAATGCGTGGCCCCTGACCTGATCTGTTGGCTGGATGAGCTGGCCCTGATTGTGGACGTGACCGTCAGATACGAG TTCGACGAGGAAAGCCTGGAACGCGCCAGGATCGAGAAAGAGTGCAAGTACCGGCCTCTGATTCCCGTGATCAGAGCCAGCAG AGTGCAGACCAAGAAAGTGACCGTGTACGGCTTCCCACTGGGAGCCAGAGGAAAGTGGCCTGCCAAGAATGAACTGCTGCTGG CCGATCTGGGCCTGAGCAAAGCCAGAACCAGATCCTTCGCCAAGCTGCTGAGCAGAAGAGTGCTGCTGCACTCCCTGGACGTGA TGCGGACCTTCATGAGATGA (SEQ ID NO: 12)

R2-1_Gav	**5' UTR**	AGGCATCTCCTTKAAGGGTAATGGTCTGGTTACATGGTCATAGCAGGTTTGTGTCAGGTACCTCCCAGTGGTTCCCGCCGGGTGS CAMAGCCCCAGGGCTGTCGGTAGCTCGATCCTGGTACAGTACGGCCAAGGGAGTTCTTCCTTGCTGTCGGGTGCCTCGCAAGCA CKTGGCAGCCCCAATCGCTTCATTGCGAAAAACACAAACGTCCTAAGGGGATGATCAGCTAGTCAGTTCTGCCGCTAGCCAAAAC TGTTTGCCACCCAGTTACAGATAGCGTCTGTGCTGACCAGCTGCCCCGCGGGCTTGGGGTGCAGTGGAGGCCGGCCCGTGGCCA GGCCGGACACGGGCCGTGGAGCCTGCTCCCAGTCCAGAGAGTTCCCCCTCGGAAGCTGCCAGGGCAGCACCAGCCGGGGAGG GCCACGGCCCCGGTCATGAGTCCCCCTCGGTGCAGAGGCCTGAGGCCGATACCACTGCCCCTGGTGTGAGCGCGCCCACCAGGG AGGGTGAACCACCCTCCACCAGGGTTTTC (SEQ ID NO: 13)
	CDS	CTGGTGAGGCTGCCTGATTCAAACCCGCCGTGCCCAATTTGCAGGGACCATGTGGGTAAACCCTCCGCGCTGGCCCTCCACTGCG TGGAGAGCCATGCGTGGGCGGATGTGCAATACCAGTGCACCCATTGCAAAAAGGTCAGTGCTAACAAACACAGCATCCTCTGCC ACATCCCATGTTGCCAGGGGAGGGTGCCCGAGTGGACCGGGAAGGACTGGGCTTGTCCTGAGTGCCCTGCCTCCTTCAATAAGA AAGTTGGCCTATCGCAGCACAAGCGGCACGTGCATCCGGTAACACGTAATGTGGAACGGGTTGCAGGGAGCCTATCGAGGGCT GGTTTAAGGCCCCAGACCAGACGCGGGTGCTGGTCGGTGGAAGAAGAGGAAACTCTCACCTGCCTAGACGCGATGTTCCGTGG TGCCCGGAACATCAACCAGCTGATCGCTGCTGAAATGGTAACGAAGATGCCTAAGCAGATCAGTGACAAACGGAGGCAGCTCG GTCTGTGTCCTGAGCAGACCACATTGGGTGGTGATGCTGAATCGACCTCCGTGGTGGAGGAAGAGTCCATGACTCCGGAGATG GAAACCCAAAGCCCAATTAACCCGCCTGGGAAAATCAGGAAGATCCTGGCCCAGAGGGCACGCCGGTGGCTGAAGAAGGGGC AGGGTCTGTCGGACAAGGTGCGAGAAGTCCTGGGCGCATGGGTGGAGGGTCAACCCAGGATTCATGCCTGGGTCGACTCAGTC TCCCTTGATGTTTTGACTTTGTTCTTGGGGGTGCCCTCAGGACCGCAGAGGGCTCCGAACAAGAAGAGGCCCAAGGAGGGTGGC AAACCAACGTCCTGGATGAACAAATGTGCCGTCAAATGGGGCACATTCCTTCGGTACCAGCACCTGTTTGGTGCCAACAGGAAG CTCCTGGTGGCGATCGTCCTGGACGGCGCTGACCGTAATCAGTGCACCCTCCTGCTAGAGGAGGTCTTCCAGGCCTACCGAGAA AAGTGGGGGCTAGAGGAAGTCCTTCGGGCCTACCGAGGAAAGTGGGAGGTAGAGTCATCTTTTGAGGGTCTCGGACGGTTCGG GGTCCGCCGGGATGCGGATAACTTCGCATTCAAGGCCCTGATCACTCCTGAGGAGGTTGTCAAACACATGATGGCAATGGCCTC GAAATCGGCTCCAGGTCCGGATAAGCTCACCCTGAGAGATCTGCGCCGCGCTGACCCCGAGGGAGATGCTCTTGCCGAACTCTT CAGCCTGTGGCTGATTACCGGCACGGTCCCGGACGGACTCAAGGAGTGTCGGTCTGTGTTGATACCCAAAACGGTGGACCGGG AGAAGTTGGGCCAGCTGGGCAACTGGCGCCCTATCACGATTGGGTCCATCGTTTTACGGCTATTCTCACGAGTGCTAACCGCACG GCTCGCCGCAGCATGTCCCATCAACCCCCGTCAGAGGGGTTTCATAGCGGCGCCGGGGTGCGCCGAGAACCTGAAGGTGCTTGA GCTTCTCTTGCGGAAGAGGAAGCGAGACAGGCAGCCGTTGGGTGTGGTATTTGTGGATCTAGCGAGGGCGTTCGATTCGGTGT CACACGATCACATTTCTTGGGTCCTAAAGGCCAAAGGGGTGGACGAACACATCGTGAATCTCATCGAAGATTCTTACCAGAAGGT TACCACGAGAGTACAAGTGTTCAATGGCGTCACCCCTCCTATCAGCATCAAAACCGGGGTTAAGCAAGGCGACCCGATGTCCCCC CTCTTGTTCAACATTGCGATGGACCCCCTGATAGCGAAGCTCGAGACAGACGGACAGGGAGTAAAAGTCGGGAGTGCCTCCCTG ACCACCCTGGCCTTCGCGGATGATCTCGTCCTGCTTAGCGACTCTTGGGAGGGCATGCTGAAGAACATCAGCATCCTAGAGGACT TCTGCAACCTCACGGGCCTACGAGTGCAACCCAAAAAATGTCAGGGGTTCTTCTTGAATCCGACATGCGACTCCTTTACGGTGAA CAACTGCGAGGCCTGGAAGATAGCCGGCCGTGAGATCACGATGCTCGGACCAGGCGAGTCGACACGATATCTGGGCTTGAATG TCGGTCCTTGGGTTGGGATCGACAAACCAGATTTGGGTACGCAACTAAGCTCCTGGCTCGAGAGGATAGGGACTGCTCCACTCA AACCGATGCAGAAGCTCTCTTTGCTGGTGCAGTATGCCATACCCAGGCTGAACTATCAGGCCGATTACGCGGGCATCGGCAGGG TGGCCTTGGAGGCTCTGGATTCTATGAACCGGAGAAAGGTAAAGGAGTGGTTCCATCTTCCCGCCTGTACCTCGGACGGTCTCCT CCACTCCCGTCACCGTGACGGGGGTCTTGGGTTACCGCGTCTGGCGAAAGCCATTCCGGAAGCGCAAGTGAGGAGGCTGATCC GCGTAGCCACTTCATCTGATGAAGTCACGCGGAAAGTATCCTACGCGTGTGGGATAAGTGACGAAGTGGAGCGGCTCTGGTTG GCGAGGGGTGGGGACATGTCCAGTGTACCGAGGTTCGAGGATCCTGAGGCCCCGAGGTCTCCGGGGGTGCAGGGCCCCTGCG AGGCTGCCCAGGAGATTCCGAGCGTAGTCCGGAAGCTTGCGATCCCCCGGCCCTCCAACTGGAGATCCAAGAAACACTCCAAAT GGGCCCAACTCAGCTGTCAGGGAGAGGGGATGGAGTTATTCTGCAATGATCCAGTCAGCAATGGCTGGAACAACAGTCGGGGA CAACTGGCGGAACACCTCCAGATCGTGGCCTTAAAACTGCGTTCAAACATTTATCCCACCAGGGAGTTTCTTGGGAGAAGCCAG GCAAGTACCAATGTAGGTTGCCGGCATTGTACACACCCTCATGAAACACTAGGGCATATCTTGGGCATATGCCCTGCCGTGCAGG AGGCACGGATCATCCGGCATAACAAGCTGTGCAAGATCCTAGCAGCTGAGGGCAAAAAGTGTGAGTGGACAGTGTATTATGAA CTGCAACTGCTTAACGCTGCAGGGGAACTGTGTAAACCTGACCTCATTTTTGTCCGAGACGGTACCGCTCTGGTTGTGAATGTCA CTGTGGGGTACGAAGGGGGCCCCGCAACCCTCCTATCCACCGCTGCAGAAAAGGCCACAAAATACCTGGATCTGAACGCACAGA TCCAGGAGCTCACAGGGGCTGAGCAGGTCACCTACTTTGGCTTCCCTATTGGAGCCAGGGGAAAGTGGCATGCTGACAACTGGC GAGTACTGTCTGAACTGGGATTGTCCAACTCCCGGAAGGAGCGGGTCGCACGGCTCCTGTCGTGGCGAGCACTGCTCGGGTCAG TGGACATGGTGAACATCTTTGCATCTAAGCACAGGCAGGAAAACCTATCGGATGACGCACTGAGCCCCAGCTGA (SEQ ID NO: 14)
	3' UTR	GAAGTTGCGAGTTCTTATGCAAGTTGAATACCACTCTKGKGACCCCAAAAAAWWAAACCCCAAAACAGTTGTGTTTAAGTGTGT TCTTGTTCGTCCCTTTGGCTTCACCTCSAAGTTGCGATCCCCCCATCTCCCCTGCGCTGCCTTTCAGAACGGCCGGTGGTGTCGAG GCTGGCGCGACCTCGGTCACCTCCAAGGCCAAGTGCCCTGGCCCCGAGTAGGACTGAGTGGCCCAGCTCGCTGGGCACCCGTCA CCATCTGGGGCAAATGGAAGGGATCTGTCCTGACCACTACCAGGCTAAGTGTGGTGCGGCCTAGCCTGCCGTAAGGTCAAGCGC CCTGCTGCCACTCAGGTATCAGTCCTCGTTCACTTGTCCCTCCTAGTACCCTCTGCCTCTGCTCTTTTGCTATCCACTATGGCCAGTG ATGTTGAGGTTGGTGCATCCTTGGTCACCTCCAGGGCCAAGCGCCTTGGCCACAGGTAGGACCTGGCACCTGCCCAGGGGGCCA GACACTGCCTGTGGCAAGGGAAAGGGAGCCGTCCCTGACCGTTACCAGGCTTGAGATGGTGCTGGCTAGCCCACCATATGTCAA GCACTCCACAGCTGCTTGAGTTTGTTGGCTTCACCTTCATCCCACCTAGTGTCTTCTGCCTCTGCACTATTTTCATCCCAACTCGTAC CTCCCCATCTCTGCGCTCCTGCTATCCCTGCAAGGACCAAGTAGGCAGGGGGGTTCATCCCCCTACCTGCAGGAGACTCAGCATA TCCATGACTTCTTGCCTCCACCGTCTTGTGGCGCTAGAGGGGTACCTCAGAGACCGGCACAACATGACCTTGACGGTTAGACAGT AGGGTCAAACAACCCTGCTGCAGGCCCAAAGGGCCAACAGCTGTGCCACGAGAGGGGAACCTTGAAGACTGGGGCAGTCTGAC CATGCTGGTTAGTCAGTTGGGTGAAATAATCCCAGCTGCAGGCCCAAAAGGGCTGACAGTCAGGTGAGGGGGTATCTCCATCTG CTCCCCACTGCCAACTACGGAGGCATGAAGTCCGTAGTGACTTCTGACCCCCACGTCTTGTGCCATGAGAAGGGAACCTTGAAGA TTGGGACAAACCGCACTTGAAAGTTACTCAGCCGGGTGAAAATAAGTCCCAGTTGCGGGCCCCTCGGGGCTGACAGTCAGGTGA GGAGGGCTGCAAAGCCCATCTCCTGACTCCAGAGGCCTGGCGTCCTAACCGACTTCTTGCCACCAATGTCTTGCGCCAGGAGAG GGCAACCTTGAAGATCGGGGCAAGCCGCACTTGATAGTTAGCCAGTCGAGTGAAACAATCTCAGCTGCGGGTCCGAAAGGACT GACTTCCAGGCGAGGGGGGGGCCTGCGGAAAACCCCCTCCATGGTACGGAGGTCTGGCATCCTAACCGACACCTTGCCACCAAT GTCTTGTGCCAGGAGAGGGGAACCTTGAAGACTGGGGCAAGCCGCAGTTGATGGTTAGTCAGTCGGGTGAAATAATCCCGGCT GCACCCTGCTGTGACTGCTAAGCCCGGTCCCCAAGGGGCATGAGGCATGTGCGCTGAGACGGGAGGGGTGACATCTGGCGATC AGCACAGCACAGACTGAAGGGAGGCACTTGCCGAGAATGCTTCTGAGGCCCCAGACTTGGGGTGGTGCAGCTTTGTCTCGTGTA

		TAGTACAGCACCCTACTGCTCCCTTTGGGCAGCAGAATTTGTCCTGACCTCTTACCCACCCGAGTCTGCGCTTTTGTTCCACCTCGC TGTCTCCCTGCTGTGCTGTTTTTCTCTCAAGTGGGTTAAATCTCAACATGATTATCTCCCACGTTTCCGCTCAAGGGCAATGCCCAA CATGACGGAGATCGTTGGTGCATGGTAGTCACGAGACCATCCGGACCCTCCAGTGGTCGCTATAGTCATTTTGTGTTGCATGGG GCATGCTGAGTCACTTAACCGAAAGACTGTAAATAACTCAAAAGAGGTACCCTCCGGGGTTCGGTAAA (SEQ ID NO: 15)
	Codon optimized	CTGGTCAGACTGCCCGACAGCAATCCTCCTTGTCCTATCTGCAGAGATCACGTGGGCAAGCCTAGCGCTCTGGCCCTGCATTGTG TGGAATCTCATGCCTGGGCCGACGTGCAGTACCAGTGTACCCACTGCAAGAAGGTGTCCGCCAACAAGCACTCTATCCTGTGTCA CATCCCCTGCTGCCAGGGCAGAGTGCCTGAATGGACAGGCAAGGATTGGGCCTGTCCTGAGTGCCCTGCCAGCTTCAACAAGAA AGTGGGCCTGAGCCAGCACAAGAGACACGTGCACCCCGTGACCAGAAACGTGGAAAGAGTGGCCGGCAGCCTGTCTAGAGCCG GACTTAGACCTCAGACTCGGAGAGGCTGTTGGAGCGTGGAAGAGGAAGAGACACTGACCTGCCTGGACGCCATGTTCAGAGGC GCCAGAAACATCAACCAGCTGATCGCCGCCGAGATGGTCACCAAGATGCCCAAGCAGATCAGCGACAAGCGGAGACAGCTGGG CCTGTGTCCTGAGCAGACAACACTTGGCGGAGATGCCGAGAGCACCAGCGTTGTCGAAGAGGAATCCATGACACCCGAGATGG AAACACAGAGCCCTATCAACCCTCCTGGCAAGATCCGGAAGATTCTGGCCCAGCGGGCTAGACGGTGGCTGAAGAAAGGACAG GGCCTGTCTGACAAAGTGCGCGAAGTGCTTGGAGCCTGGGTGGAAGGACAGCCTAGAATCCACGCTTGGGTCGACAGCGTGTC CCTGGATGTGCTGACACTGTTTCTGGGCGTGCCAAGCGGACCTCAGAGAGCCCCTAACAAGAAGCGGCCTAAAGAAGGCGGCA AGCCCACCAGCTGGATGAACAAGTGTGCCGTGAAGTGGGGCACCTTCCTGAGATACCAGCACCTGTTCGGCGCCAACCGGAAAC TGCTGGTGGCCATTGTGCTGGACGGCGCCGATAGAAATCAGTGCACCCTGCTGCTGGAAGAGGTGTTCCAGGCCTACAGAGAG AAGTGGGGACTCGAGGAAGTGCTGCGGGCCTATAGAGGCAAGTGGGAAGTCGAGAGCAGCTTCGAAGGCCTGGGCAGATTTG GCGTGCGGAGGGACGCCGATAACTTCGCCTTTAAGGCCCTGATCACCCCTGAAGAGGTGGTCAAGCACATGATGGCCATGGCTA GCAAGAGCGCCCCTGGACCTGATAAGCTGACCCTGAGAGATCTGCGGAGAGCCGATCCTGAAGGGGATGCTCTGGCCGAGCTG TTTAGCCTGTGGCTGATCACAGGCACCGTGCCAGACGGCCTGAAAGAATGCAGAAGCGTGCTGATCCCCAAGACCGTGGATCGG GAAAAGCTGGGGCAGCTCGGAAATTGGAGGCCCATCACAATCGGCAGCATCGTGCTGCGGCTGTTCAGCAGAGTGCTGACAGC TAGACTGGCCGCTGCCTGTCCTATCAATCCCCGGCAGAGAGGCTTTATCGCCGCTCCTGGCTGTGCCGAGAATCTGAAGGTTCTG GAACTGCTGCTGCGGAAGCGGAAGAGGGATAGACAGCCTCTGGGCGTCGTGTTCGTGGATCTGGCTAGAGCCTTCGACTCCGT GTCTCACGACCACATCAGCTGGGTGCTGAAGGCCAAAGGCGTGGACGAGCACATCGTGAACCTGATCGAGGACAGCTACCAGA AAGTGACCACCAGAGTCCAGGTGTTCAACGGCGTGACCCCTCCTATCAGCATCAAGACCGGCGTGAAGCAGGGCGACCCTATGA GCCCTCTGCTGTTCAATATCGCCATGGATCCTCTGATCGCCAAGCTGGAAACAGACGGCCAGGGCGTGAAAGTGGGATCTGCCT CTCTGACCACACTGGCCTTCGCCGATGATCTGGTGCTGCTGAGCGATAGCTGGGAGGGCATGCTGAAGAACATCAGCATCCTGG AAGATTTCTGCAACCTGACCGGCCTGAGAGTGCAGCCCAAGAAGTGCCAGGGCTTCTTCCTGAATCCTACCTGCGACAGCTTCAC CGTGAACAACTGCGAGGCTTGGAAGATCGCCGGCAGGGAAATCACAATGCTCGGCCCTGGCGAGTCCACCAGATACCTGGGAC TGAATGTCGGCCCCTGGGTCGGAATCGACAAGCCTGATCTGGGCACACAGCTGAGCAGCTGGCTGGAAAGAATCGGCACTGCC CCTCTGAAGCCCATGCAGAAACTGAGCCTGCTGGTGCAGTACGCCATTCCTCGGCTGAACTACCAGGCCGATTATGCCGGAATTG GCAGAGTGGCCCTGGAAGCTCTGGACAGCATGAACCGGCGGAAAGTCAAAGAGTGGTTCCATCTGCCTGCCTGCACCTCCGATG GCCTGCTGCATAGCAGACACAGAGATGGCGGACTGGGCCTGCCTAGGCTGGCTAAAGCTATCCCTGAAGCTCAAGTGCGGAGA CTGATTAGAGTGGCCACCTCCAGCGACGAAGTGACCCGGAAAGTGTCTTACGCCTGCGGCATCAGCGACGAGGTGGAAAGACT GTGGCTTGCCAGAGGCGGCGATATGTCTAGTGTGCCTAGATTCGAGGACCCCGAGGCTCCTAGATCTCCTGGTGTTCAGGGCCC TTGCGAGGCCGCTCAAGAGATCCCTTCTGTTGTGCGGAAGCTGGCTATCCCCAGACCTAGCAATTGGCGGAGCAAGAAACACAG CAAGTGGGCACAGCTGTCCTGCCAAGGCGAAGGCATGGAACTGTTCTGCAACGACCCCGTGTCCAACGGCTGGAACAATAGCA GAGGACAACTGGCCGAACACCTCCAGATCGTGGCCCTGAAACTGCGGAGCAACATCTACCCTACCAGAGAGTTCCTGGGGAGAA GCCAGGCCTCTACCAATGTGGGCTGCAGACACTGCACACACCCTCACGAAACACTGGGCCACATCCTGGGCATTTGCCCTGCCGT GCAAGAGGCCAGGATCATCAGACACAACAAGCTGTGCAAAATCCTGGCCGCCGAGGGCAAGAAATGCGAGTGGACAGTGTACT ACGAACTGCAGCTGCTGAACGCAGCCGGCGAACTGTGCAAGCCCGATCTGATCTTCGTTAGAGATGGCACAGCCCTGGTCGTGA ACGTGACCGTGGGATATGAAGGCGGACCTGCCACACTGCTGTCTACCGCCGCTGAGAAGGCCACAAAGTACCTGGATCTGAACG CCCAGATCCAAGAGCTGACAGGCGCCGAGCAAGTGACCTACTTCGGCTTTCCTATCGGAGCCAGAGGAAAGTGGCACGCCGACA ATTGGAGAGTGCTGTCTGAGCTGGGGCTGAGCAACAGCCGGAAAGAGAGAGTTGCCAGACTGCTGAGTTGGAGAGCCCTGCTG GGATCCGTGGACATGGTCAACATCTTCGCCAGCAAGCACCGGCAAGAGAACCTGAGCGACGATGCCCTGTCTCCTAGCTGA (SEQ ID NO: 16)

R2-1_GFo	**5' UTR**	AGACTTAAGTGAGTTTGGTTACAACTGGGCATAGCTGCAGAGACCGCGCCTCCTCGCGGCCCCGCTGGTAAGCCCTTAACAGGG TGACTAAGTCGGTCTCTGCCCCAGTCCGGGAGTCGATGGGACTCACCAGCCCAACGATTCCTTCCAAAATTTCGGTGAAACAAAT TTCTCGGTGCAAGTCGCAAGGCTTGTCACCCGAAACCTAGCCCCCCGGTCGGTCAGGGGCAACGGGTTCGGAAGTGGG (SEQ ID NO: 17)
	CDS	ATGGCCACCCACCCCGTTCCCGCAGACGAATCCGGCCATGAATCTGATCCATTCCTTGTAGGGAGGAGCTGCGGACAACCGGCA CGCCTTACTAGGCAATCGGTTGGCACCCAGACCTCCCGAGATGATATTTTACCATCTAAAACCACCAAATTGACAGAGAATGAAT TGGACTTGCTGGTGAACTTTTCTTTAGAATTGTATAGGTCAGATCTGCAGGGATTTGTGCAGGAGGGGATTCATTTTTCTGTGAA TAGGGAGGTGTTAGAGGGGTTTCCTGAGGTGTATGAACAACCTGCACCACAACCGGCAGTAGGGGACGATTTAAACACCAGTCT CCCACCGGACAATAATATATGCGTACTTGAGAAGGGTAGCAGTGAAGCAGTGGAGGATGGCACACCGGAGGTAGCGCACCCCG TGCCTGAAACCCAGGGCAAAGAGTCACCGAATAACATCGTGATGGTAACTCTTCCCAACAAAAATCCACCATGTCCTTGCTGTAG GGTCAGACTGCATTCAGTACTGGCTCTGATTGAACATCTTAAGGGGTCGCATGGGAAGAAGAGGGCATGCTTTAGGTGTGTCAA GTGTGGGAGGGAGAACTTTAACTATCATAGTACTGTTTGTCACATCGCAAAATGCAAGGGACCAAAAGTTGAGAAGGCCCCAGT GGGAGAGTGGATCTGTGAGGTATGTGGTAGGGACTTTACAACCAAAATCGGCCTGGGACAACATAAAAGATTGGCACATCCCTT GGTTAGAAACCAAGAAAGGATCGATGCTTCCCAACCGAAGGAGACATCAAACAGAGGAGCCCACAAGAGATGTTGGACAAAAG AGGAGGAGGAGATGCTGATAAAGTTGGAGGTACAGTTCGAGGGACACAGAAACATCAATAAGCTTATCGCGGAACACTTAACA ACTAAAACATCCAAACAGATTAGTGATAAAAGGAGACTATTACCCAGAAAACAATTAACAGATCTAAGTAAGGGAGTGGCTGGA CAGAAGGTGCTGGACCCAGGACTGAGTCATCAACCCCAGCTGGGGGTAGTTGACAATGGACTTGGTGGGGGTCATCTGCCAGG GGGGCCAGCTGCTGAAGGAAGAACAATAGAGCCATTAGGACACCACCTTGATAAGGATAACGGTCACCGGGAAATCGCTGACC AGCACAAGGCAGGGAGGCTGCAGGCCCATTACCGAAAGAAGATAAGGAAGCGCCTTTCAGAAGGGATGATTAGCAACTTCCCC GAAGTATTTGAACAACTACTGGACTGCCAGGAAGCACAACCATTGATCAATCAAGCAGCGCAGGATTGCTTTGGATGCCTGGAT TCAGCAAGCCAGATAAGGAAGGCGCTCCGAAAACAGAACACACAGAAAGACCAGGGGGATCAACCCAAAAGACCAGCTCAGA AGTGGATGAAAAAAAGAGCAGTTAAGAGGGGTCACTTCCTCCGCTTTCAGAAATTATTTCATCTTGACAGGGGGAAATTGGCAA AGATTATTTTGGACGACGTAGAGTGTTTGTCCTGTGATATACCACCCAGTGAAATTTATTCGGTATTCAAAGCCCGATGGGAAAC

		ACCTGGACAGTTTGCTGGCCTTGGGGATTTCGAAATTAATAGGAAGGCGAACAATAAAGCCTTCAGGGACTTAATTACGGCCAA AGAAATTCTCAAAAATGTGCGGGAGATGACCAAGGGCTCGGCCCCAGGTCCAGATGGGATCGCGCTTGGGGACATCAGGAAGA TGGACCCTGAGTACACCCGGACCGCCGAACTCTTCAACTTATGGTTAACATCTGGTGAGATCCCGGACATGGTGAGGGGGTGCA GAACTGTGTTAATCCCCAAATCGTCAAAACCGGAACGCCTGAAGGACATCAATAACTGGAGACCCATCACGATTGGATCCATCTT GCTGAGACTTTTCTCCAGGATCATAACAGCGAGGTTAACAAAGGCGTGCCCCCTCAACCCTAGGCAAAGAAGCTTCATCAGTGCG GCAGGATGCTCCGAGAACTTGAAGCTCCTGCAAACCATAATTCGGACTGCTAAAAATGAACACAGACCACTGGGTGTTGTATTCG TGGACATCGCCAAGGCCTTTGACACCGTGAGCCACCAACACATCATACATGTATTGCAAAGGAGGAGAGTGGACCCCCACATCA TTGGATTGGTGAAAAATATGTACAAAGACATCAGTACGGTTATCACCACAAAGAAGAACACATACACGGACAAAATCCAGATCC AGGTTGGAGTGAAGCAAGGTGATCCGCTTTCGCCCCTTCTATTCAACCTGGCGATGGACCCCCTGTTGTGCAAGCTGGAAGAAC ACGGCAAAGGATTCCACCGAGGACAGAGCAAGATAACAGCGATGGCATTCGCTGATGACCTGGTCCTGTTGAGCGATTCCTGG GAAGACATGAATGCGAACATCAAGATACTGGAGACCTTCTGCGACCTCACCGGTCTCAAAACACAGGGTCAAAAGTGCCACGGC TTCTACATCAAGCCTACAAAGGACTCTTACACCGTCAACAACTGCGCTGCGTGGACCATCAATGGCACACCCCTGAACATGATCA ACCCCGGGGAATCAGAGAAATACCTCGGCCTGCAGTTTGACCCCTGGGTGGGAATTGCAAAGACCAGCCTCCCCGAAAAACTGG ACTTCTGGCTCGAACGCATTGATCGAGCTCCACTCAAACCATTTCAGAAACTGGACATTCTTAAGACATACACCATACCTCGACTG ACCTACGTAGCTGACCACTCAGAGATGAAAGCGGGGGCCCTTGAAGCCCTTGACCGGACAATTCGATCGGCGGTCAAGGACTG GCTGCACCTACCTTCGAGCACCTGTGATGCCATCTTGTACACGAGCATGAAGGACGGTGGTTTGGGAGTGACCAAATTGGTGGG ACTGATTCCGAGTGTACAAGCCCGGAGGCTGCACAGGATTGCGCAGTCACCGGAGGAGACGATGAAAGACTTCCTGGAAAAGG CCCAGATGGAGAAGATGTACGAGAAATTGTGGGTCCAAGCTGGAGGGAAAAGAAAGAGGATGCCGTCAATTTGGGAAGCGCT CCCGGAGGTTGTACCATCCATAGACACAGCCACAACTTCGGAGTGGGAAGCACCGAACCCTAAAAGTAAGTACCCTAGACCTTG TAATTGGCGCAGAAAAGAATTTAAAAAGTGGACTAAATTAATAGCCCAGGGCTGGGGAATTAGGTGTTTTAAGGGGGACAAAA TTAGTAACAATTGGATTCGACATTATAGATACATACCTCACAGGAAACTTCTCACTGCCATACAGCTCCGGGCCAGTGTGTACCCC ACAAGGGAATTTCTCGCGCGGGGGAGGGAAGATAACTGTGTTAAGTCTTGTAGGCACTGTGAGGCGGCAGAGGAGTCCTGTGC CCACATCATCGGCATGTGTCCAGTCGTGAGGGATGCCCGAATCAAGAGGCACAATCGCATTTGCGAGAGGCTGATGGAGGAGG CGGGGAAGAGGGACTGGACGGTGTTTCAGGAGCCGCACATAAGGGACGTCACCAAGGAACTGTACAAACCGGACTTGATATTC GTGAAAGAAGGCCTTGCACTTGTTGTGGATGTTACAATACGGTTCGAGTCAACCAAGACAACGTTGGAGGAGGCTGCTGCAGA GAAGGTGAACAAGTACAAACATCTGGAGACCGAAGTACGGAACCTCACCAACGCTAAGGACGTTATCTTTATGGGGTTTCCCCTT GGAGCGCGGGGACAATGGTACAATAAGAACTTTGAACTTTTGGACACTCTTGGCCTCCCCAGATCGAGGCAGGACATTATTGCA AAGACTTTATCCACGGACGCGCTCATTTCATCTGTGGACATTATACATATGTTTGCCAGTAGAGGCAGAAGACAGCATGCTTAG (SEQ ID NO: 18)
	3' UTR	GGTAGATAATCTTTGTATAGTGGGGGGGGATCTCATGTACCGGGTTTCTTTTATTTGATTTTCAATAAAACAGACGGTAGCTAGG TTCGCAAGGCAGCCACAAGCCAAAGATAGGTAGGGTGCTCATAGTGAGTAGGGACAGTGCCTTTTGATTCACAACGCGTCAATA CCATCTGACACGGATACCCTTACCGGACTTGTCATGATCTCCCAGACTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGA AAAGGAACATATATTAATTATATGTGTTCGGAAAA (SEQ ID NO: 19)
	Codon optimized	ATGGCCACACATCCTGTGCCTGCCGATGAGTCTGGCCACGAGAGCGATCCATTTCTCGTGGGCAGAAGCTGTGGCCAGCCTGCC AGACTGACAAGACAGTCTGTGGGCACCCAGACCAGCCGGGATGATATCCTGCCTAGCAAGACCACCAAGCTGACCGAGAACGA GCTGGACCTGCTGGTCAACTTCAGCCTGGAACTGTACAGAAGCGACCTGCAGGGCTTCGTGCAAGAGGGCATCCACTTCAGCGT GAACAGAGAGGTGCTGGAAGGCTTCCCCGAGGTGTACGAACAGCCTGCTCCTCAACCTGCCGTGGGCGACGATCTGAATACCTC TCTGCCTCCTGACAACAACATCTGCGTGCTGGAAAAGGGCAGCAGCGAGGCCGTGGAAGATGGAACACCTGAAGTGGCCCATC CAGTGCCTGAGACACAGGGCAAAGAGTCCCCAAACAACATCGTGATGGTCACCCTGCCAAACAAGAACCCTCCTTGTCCTTGCTG CAGAGTGCGGCTGCATTCTGTGCTGGCCCTGATCGAGCACCTGAAGGGCTCTCACGGAAAGAAGCGGGCCTGCTTCAGATGCGT GAAGTGCGGCAGAGAGAACTTCAACTACCACAGCACCGTGTGCCACATTGCCAAGTGCAAGGGCCCCAAGGTGGAAAAGGCCC CTGTTGGAGAGTGGATCTGCGAAGTGTGCGGCAGGGACTTCACCACCAAGATCGGACTGGGACAGCACAAGAGACTGGCTCAT CCTCTCGTGCGGAATCAAGAGCGGATCGATGCCAGCCAGCCTAAAGAGACAAGCAACAGAGGCGCCCACAAGCGGTGCTGGAC CAAAGAAGAGGAAGAGATGCTGATCAAGCTCGAGGTGCAGTTCGAGGGCCACAGAAACATCAACAAGCTGATCGCCGAGCATC TGACCACAAAGACCAGCAAGCAGATCAGCGACAAGCGGAGACTGCTGCCCAGAAAGCAGCTGACAGACCTGTCTAAAGGCGTG GCCGGACAGAAAGTGCTGGATCCCGGACTGTCTCACCAGCCTCAGCTGGGCGTTGTGGATAATGGACTCGGCGGAGGACATCT GCCTGGTGGACCAGCTGCTGAGGGCAGAACAATTGAGCCTCTGGGCCACCACCTGGACAAGGACAATGGCCACAGAGAGATCG CCGACCAGCACAAAGCTGGCAGACTGCAGGCCCACTACCGGAAGAAGATCCGGAAGAGACTGAGCGAGGGCATGATCAGCAA CTTTCCAGAGGTGTTCGAGCAGCTCCTGGACTGCCAAGAAGCCCAGCCTCTGATCAATCAGGCCGCTCAGGATTGCTTCGGCTGC CTGGATTCTGCCTCTCAGATCAGAAAGGCCCTGCGGAAGCAGAACACCCAGAAGGATCAGGGCGACCAGCCTAAGAGGCCCGC TCAAAAGTGGATGAAGAAACGGGCCGTGAAGAGAGGCCACTTCCTGCGGTTCCAGAAACTGTTCCACCTGGATAGAGGCAAGC TGGCCAAGATCATCCTGGACGACGTGGAATGCCTGAGCTGCGACATTCCTCCTAGCGAGATCTACAGCGTGTTCAAGGCCAGAT GGGAGACACCTGGCCAGTTTGCTGGCCTGGGCGACTTCGAGATCAACCGGAAGGCCAACAACAAGGCCTTCAGGGACCTGATC ACCGCCAAAGAAATCCTGAAGAACGTGCGCGAGATGACCAAGGGCTCTGCCCCTGGACCTGATGGAATTGCCCTGGGCGATATC AGAAAGATGGACCCCGAGTACACCCGGACCGCCGAGCTGTTTAATCTGTGGCTGACAAGCGGCGAGATCCCCGACATGGTTCGA GGCTGTAGAACCGTGCTGATCCCCAAGAGCAGCAAGCCCGAGAGACTGAAGGATATCAACAACTGGCGGCCCATCACCATCGG CAGCATCCTGCTGAGACTGTTCAGCCGGATCATCACAGCCCGGCTGACAAAGGCCTGTCCTCTGAACCCTCGGCAGCGGAGCTTT ATTTCTGCCGCCGGATGCAGCGAGAACCTGAAGCTGCTGCAGACCATCATCAGAACCGCCAAGAACGAGCACAGACCCCTGGGC GTCGTGTTCGTGGATATCGCCAAGGCCTTTGACACAGTGTCCCACCAGCACATCATCCACGTCCTGCAGCGGAGAAGAGTGGAC CCTCACATCATCGGCCTGGTCAAGAACATGTACAAGGACATCTCCACCGTGATCACCACGAAGAAGAACACCTACACCGACAAG ATCCAGATCCAAGTGGGCGTGAAGCAGGGCGATCCTCTCAGCCCTCTGCTGTTTAACCTGGCCATGGATCCACTGCTGTGCAAGC TGGAAGAACACGGCAAGGGCTTCCACAGAGGCCAGAGCAAGATTACCGCCATGGCCTTCGCTGACGACCTGGTGCTGCTGAGC GATAGCTGGGAAGATATGAACGCCAACATCAAGATCCTGGAAACCTTCTGCGACCTGACAGGCCTGAAAACCCAGGGCCAGAA GTGCCACGGCTTCTACATCAAGCCCACCAAGGACTCCTACACCGTGAACAATTGTGCCGCCTGGACCATCAACGGCACCCCTCTG AACATGATCAACCCCGGCGAGAGCGAGAAGTACCTGGGACTGCAGTTTGACCCTTGGGTCGGAATCGCCAAGACAAGCCTGCCT GAGAAGCTGGACTTCTGGCTGGAACGGATCGACAGAGCCCCACTGAAGCCCTTTCAGAAACTGGACATCCTCAAGACCTACACA ATCCCCAGGCTGACCTACGTGGCCGACCACTCTGAAATGAAGGCAGGCGCTCTGGAAGCTCTGGACCGGACAATTAGAAGCGCC GTGAAGGACTGGCTGCATCTGCCAAGCAGCACCTGTGACGCCATCCTGTACACCAGCATGAAGGATGGTGGCCTGGGAGTGAC CAAGCTCGTGGGACTGATTCCTAGCGTGCAGGCCAGACGGCTGCACAGAATTGCTCAGAGCCCCGAGGAAACCATGAAGGACTT TCTGGAAAAAGCCCAGATGGAAAAGATGTACGAGAAACTGTGGGTGCAAGCCGGCGGAAAGCGGAAGAGAATGCCCAGCATT TGGGAAGCACTGCCCGAGGTGGTGCCTAGCATCGATACAGCCACAACCAGCGAGTGGGAAGCCCCTAATCCTAAGAGCAAGTA

		CCCCAGACCTTGCAATTGGCGCCGGAAAGAATTCAAGAAGTGGACCAAACTGATCGCCCAAGGCTGGGGCATCAGATGCTTCAA GGGCGATAAGATCTCCAACAATTGGATCCGGCACTACCGGTACATCCCTCACCGGAAACTGCTGACCGCCATCCAGCTGAGAGC CAGCGTGTACCCCACCAGAGAGTTTCTGGCTCGGGGCAGAGAAGATAACTGTGTGAAGTCCTGCAGACACTGCGAGGCCGCCG AAGAATCTTGCGCCCACATCATTGGCATGTGCCCCGTCGTCAGAGATGCCCGGATCAAGCGGCACAACAGAATCTGCGAGCGGC TGATGGAAGAGGCCGGCAAGAGAGACTGGACCGTCTTTCAAGAGCCTCACATCCGGGACGTGACCAAAGAGCTGTACAAGCCC GACCTGATCTTCGTGAAAGAAGGCCTGGCTCTGGTCGTGGACGTGACAATCAGATTCGAGAGCACCAAGACCACACTGGAAGAA GCCGCCGCTGAGAAAGTGAACAAGTACAAGCACCTGGAAACGGAAGTGCGGAACCTGACCAACGCCAAGGACGTGATCTTCAT GGGATTCCCTCTGGGCGCTAGAGGCCAGTGGTACAACAAGAACTTCGAGCTGCTGGACACCCTGGGCCTGCCTAGATCCAGACA GGACATCATTGCTAAGACCCTGTCCACAGATGCCCTGATCTCCAGCGTGGACATTATTCACATGTTCGCCAGCAGAGGCAGACGG CAGCATGCTTAA (SEQ ID NO: 20)
R2-1_IS	**5' UTR**	GTTCCAAAGGAAGGCACTCCTTTGGTTCGTGATGAGATGTTCATGGTGCTTGCCTAGCTGGAGAAATCCGACTCACACCTGCACG TGGTCCCTGCCGCCTGCCAGTATGCCGAGGAAACGGGTGCAACTTAATCCGTGGATACTGGTAGCAACGTGAGCAACGGTACGG TCCTTCGCGGACCACCCTGGGCGTTCGGGTTGCCAGCCCGTTCGCCCGAAATATCTTGGCCCTGAAACTAAAAGAAAA (SEQ ID NO: 21)
	CDS	ATGCAATGCACCAGCCGACTGGCTGATGCACCAAGATTTGCCCGAGTGGGCGTCGAGGGTGAAGGTGTCGGTGCGTCTGGTAA CGGCACTGATGCGCAGTTATGGTATGGCTGCACGGGCTGTGACGAAGCCTTCTCGTCCCTGCGAGGACTGAGAATTCATGCGGC CCAAAAGAAACATGGAAACCAAGATGGCCTTCTCCGCCTGCCGGCGGGACGGCCCCGAAAACGACGAGTGGGGAAGAGCACCA CAGCGGGTGCTTCGGACCGGGTGACCACGGATCCAGTGCCTGCTCCAGTTCCTGAATCTCCTGGGCTGCTGCCTGGGCTACCTG GACCATCGCTGCCTGGGTGCTCGGACCTGCCGCCTGGGGTGCTGCCTGGAGGGTGGTCTGCATCACCTGGGCCTCTCTCCTGGC CTCCTTCCCTGGATGCCGGGCCTCTGCCTGGACCTTCAAGAGTATCACCTGGACCTTCAAGACCTTCGCCCGGGAAGCCGACCGG GCCTCCTTCCCTGGATGCCGGGCCTCTGCCTGGACCTTCAAGAGTATCACCTGGACCTTCAAGACCTTCGCCCGGGAAGCCGCCC GGGACGCCTGAGCCGCTGCCTGGATCTCCTGGCGGTCGGCGCGGGGTGTCCCCGGGACAGCCCGGGTCACGGACCGACCCCTC AAGCTCTGCTGGCGCCGGACACTTCGTATGCCCGCAGTGCAGCAGAGCCTTCTCAAGCAAGATTGGCATGTCTCAACACCAAAAA CATGCCCACCTCGAAGAATACAACGCGGGCATTAACATCACCCGTACCAAAGCCCGGTGGGACCCCGAGGAGACCTATCTTCTG GCCCGCCTGGAGGCCACCCTCAACCCAGACCACAAGAACATCAACCAGACGCTGCACGCCGCGCTGCCCCGCGGTTCCTGTCGA ACCCTGGAGAGCATCAAGGCCCACCGAAAGCAGGCGGCTTACAGGGACCTGGTGACGAGCCTGCGGTCAGCCAGGGAGAGCA GCGAGGCGCAGCACGTTCCGGACCGGCCCCTGGAGACCCCGGAGCCCCAGACACCAGCGAACCCTCAAAGAGACTCGAAGCAG GCAGTCATCGAAGCGCTGCAATCCCTCATCGGCCGAGCACCACCAGGCTCCTTCCAGGGAGCGCGTCTCTGGGACATCGCGAGG CAAGCCACAAGGGGGACGAACATCCTCCCACTCCTGAACAGCTACCTGAGGGATGTTTTCACCCTCCCCACAAAGCCAACAAGAA AGAAACCTGCAGTGCGGCCCGCCCGGAGCCGCAGAAAACAAAAGAAACAAGAGTATGCCAGAACACAAGATCTATTCAGGAAA AAGCAGTCCGACTGTGCCAGGGCGGTCCTGGACGGCCCCACGTCGTCATCGGTCCCTGGAACGGGCGCCTTCCTGCAAACCTGG CGAGAGATCATGACGGGGCCCAGCCCTGCACTCGAGGCACCGCCTCTACCTACCCGGGGGGAAGTCGACCTGTTCTTCCCGGCG ACGGCGCAAGAGATCCAGAGCGCTGAGATAGCCGTCAACTCGGCTGCTGGACCCGACGGGTTCTCAGCCCGTCTCCTCAAGTCC GTCCCGGCCCTCCTCCTAAGGGTCATGGTTAACCTTCTGCTCCTCGTCCGACGTGTCCCGGCGGCCCTCCGGGACGCGAGAACGA CCTTCATCCCGAAGGTCCCCGATGCAGTGGACCCCTCCCAATTTCGCCCAATAACGGTGGCCTCCGTTCTCCAGCGCCTACTACAT CGCATCCTGGCCAAGAGGGCGCTGGAGGCCATTCCCCTCAACTTTCGACAAAGAGCCTTTCAGCCGGTGGATGGCTGTGCCGAG AATATATGGCTGCTGTCCACCGCGCTCAACGAGGCAAGAACCAGACGGCGCCCGCTACACATGGCGAGCGTCGACCTAACCAAG GCATTCGACCGGGTCACCACGGATGCCATCCTGAGGGGCGCAAGGCGCGCCGGGCTGTCCGGGGAGTTCATCGGATACCTGAA GGAGCTCTACACAACATCCAGGACCCTCCTGCAGTTCCAGGGAGAGAGCCTGCTTGTCGAACCCACGACCGGCGTGCGACAGGG CGACCCACTGTCGCCCATCCTCTTCAACCTGGTCCTGGACGAGTACCTCTCCTCCCTGGACCCGGACATCTCCTTCGTCTCGGGCG ACTTGCGCCTCGATGCGATGGCATTCGCTGACGACTTGATCGTCTTTGCCTCAACCCCAGCCGGCCTGCAGGATCGGCTCGATGC CCTCGTCGAGTTCTTCGACCCAAGGGGGCTCAGGGTGAACGTGAAGAAGAGCTTCACGCTATCGCTGCAGCCGGGACGAGACA AGAAGGTCAAGGTGGTGTGTGACCAGATCTTCACCATCGGAGGAACCCCACTCCCAGCCTCCAAGGTCGCAACCCCTTGGCGCT ACCTGGGGATGACCTTCACCCCCCAGGGCTCAATCAACAAGGGCACCAGCGAGCAGTTGGACCTACTGCTCACGAGAACCAGTA AGGCCCCCCTCAAGCCACAACAGAGGCTGGTGGTCTTAAGAAACTACCTGCTCCCGAGGCTATACCATCGCCTCGTGCTTGGACC TTGGTCGGCCGCCCTCCTACTGAAGATGGACACCACCATTCGAGGAGCCATTAGACGCTGGATGGATCTCCCGCACGACACACC GCTGGGTTTCTTCCACGCCCCAGTAACGGAGGGAGGCCTAGGAATCAACTCCCTGCGAGCATCAATTCCAGCCATGGTGCTCCAA CGGCTGGATGGACTTCACTTCAGCACGCATCCCGGAGCTGAGGTCGCCATCCAGCTGCCGTTCCTGACAGGACTCCATCGAAGA GCGGAAGCGGCGGCCCAATACCAGGGACAGAGACTACTGTCCAAAGCGGACGTCCACCGGATGTGGAGCGCAAGACTCCACG GGAGCTGCGACGGAAGACCCCTTCGGGAGTCCAAGAGAGTGCCGGCTGCCCATCGTTGGGCCGCGGAAGGCACCAGACTACTC TCGGGAAGGGACTTCATCTCGATCACGAAACTCAAGATAAACGCGCTACCTACACTCGAGCGCACCAGCCGGGGCCAGCACAAG GACATCCAGTGCAGAGCTGGCTGCCAGGCTGTTGAATCCCTGGGCCACGTCCTACAAGCTTGCCATCGAGGACACCGTGGCCGA ATCCGGCGGCATGATAACATTGCCCGCTACGTCTGCGGCCGACTGACCCAGATTGGCTGGGCGGTGAAGTGGGAGCCCCACTAC TCTGTCGCTGGAAGGACCCTCAAACCTGACATCGTTGCCCATCGTGGAGCCGAGACTGTCGTGCTCGACGCCCAGGTCGTCGGC ACCAGCATGCGACTGGGCTTCCACCACGCTCAAAAGAAAGAAAAGTACTCTCTCCCAGACCTCCTCCACCAAGTCTGCGAGGGAC GGAGAGACGCAGCCCGGGTGTCAACAATCACCCTCAATTTTCGAGGTGTTTGGGCACCTGAGAGCGCCCAGGACCTGAAGTCCC TGGGCCTGACGGACAACGACCTAAAGCTTCTCACCGTCCGCTGCCTCCAGGGCGGCGCGCAGTGTTTCCGGCTGCACCGCCGAA TGACCACCGTGGTGAAGGCCACGGGCGATGAAGCCAACGCCCTCCCCGCCCATTCGGGCTTGCCGCCAACACAGCTTGGTGGCC GAACCCTGGGTCCCTCTGCCCACAATCAGAGTGCAAGAACTACTTAG (SEQ ID NO: 22)
	3' UTR	TGTGACGGAGTCCTCAAGCCCCCACAAGTGCCTGCCAGGTGGCAGGAAAGGGCAACTACTGGTGAGCGACCCAAGCAAGGCGG AGCCAAGACCAAGCTGGAGCCAAGAGCAACTCCAGGAGGCAGGGGTGGATATCAAGAGCAACCCCAAGGGACACAGACCACG GGCAACTACTGGTGAGCGCCCAAGACAGGGGTGGATATTAAGAACAGCCCCACAAAGTGTTACCTATATTAACAATAAAGTTGA AGCCTCAACCACGCATTGCGGGTTAGATGGCGTGGCTTGGCCCGCCGCCATGATGAGCTGGAACCCTCCACCTGGTGGGCCGCA CGAGACCACCGGCTCTTTCTACTAAGGCCGGTCTCCGTGACTGCGGTTGGGATAAACTCCAAGCACTGAGCGGTAAAAAAAAAA AAAAAAAAAAAAAAAAA (SEQ ID NO: 23)
	Codon optimized	ATGCAGTGCACCAGCAGACTGGCCGACGCTCCTAGATTTGCCAGAGTGGGCGTTGAAGGCGAAGGCGTGGGAGCCTCTGGCAA TGGAACAGATGCCCAGCTTTGGTACGGCTGCACCGGATGTGATGAGGCCTTCAGCTCTCTGCGGGGCCTGAGAATTCACGCCGC TCAGAAGAAGCACGGCAACCAGGACGGACTGCTGAGACTTCCTGCTGGCAGACCCAGAAAGCGGAGAGTGGGCAAGTCTACAA CAGCCGGCGCTAGCGACAGAGTGACCACAGATCCTGTGCCTGCTCCTGTGCCAGAGTCTCCTGGACTGCTTCCAGGACTGCCTG

		GACCTTCTCTGCCTGGCTGTTCTGATCTGCCTCCTGGTGTTCTTCCTGGCGGCTGGAGTGCTTCTCCAGGACCACTTTCTTGGCCTC CAAGCCTGGATGCTGGACCTCTGCCAGGACCTAGCAGAGTTAGCCCTGGACCAAGCAGACCCTCTCCAGGCAAACCTACAGGCC CACCATCTCTGGATGCAGGTCCACTTCCTGGGCCTTCCAGAGTTTCCCCTGGGCCATCTAGGCCAAGTCCTGGCAAACCTCCTGGC ACACCTGAACCACTGCCAGGATCTCCAGGTGGCAGAAGAGGTGTTAGCCCAGGCCAGCCTGGCAGCAGAACAGATCCTAGTTCT TCTGCCGGCGCTGGCCACTTCGTGTGTCCTCAGTGTAGCAGAGCCTTCAGCAGCAAGATCGGCATGAGCCAGCACCAGAAACAC GCCCACCTGGAAGAGTACAACGCCGGCATCAACATCACCCGGACCAAGGCCAGATGGGACCCCGAGGAAACATACCTGCTGGC TAGGCTGGAAGCCACACTGAACCCCGACCACAAGAACATCAATCAGACCCTGCACGCTGCCCTGCCTAGAGGCTCTTGTAGAACC CTGGAATCCATCAAGGCCCACAGAAAGCAGGCCGCCTACAGAGATCTGGTCACCAGCCTGAGAAGCGCCAGAGAGTCTAGCGA AGCCCAGCACGTGCCAGACAGACCTCTGGAAACACCCGAGCCTCAGACACCCGCCAATCCTCAGAGAGATAGCAAACAGGCCGT GATCGAGGCCCTGCAGTCTCTGATTGGTAGAGCCCCTCCAGGCAGCTTTCAGGGCGCTAGACTGTGGGATATCGCCAGACAGGC CACCAGAGGCACCAACATTCTGCCCCTGCTGAACAGCTACCTGCGGGACGTTTTCACCCTGCCTACCAAGCCTACCAGAAAGAAA CCTGCCGTGCGGCCTGCCAGAAGCAGACGGAAGCAGAAGAAACAAGAGTACGCCCGGACACAGGACCTGTTCAGAAAGAAGC AGAGCGACTGCGCCAGAGCCGTGCTGGATGGACCTACATCTAGCTCCGTGCCTGGAACCGGCGCCTTTCTTCAAACCTGGCGCG AGATCATGACAGGCCCATCTCCAGCTCTGGAAGCCCCTCCGCTTCCTACAAGAGGCGAGGTGGACCTGTTTTTCCCCGCCACCGC TCAAGAGATCCAGTCTGCCGAGATCGCCGTGAATAGTGCCGCTGGACCTGACGGCTTTAGCGCCAGACTGCTGAAGTCTGTGCC AGCTCTGCTGCTGAGAGTGATGGTCAACCTGCTGCTGCTTGTGCGCAGAGTGCCTGCCGCTCTGAGAGATGCCAGAACCACATTC ATCCCCAAGGTGCCCGATGCCGTGGATCCCAGCCAGTTCAGACCAATCACAGTGGCCTCCGTGCTGCAGAGGCTGCTGCATAGA ATCCTGGCCAAGAGAGCCCTGGAAGCTATCCCTCTGAACTTCCGGCAGAGGGCCTTTCAGCCTGTGGATGGCTGTGCCGAGAAC ATCTGGCTGCTGAGCACAGCCCTGAACGAGGCCAGAACTCGTAGAAGGCCTCTGCACATGGCCTCTGTGGATCTGACAAAGGCC TTCGATCGCGTGACCACCGACGCCATTCTTAGAGGTGCTCGGAGAGCTGGACTGTCCGGCGAGTTTATCGGCTACCTGAAAGAG CTGTACACCACCAGCAGGACCCTGCTGCAGTTCCAGGGCGAAAGCCTGCTGGTCGAACCTACCACAGGTGTCAGACAGGGCGAT CCTCTGAGCCCCATCCTGTTTAACCTGGTGCTCGACGAGTACCTGAGCAGCCTGGATCCTGACATCAGCTTCGTGTCCGGGGACC TGAGACTGGATGCCATGGCCTTTGCCGACGACCTGATCGTGTTTGCCTCTACACCAGCTGGCCTGCAGGATAGACTGGACGCCCT GGTGGAATTCTTCGACCCCAGAGGCCTGCGCGTGAACGTGAAGAAAAGCTTCACCCTGAGCCTGCAGCCTGGCCGGGATAAGA AAGTGAAGGTCGTGTGCGATCAGATCTTCACCATCGGCGGCACACCTCTGCCTGCCAGCAAAGTTGCTACCCCTTGGAGATACCT GGGCATGACCTTCACACCCCAAGGCAGCATCAACAAGGGCACCAGCGAACAGCTGGATCTGCTGCTCACCAGAACCAGCAAGGC CCCTCTGAAGCCTCAGCAGAGACTGGTGGTGCTGCGGAACTACCTGCTGCCAAGACTGTACCACAGACTGGTGCTTGGCCCTTG GAGTGCTGCCCTCCTGCTGAAGATGGACACCACAATCAGAGGCGCCATCCGGCGGTGGATGGATCTGCCACATGATACCCCTCT GGGCTTCTTTCACGCCCCTGTGACAGAAGGCGGACTGGGCATCAATAGCCTGAGAGCCAGCATTCCCGCCATGGTGCTCCAGAG ACTCGATGGCCTGCACTTCTCTACACACCCTGGCGCTGAGGTGGCAATCCAGCTGCCATTTCTGACCGGCCTGCATCGGAGAGCA GAAGCCGCTGCTCAGTATCAGGGACAGCGCCTGCTGTCTAAGGCCGACGTTCACAGAATGTGGAGCGCTAGGCTGCACGGAAG CTGTGATGGCAGACCACTGCGCGAGAGCAAAAGAGTGCCAGCTGCTCATAGATGGGCCGCTGAGGGAACAAGACTGCTGTCCG GCAGGGACTTCATCAGCATCACCAAGCTGAAGATCAACGCCCTGCCAACACTGGAACGGACCAGCAGAGGACAGCACAAGGAC ATCCAGTGCAGAGCCGGCTGTCAGGCCGTTGAATCTCTGGGACATGTGCTGCAGGCCTGTCACAGAGGACACAGAGGCAGAAT CCGGCGGCACGACAATATTGCCAGATACGTGTGTGGCCGGCTGACCCAGATTGGATGGGCCGTGAAATGGGAGCCCCACTACTC TGTGGCTGGCAGAACCCTGAAGCCTGACATCGTGGCTCATAGAGGCGCCGAGACAGTGGTGCTTGATGCTCAGGTTGTGGGCA CCTCCATGAGACTGGGCTTTCATCACGCCCAGAAGAAAGAGAAGTACAGCCTGCCTGACCTGCTGCACCAAGTGTGCGAAGGCC GAAGAGATGCCGCCAGAGTGTCCACAATCACCCTGAACTTCAGAGGCGTGTGGGCCCCTGAATCCGCTCAGGATCTGAAAAGCC TGGGCCTGACCGACAACGACCTGAAGCTGCTGACCGTGCGTTGTCTGCAAGGCGGAGCCCAGTGCTTCAGACTGCACAGACGG ATGACCACCGTCGTGAAAGCCACAGGCGACGAGGCTAATGCCCTGCCAGCTCATTCTGGCCTGCCTCCAACACAGCTCGGAGGC AGAACACTGGGACCCAGCGCTCATAATCAGAGCGCCCGCACAACATGA (SEQ ID NO: 24)
R2-1_PM	**5' UTR**	CTATTAATGGGATGAAGAAGGGGGACACGAGTTTGTGTGTGCATCCAGTTTCCATGGTGCATGCAGGAGTGGTGGTTTAAATGG CGAGACTCTACAGGGCTTCCATGGCTACACGGGATGCAAGGCATCAGACATTTTGGCACAGGCAATCCTTTTGGTCTCTACCGCA ATCATGTCTTAGACCTCAGTAGCGACCACTACAACCACAGTGGTGACTGCTGTTGAGTGAAGGACGACTGAGCGCTGGATAACA ACTTTCTTGCGTGGCCCAACATCGAAGCAACCACTTCGGAGCTGGCACAAGGCAAGAGGGCAGCCCAAGGTGTGAATCATCTCA ACTTCACTGCAGGAAGAAATGCTGTGCAAGGATGAGTGTGAACGACACCAACGGGATTGTTGCTGACCAGGAGGTGCCAACCA AATTTGAATGGATTGACTTTGGGCCTGGTTTCTCCTGCGTGTATTGCACGGAAAAACAAGTGGCTACACGTGTGGCCGTCGTGTC CTGGGGTTTCGCAACACAACTCCACAAGATCGACAACTATGAGGATGACAATGTACTTAAAGAACAAAGAGACTGACGCCAAAG GGGATTTAAACCGCCAAATCGTACATTGGGTCTCACTACAATTTTTTTACGTGTATTTATTTTCCTAAGTGTCTGTACTTGCCATTCT TCGCTGCTTTTTCTGCATTAATTGCATATCGTATGCAAATAAGCGAATTAACCACCACCGTGCAACTATATGCAGATGTTACAGCT GAGCCCTCTATCATACCGGTGTACTAATCTGGTATGGTGTTGGCATGCTATGCTTGCGTAACGACCTTTGCTGATTGGTTCAGTCG GCTGATGGTGGGTTCAGGCGAAACATTTGTATATTGGTTTAATCAAACCGAAACACTAAAATTTTGAACACAGTTTTCCATTACAC CAGTTGTATTGCTAGAAGTGCAAATCGAAGGAGTCAATTTTGACCGACGATTAGCTGCCGATGTGCGGTGAAAAAGCTGATCAC AATAGCATACACTTGGGCCGACAACCCCGTGTGCTATAAACGTAAGTCGCGAATTATAAAGAAAACAAACCGGACGGACTACTC GGTGACGAACTAACATCGCTC (SEQ ID NO: 25)
	CDS	ATGAATGAGCGATTAACAGACGAGCTGACTACGGAATTTATCCTTTCGGACATGTTTTTATGGGACTACCCATGCACAGATCAGA ACAAATGTTATCCATGCAATCTTGTTTTCCTAGACCACAGAACTTGGTCATCACATATGGCACGGGTACATCCACATGCAAACAAA ACGTATAAATGTCGAATTTGTAATCGCACAGCAGATAGCATACACAAGATAGCGTCACACTACGGAAGAACTTGCAAAAGTTTAA TAGGTAAAACTAATGCTATAACCACCACAATTGATGAAACACTATTTAGTTGTTTACATTGCAGCAGAGGTTTTACTACGAAAACA GGTTTAGGGGTACATACTAGACGAACTCATCCGACAGAACATGAGGCTATACTACAGCAAAACACACCAGGAAGGAAAGTTAG ATGGGGAGAAGAAGAGGTAGAAATTATGGCCCATAAAGAAGCCCAACAGAAGGATGAGGACATAAACATGAATCAACTAATAC AGAACTCAGTTATGCCACACAGAACGCTAGAAGCGATTAAAGGGAAGCGGAGAAATATCAAGTATAAGGAATTGGTAAGGACT TTGAAAGAAACTACCTATAAGGTAGAAAATCAATGCCTTGTTAACTTAGTTTTACCGACAACATCGGAAATAACAACTACACCTTC GGAAGGAGATCAGCCAGCAATAAGGGCCGAAAAAGAACAATCACCGACAGCAGCTGAGGATCTTCAGGTCATAATTAACGATC TAAAGAGCCAGAATTTTAGCCACAATCAGGCGTTACTGCTACTCAATTCTCATGTAGAAAAGTTTTTAAATCGAAGTAAACCAATT AAAAGGAAAGATCACGTAAACCAACAGGAGATAGATGAGAATAGGCATCGAAGACAATCAAAGCAAACTAAATACAGGAGATA TCAATATTTATACCATACGAACAAGAAAGCTCTATTAGACGAGATTACTTCAGATAGATCGGGGCCAAGTATATACCCAACTGAG GAAAGCATACGGGGAACATTCGTTACTTTATTCGAGTCAAACTCTCCTCCAGATAATATACCCTCTAAATTAAAAAACGACCAATC CTGCATCGATATCGTAAAAGCAATCACCTTAGATGAGTTGATTAAGACCCTAGCAATTATGAAGGATAAGTCACCTGGACAGGAC

		AACATTACTCTTAGCGATCTTAGGACTTTACCAATAAAATATTTACTAGATATCTTAAATATCATCCTTTACATACAGGATATACCA CAAATATGGAAACAGCACAGAACAAGACTTATCCCGAAAACTAAAGAGGAATTAGAAAAACCCTCAAATTGGAGACCCATAACC ATCTCATCAATTGTAATTAGGCTATTACATAAAATTTTAAGTTATCGTCTAGGACAGCAATTAAAGCTTAATTACAGGCAGAAAGC ATTCCTCCCGGTAGACGGATGTTTCGAAAATAGTGCATTACTACACTTCATCATACACAACGCCAGGCAAAAGCACGAAAACACG CAAATAGTGTCAATAGACCTCAGTAAGGCATTCGATTCTGTCAGCCACGAATCGATTATTAGAGCCTTAAACCGATTTAACTTATC AAAGGAATCCATAACGTACTTGACCAACATCTATAAGTGTAATCTAACTGATATTGTATTTGGATCGACAATAATGCGTAACATAA ATCTAAAAAGAGGCGTAAAGCAAGGAGATCCACTTTCACCGTTACTATTTAACATGATTATGGATGAATTATTAGATAACTTGCC GACATATATAGGAGTTAATGTAGGAAATCAGAAAGTAAATTCTATGATGTTTGCAGACGACCTTATCCTATTTGCAGAAACGGAA TGTGGCATGAATAAACTCTTAGATATAACTACTAAATTCCTCGATGACAGACACTTGAAAATAAATATAAACAAATGCAATTCGTT AAGATTTATCAAGTACGGCAAACAGAAGACATTTAGTGTTGCAACGACATCATCGTACTTTATAAATAACGAACCCATTAATCCG GTATCATATGTAAAGGGATTCAAATATCTAGGCATTGAATTTGACCCAAGGGGAAAACGATCTATAAGCTGTAACCTGCTCGCAG CAATGTTAAACAAACTGACCAGAGCACCGTTAAAGCCAGAAAAGAAAGTATATTTAATCAATAACAATTTAATACCTCGTATTATT CATCAATTGGTCCTCGGAAAAGTTACCAAGGGTTTATTGATGTCACTTGATTCTGAAATTAGGAAAACAGTAAAGCTTCTGCTCA GGTTGCCACACGATACGCCCGACAGTTTCTTTTATACATCAGTATCCAACGGAGGAATGGGTATAAGAAATTTATGCGACTCAGT TGCACTATCTATAATAAACAGACACAACAAATTGATAACTTCAGATGATCTAGTAATAAGAGCATTATCACAACAATCATACACTA TTGCAACGTTAAAACAGGCCCATATCATTGCAGGCTCCAAATTTCCTTCGAAATCTTTAAATCAGAACAAATGGTCAAATAAACTA TATCAAACAACAGATGGTCGGGGGTTGGTATACTGCCAATCTCAAACAGAAAACAATTCATGGATAACAGGGAATCATAGAACA ATAAAATCGTATAATTACATAGACATGGTTAAACTAAGGATTAATGCACTACCGACTAAATCGAGATGCAATCGAGGGACGTTAG AGACCAAGCAATGTAGATTTAAATGTCGAAGTATTAACAACCAAATTTCAGAGGAAACATTGGCACATATCTTGCAAAAGTGTGA TCGCAGTCATTATTCAAGAATCGCAAGGCATGATTCTTTGGTGCAATTTCTGGCAACGGCCGCACAAAAACTAAACTGGGAAGTG ATCAAAGAACCCACTTTACCGAGCGATACAAATAAGGCAAAACCGGACTTAATTTTAGTAAGAGACTCTCATGTCTTGATAGTAG ATGTGGCAGTTCCGTGGGAGTCTCGATCATTGGCACATGCATACGATTTTAAGGTGAAAAAATACGCTACTGACAAAAAAATGC AAGCATATTTAAAAACTATATATCCGGAAAAAGAAATTAGAACGGAGGCTTTAATCATATCTGCACGTGGGGGCTGGTGCGCTTT AAATAATATGGTAACAAAAAAGGTGGGATTGTCAAGTGCATGGGTAAAATTAGCATTGATCAAGGTCATGGAGGGTTCCGTAAA GATATGGCGCTCTTGGAGCAAAGGATAA (SEQ ID NO: 26)
	3' UTR	TTTAAGGTAAAATCGTGGGATTGTTTTGATGGCAATCTGCCTAGTCGCGGCCTTCCATTTTGGGTAGGCAGCAGACCCATCTATAT AACAAACTACTTTGCCTTTCATAGGGGTACCCGACCCTACCAACTTTCGGGGAAGTAAAAGAAA (SEQ ID NO: 27)
	Codon optimized	ATGAACGAGAGACTGACCGACGAGCTGACCACCGAGTTTATCCTGAGCGACATGTTCCTGTGGGACTACCCCTGCACCGACCAG AACAAGTGCTACCCTTGCAACCTGGTGTTCCTGGACCACAGAACCTGGTCCAGCCACATGGCCAGAGTGCACCCTCACGCCAACA AGACCTACAAGTGCCGGATCTGCAACCGGACCGCCGACAGCATTCACAAGATCGCCTCTCACTACGGCCGGACCTGCAAGTCTCT GATCGGCAAGACCAACGCCATCACCACCACCATCGACGAGACACTGTTCAGCTGCCTGCACTGCAGCAGAGGCTTCACCACAAA GACAGGCCTGGGCGTGCACACCAGAAGAACACACCCTACAGAGCACGAGGCCATCCTCCAGCAGAATACCCCTGGCCGGAAAG TCCGCTGGGGCGAAGAGGAAGTGGAAATCATGGCCCACAAAGAGGCCCAGCAGAAGGACGAGGACATCAACATGAACCAGCT GATCCAGAACAGCGTGATGCCCCACAGGACCCTGGAAGCCATCAAGGGCAAGAGAAGAAACATCAAGTACAAAGAACTCGTGC GGACCCTGAAAGAAACCACCTACAAGGTGGAAAACCAGTGCCTCGTGAACCTGGTGCTGCCTACCACCAGCGAGATCACCACAA CACCTAGCGAGGGCGATCAGCCTGCCATCAGAGCCGAGAAAGAGCAGTCTCCTACAGCCGCCGAAGATCTGCAAGTGATCATCA ACGACCTGAAGTCCCAGAACTTCAGCCACAATCAGGCCCTGCTGCTGCTGAACAGCCACGTGGAAAAGTTTCTGAACCGCAGCA AGCCCATCAAGCGGAAGGACCACGTGAACCAGCAAGAGATCGATGAGAACCGGCACCGGCGGCAGAGCAAGCAGACCAAGTA CAGAAGATACCAGTACCTGTACCACACCAACAAGAAAGCCCTGCTGGATGAGATCACCTCCGACAGAAGCGGCCCCAGCATCTA CCCAACCGAAGAGAGCATCAGAGGCACCTTCGTGACCCTGTTCGAGAGCAACAGCCCTCCTGACAACATCCCCAGCAAGCTGAA GAACGACCAGAGCTGCATCGACATCGTGAAGGCCATCACACTGGATGAGCTGATCAAGACCCTGGCCATCATGAAGGACAAGA GCCCAGGCCAGGACAACATCACCCTGTCCGACCTGAGAACCCTGCCTATCAAGTACCTGCTGGACATCCTGAACATCATCCTGTA CATCCAAGACATCCCGCAGATCTGGAAGCAGCACCGGACCAGACTGATCCCCAAGACCAAAGAAGAACTGGAAAAGCCCAGCA ACTGGCGGCCCATCACCATCAGCAGCATCGTGATTCGGCTGCTGCACAAGATCCTGAGCTACAGACTGGGACAGCAGCTGAAGC TGAACTACCGGCAGAAAGCCTTCCTGCCTGTGGACGGCTGCTTCGAGAATAGCGCACTGCTGCACTTCATCATCCACAACGCCAG ACAGAAGCACGAGAACACCCAGATCGTGTCCATCGACCTGAGCAAGGCCTTCGACTCCGTGTCTCACGAGTCCATCATTCGGGCC CTGAACAGATTCAACCTGTCCAAAGAGTCTATCACCTACCTGACCAACATCTACAAGTGTAACCTGACGGATATCGTGTTCGGGT CCACCATCATGCGGAACATCAACCTGAAGCGGGGCGTGAAGCAAGGCGATCCTTTGAGCCCTCTGCTGTTCAATATGATCATGG ACGAGCTGCTGGATAACCTGCCTACCTACATCGGCGTGAACGTGGGCAACCAGAAAGTGAACTCCATGATGTTCGCCGACGACC TGATCCTGTTTGCCGAGACAGAGTGCGGCATGAACAAGCTGCTCGATATTACCACCAAGTTTCTGGACGACCGGCACCTGAAGA TCAATATCAACAAGTGCAACAGCCTCCGGTTCATTAAGTACGGCAAGCAGAAAACCTTCAGCGTGGCCACCACCAGCAGCTACTT CATCAACAACGAGCCTATCAACCCCGTGTCCTACGTGAAGGGCTTTAAGTACCTGGGCATCGAGTTCGACCCCAGAGGCAAGCG GAGCATCAGCTGTAATCTGCTGGCCGCCATGCTGAACAAACTGACAAGGGCCCCTCTGAAGCCCGAGAAGAAGGTGTACCTGAT CAACAACAATCTGATCCCGCGGATCATTCACCAGCTCGTGCTGGGCAAAGTGACCAAGGGCCTGCTGATGAGCCTGGACTCCGA GATCAGAAAGACCGTGAAGCTGCTGCTCAGACTGCCCCACGATACCCCTGACAGCTTCTTCTACACCAGCGTGTCCAATGGCGGC ATGGGCATCAGAAACCTGTGCGATTCTGTGGCCCTGAGCATCATCAACCGGCACAACAAGCTGATTACCAGCGACGACCTCGTG ATCAGAGCCCTGAGCCAGCAGAGCTACACAATCGCCACACTGAAGCAGGCCCACATCATTGCCGGCAGCAAGTTCCCTAGCAAG AGCCTGAATCAGAACAAATGGTCCAACAAGCTCTACCAGACCACCGATGGCAGAGGCCTGGTGTACTGTCAGAGCCAGACCGAG AACAACAGCTGGATCACCGGCAATCACAGAACCATCAAGAGCTACAACTACATCGACATGGTCAAGCTGCGGATCAACGCTCTG CCCACCAAGAGCAGATGCAACAGGGGCACCCTGGAAACAAAGCAGTGCAGATTCAAGTGTCGGTCCATCAACAATCAGATCAGC GAGGAAACCCTGGCTCACATCCTGCAGAAGTGCGACAGATCCCACTACAGCCGGATCGCCAGACACGATAGCCTGGTGCAGTTT CTGGCCACCGCCGCTCAGAAACTGAACTGGGAAGTGATCAAAGAGCCCACACTGCCCAGCGATACCAACAAGGCCAAGCCAGAT CTGATCCTCGTGCGGGATAGCCACGTGCTGATCGTGGATGTTGCCGTGCCTTGGGAGAGCAGATCTCTGGCCCACGCCTACGAC TTCAAAGTGAAGAAGTACGCCACCGACAAGAAGATGCAGGCCTACCTGAAAACAATCTACCCTGAGAAAGAGATCCGGACAGA GGCCCTGATCATCTCCGCTAGAGGCGGATGGTGCGCTCTGAACAACATGGTCACCAAGAAAGTGGGCCTGAGCAGCGCCTGGG TTAAGCTGGCACTGATCAAAGTGATGGAAGGCAGCGTGAAGATCTGGCGGAGCTGGTCTAAGGGCTGA (SEQ ID NO: 28)

R2-1_SSa	**5' UTR**	AATCTTTAACCCCGGACTCTTGGGGTTCTTACGACTCTGTATGAGGAACAGTCGAAGAGAGGGCGCTACCAATCCAAGTATATGT CCCAAGAGGGCTGGGACAGGGTGGAAGAGTGCACCTCGCGATCTGGGGCAGGGAAGGAATGGAGAGAAGTCGAAGAAGGCT TGTAGAGAAGGGGCTCTCCTAGATCCTAACCTGTATGACGCCCGTAAAACGGTGACCCCAGTAGCGAATAAAGGAGGCAGGTG

		ACAATAGAGGGCAGGGCCGACTTCCCAGGTTTACATTGTTGTACTTGTCAACATAAAGAGGTGTCTCAATAGTTTGAATCAACAA GGGAGAGGAATACCGACCTGCTCCCTTGGGGCGGGGGTACTGGTCTTAGCCCGGTTCCCCGCAAGTTTCCTTTGCCTGGGATGT GCCTGACTGGCTCCATCCCCTTTCCCCATTAGGCACGGCTAGATGACGCACCGATGGGCGGGTGTGTAGGTCGCTACCGAAGGG GACTGGGGGTGTCCGGTGAACCAGGACTTCCCAAAATGGTCTCACATTTTTAAGCGGCTTGAGTATCGCCCAGTATCCTCGCGCG GCACTGGGAACCCAGTCAACCGCTCTGTGCCCCGGCGCAGGCGGGGGTTTAATGTCTCCCCGGCTTCACCGGCGCTTCGGCGAC GACGCAGAGGAGCACCCGGAGGCCCCCATGAACTTAAACCAACCTATCTTGAAATATGGCCTCTCGTTCGGGTGAAGGGCAGGT GGGAAGAGAGGGCTGCCTCACGATAAACACCTAGTCAATAGCCAGTCGGGAAAAATGTGGAATGTTAGGACAGGGAGGTAAG GGAGGCGGCTTTTTCGTAAGGCTCCTTCAACCCCTACCTGTAGTCCACCTATTGCAGGTGTTGACAACATGCAAGATGACCTGCCT CGTTACGGGTCGCGTATCATTGCTACAGGTCGTGTGCCGCTTCTAAGAGGATAGTAAGGAGAGGTTATAGGGAGGTCCTGTTAG GGCTTCCTCAACCCCTCTCTATGCGATTCCTTACAGGAGTGGATCGAGAAGTCCCGGACGTAATACACCCTGGAGGTAAGGGAG TGGCCTTCTAGTAGGGCTGCTTCAACCCCTCTGATGGGAGTGTACCGGGAACCTCGACTTGTAAGCACAGGTTAGTATGGGAGC AGAAGGGGGAGCCGTAATGGGCTTCTCTTTCACCTGCTTACATAATACCTGTGGTGCATGTATCTAGGTCTTGGCGGGAGAGTAC TGAGAGACAAGGTTGAGACCCCAAGATTGGGTCTCCCTAGCCTCTATAGCTGCGACTCTTAGCGGGGATATGGAGTAACATGTA CCAAGGGAGTGAATAATAAAGGAATTGACGGGGTACAAGTGACTGTTGGCCCGAATCCTAGCCACTTGATGACCAGGGATATAT TACAGACTGAGAGCGAATCTAGAGACGTGAGAATAAAGTGAGAATTGGTTGAGCGAATACAGAGGAAG (SEQ ID NO: 29)
	CDS	ATGAGCGGAAAGAGAATTGTTGAGATGAGCGGTTGCGATGAGAAAATCTGCCAAAATAAGCATTGCTTGAAGCGCAGATGGGC GTGGATTTCCGGCCCAAAGGGGGAGACATCTCCTCCTCGCAAGAGAGGAACTTGCGAAAACGTATCCTTCCAGGATAAATCTCA TGCCTCGGACCCAGATCCTCTCAAAGCCCCGGAAGCGAGAGAGGACGCGGGTTCGGTGGCGCCTCAGTGGGTTGGCGAAATAA AGACACCTAGCCTAACATCGCGGGATGGGGTTAGTGAAGTCGTGCTGCCACCACAGCCAGTTCATGCAGAGGGAGTATCCCCAG CCAGCGACTCAAAGGACAAGGCTACGAAGATCACCCTGCTGATATCACTCCCCGTCTGTGATTTGAGATGCGGACGATGTGAAA GACCATTAGAGACCGTTGGGAAGGCGGTGAGACACTTTGCTGTGGCTCATCCGACGGTGTCGGTGGTTTTTAAGTGTCAAAAGT GTGAGAAGAGCAGCAAGAATAGCCACTCCATCTCCTGTCATATCCCTAAATGTAAGGGAATGACAGAGACCCGGACGGATGTGG AAGGTGATCACGGCTGTGATCATTGCCAGGAGAAGTTTACAACGGCTATGGGGCTGACTCAACACAAGAGACACAGACACATCG TCCAGTATTGTAAGGAGAAGGAGGGGGAGATGACAGCAAGAAGGAAGGGTGAAGTCGAAGCAGTCAAATGGAGCGAATGGG AAGAAAGTGAGGTGGCAAGGTTGAGCGATGGACTGGCTGGGCTAAAAATGATCAACAGGCGAATCGCAGATAGCCTGGGGAC TGGAAAAACTGCGGAACAGGTGAGGCAGAAAAGACGTAGAATGAGACCTGAGAAGGTACGGTGTGACAAACCTAAGGAGGCA AAAGATAAGAGCAATCTTATCAAAATGCTGTCCATACCGAGTGCAACACCAACACCCCAAACTGGGCTCAAAGGATTCCTCCTTG GAGAACTAAATGGGGTTGCTACCAAAGGTGAAGTACAGATTGGGGGAGTTACGTTGTCCCTAAGGGGGGTAGAGCAAGACTCA GCCTTGCTCAATACGTCGGCCCTTGAACTGCAACGATTGTTGGGAGGGAGAGCGGGGAGCGCGAATCCGTTAAGCTTACAGAG AGAACGTGAGACCACATTGCCAAGTGAAAGAAGGAAAACGAAACAAGGAGAGTACCGGAGGGTACAAAAGATGTTTCGAAGC AATGAAAAGAAAATTGCTAAATACATTCTAGATGGCAATGGAGATGGAGAAGCGGCCTCCCCGCCACTCGAGATCGCTTTGGCG TTCAAAAGTAGGTGGGAGGAAGTGGAAACATTTCATGGGCTTGGCCAGTTCTACTCGAGGGGGGAAGCGGATGGTGTGGTCTT CCGGTCACTTATCTCAATGAGCGAAGTATGTGAAAACCTGGGGGCAATTAAAAACAACACAGCAGCTGGGCCAGATGGGATAAC AAAACCGGCATTGCTTGAATGGGATCCCACTGGTGCGAAACTGGCCGCTATCTTTTCCATATGGTTGACATCGGGCACCCTGCCT GGGCCTTTTAAGAAGTGCAGAACAACCTTAATACCCAAGACCGATGACCCGATTTTACTCACCCAGGTGGCTGGGTGGAGGCCT CTCACTATCGGGTCGGTGGTTTTGAGGCTATACTCTCGCATCCTGACACACAGGCTGGAGCGGGCGTGTCCCATTAACCCGCGCC AGAGAGGATTCATTTCCTCACCTGGGTGTTCGGAAAACTTGATGATCTTGGGAGGTCTAATCAAGAGAAGTTGGGCGAAGGGCG AGAGGCTGGCGGTAGTGTTGGTTGACTTTGCGCGTGCGTTTGACTCTGTGAGCCACTCGCACATCTTGGAGATACTCAGACAGA GAGGGCTTGATGAACATATCATCGGAATCGTAGGTGACTCGTACACCGATGTAACGACCACAATTACAGTCAGTGGGGAGCAGT CCCCTCCCATTGACATGAGGGTAGGGGTTAAGCAGGGAGACCCGATGTCTCCGCTACTGTTTAATCTAGCCCTGGACCCAATGAT CGACACCCTTGAACGCTACGGCTTGGGGTACAGGATGGGCGAGCAGCAGATCACGGCCCTAGCTTTTGCTGACGATCTGGTTCT GGTGAGCGACTCGTGGGAGGGCATGGCGTGCAATATCCGTATTCTGGAAGAATTTTGTCGACTGACTGGGCTGAGGATTCAGCC TAGGAAGTGTCATGGTTTCCTCATACAGAAGATTCAGAGGGCGAGATCGGTAAACCTCTGTAAGCCCTGGATAGTGTGTGGTGA AGAACTACATATGGTCGGGCCGGAAGAGTCGGTTTCCTACCTGGGTATGAAGGTGAGCCCATGGCATGGCATTATGGAGCCAG ATCCTGTCGAACGACTCTGTAACTGGATCAGTTCGATTGGGCGGTCACCGCTGAAGCCTTCTCAGAAGGTGAGGATGTTGAATGT TTATGCTGCCCCGAGGATGACTTATCAGGCGGATCATGGCGGGCTGGGGCCAATTGTCCTGAATGTACTCGATGGGATGATCAG GAAAGCAGTGAAGGTGTGGCTACACCTTCCGCTGTGTACCTGTGATGGGCTACTTTACTCTAGATGCCAGGACGGTGGACTGGG CATAGTAAAATTGGCTTGTCAAATCCCTTCTATCCAAGCTAGAAGGGTCTACCGCCTGTGGCATTCTAAGGAAGCCATAACGCGG GTAGTCACCCGAAGGACGGTCGAAGCGGAAGAGTACCGTGGGATGTGGCTGAGAGCCGGTGGGAGTGAGGCAGGTTTGCCTC CCCTGGAAGATAGGGAAGAAGGTGCTGTACAGTGTACAGACACTGCCGGTTCGGTGAAGCCGAAAAACCCAGTCATTCCCGATT GGAGGCGAGCTGAGTTCCTCAAGTGGCAAAACCTGACAGCGCAAGGGGTTGGAGTGCAGGTCTTTGGCGGTGACAAAAACAGC AATCACTGGATGGCGAATCCGGAAACGTTGGGATCGAAAGAGCGCCACTATATTGCAGGTCTACAGTTGAGGGCCAATGTATAT CCAACTCGCGAGGCACTGTCCAGGGGCAGGCCGGACTTACCTAAAGTCTGCCGGCAATGCCTAGCAGGAACTGAGTCTTGCGCG CATATTCTCGGGCAGTGTCCTGCAGTGAAGGATAGCCGCATCAGGCGGCATCATAAACTGTGTGACCTGCTAGCAAGTGAAGCC GAAAGCGCCGGATGGACCGTCATCAGAGAAATGTGTTGCAGAACTCGTGCCGGAGCTTTGCGGCGTCCAGACCTGGTGTTTGTG AAAACCGGTTTTGCTTTGGTGGTGGATGTTACTGTGCGGTACGAGATGGCCTATGATACGCTCATGGGTGCGGCTGCCGAGAAA GTTGCTCGGTACACCCCAATTACTCCATATGTTGCGATGACCCTGAAGGCAAGGAGAGTCAAGGTGTTTGGCTTTCCACTGGGAG CCCGAGGCAAATGGCCGGGAAGCAACGACCGGCTGCTGAAGGCTATGGGTGTTGGTGGCGGCAGGAGGAAACAGCTGGCCAA GTTGTTTAGCAGGAGGGCGCTCCTGTACTCCTTAGATGTCCTTAGGGACTTCTACCGGGCGGAGGGAGAAACGGGGGACTTGG ATGATGAGAGCGTGGATGATCATCTATAG (SEQ ID NO: 30)
	3' UTR	ATCCGTTTGTTATGATTGGAGGGAGCCTGCCGAGTGGTATGAGCGCTCCAACTATTGAACCCATATGATTCCCGAGGCCTGGCCA GACGCCTAGATGCCTGCCACAATTGAACGCAGCCCTAGCTTGCTAGGAGATCCATAGGAACTGGCCTATGGGGCGTCATGACGG TTGAAGTTCCTCCATAGCGTGCTTGGGAGGGGACGACAATGACGAGTCATGACGTACCGAGAGAACCCCAACCCAGGTTGGGG GAGAGAGCCAGCAAGAGCGGAGATGCTTGGTATACCAAGCTAGCAGAGAGAGGGTTGAAGAGGATGACTACTGGGCTCAGAG TCATCTCACCCTAAAAGGCGGTGGGGCATCGGTTGAACACCTACCCATACCGGGATGGGAGGTGGTAGGCCGAAAAAGAACAG GAAGATGGTGGAGTAAGTTGAGAGCGGTTGCTCGGGAAGTTATGTTGTGATAACTCCATTAAGGCCGGTGGGCATGGTGCGGA TAATGGAAACTATAAAAACAATAAAAAGAAAGACCAAAAAAATGTTCTGTTATGATGCCTTACACATGTCTGGGAGACCCCATAA GGGTCTCCCCTTATACTTCACTGGGAAACCCCATAAGGGTATCCCCCTATATTTACTGGGAGACCCCATAAGGGTCTCCCCCTATA GATGTAGAGCGTAAGGGGTCTCCAAAGTACCGGCCGATATGGCCTTATGGCAAACTCTGGTGGTAGGGACAAGGAGGTAAGGG CAGTGCCAACCCCTACTTGATCGGGACCATCCAGGGAATGCCATCCTCCCGCGAAGGTGATGTGGTGAGGTAAGGGGGGAGCC CGTCTTCGAGTTTCCCCAACCCCTACCCACAGGTGAGAGGAGGAGAAGAGGAATCTGTCCCCAACGGGAGGAGGGTGAGGTGT

		AAGGGGGAGACCTTCTAGTAGGGTCTTCTCAGTCGCCTGACGTCCTGACTGTGGGGTGGATCAGTACCCTACAGGTGAGACCGG TGAGGTAAGGGTGTGGCCCTCTTGAGGGCTGCGCCAACCCCTACTCGAGGTAACCTGAGGGAGTGGTGGAATGGCGGCATGTT AGTGCTGGGACTTGATTGCGAGGGTTTAATGAGAGTGGCCTGCTGAGAGCAACACTTGTGGTGCTTAAAGCGGGGCGGCCCAT GACCACCGTGAGATAGGACACTGCACAGTGCAGCCATGAGGTTCCTGGAGGATGATGCGATGAGGTGGGGGCCTCATCAGCCC CTCCTGGCAGGGCGTCGGCCAGGGAAACTAAATGTCTCTAGCATGTCAGTGCAGTGAGGTAAGGGGAGAGCACTCTAGTAGGG CTCTTCCAACCCCTACCTGTAGGTCACCTGGTCCAGGTGTCGATGATGTGAAAACAAGAGCTACTTTGGTACCGGTCTGTTGCAA AAAGGGTTCTGCAGAGGACGACGGCTATCCCTATCGGGAGGGAATAGTCGGTCCCAGGTAGTGGAAAATGGGGCTTTCCACTG AGCATGAAAATGTGGTAGAGGTTGCGTCCAACCCAATGATTTGCAGCAGAGCTCTTGGACACGAAGTCTGTATAGTCCCATGCA GGCAGCCAACCAGAGAATGGTGGCAAGACCCCAGCTCCGTATGGGAGGGGAGGGCCAAGATATACGGAACGGCTGCTAAAGC GTTCTGCCGGTGTCAGTCTAATCACAGACAGCTGTGACGAAACAAAGTATGGGTTCCGACATGCTTGGTCAGCTCTTAGCCGCAA GGCTTAAATCGAACGCAGCCCGCCGAGAGTGAACATTAAACGGGGATGGAATGTGTCTAGCGGTTACGTACTACCAGGGCTCA GGTTCGCCTGAGCCGAGGCTCTACACGTCATGGTGGGAGTTCTCCCCACGCTCGTGAGGGCATGTAGTGGGATGGCATGTGGCG GACCATCAGCTGGCACTACCAGGCCTCGGGCTTGCCCGAGTGCGGGACCTCACACATTGTAGGTGTGCTTGTCCCCCCTACGTTC GAAGACTTGAGGCGGAGAATACTCATAGGCCCCACGGCAAAGGGACACAACACGGAGGCTTGTGTCCGACGAGCCGTGGACTC CTATAGACAGCCCGGGATATCACTGGGCACGCTCATACTGAAGAAATTCGATGAACCGGGCCTACCGGAGCAAATGCACTCTAA TCGCCTTTGTGGGCGACTGTGGCCCCCTCATGCGAGTGAGGAATATCATAAACTGCAATGGTTCAAAAAGTGATTCCTATGGCTC GTCGGGGAGGGCTGACTGGGGCAAGCAAATGATTGAAAGGGGAAGAACCTTTTTCAACTGTTTCTTGCCAAGCCCGGTTGATG GTGGCGCTAGTAATTGCGACGGGAAAATGCGGTTTAAGTCTCCGAAGTAGTGCGTAGCACCGGATGTCGACCGGGTGTAAAAG CCCTTCGTAAAGTCCCTGGGGAGGTCAGTCCTGGGGCTACTGATGCGCAGTATGTAATTCGCAGAATAGGGCCATCGATACCGC CTGCGTGACTCGACTGGGTTTCCACTTGAGGATATCCGACCGTAGCGTGCACCCTCTTGTAGTTGCGCCGGAAACGGCTGTGTTC CCTCACGTATGTGAGGAAACTCAACAATGTGAGTGGGTAAACGGCGGGACGAACTATGGCTCTCGT (SEQ ID NO: 31)
	Codon optimized	ATGAGCGGCAAGCGGATCGTGGAAATGAGCGGCTGCGACGAGAAGATCTGCCAGAACAAGCACTGCCTGAAGCGGAGATGGG CCTGGATCTCTGGACCTAAGGGCGAGACAAGCCCTCCTAGAAAGAGAGGAACCTGCGAGAACGTGTCATTTCAGGACAAGAGC CACGCCAGCGATCCCGATCCTCTGAAAGCCCCTGAGGCCAGAGAAGATGCCGGATCTGTTGCCCCTCAGTGGGTCGGAGAGATC AAGACCCCTAGCCTGACCAGCAGAGATGGCGTGTCAGAAGTGGTGCTGCCTCCTCAGCCTGTGCATGCTGAAGGTGTTAGCCCT GCCAGCGACAGCAAGGATAAGGCCACCAAGATCACCCTGCTGATCTCCCTGCCTGTGTGCGACCTGAGATGTGGCAGATGCGAG AGGCCACTGGAAACCGTGGGCAAAGCCGTGCGGCATTTTGCCGTTGCTCACCCTACAGTGTCCGTGGTGTTCAAGTGCCAGAAG TGCGAGAAGTCCAGCAAGAACAGCCACAGCATCAGCTGTCACATCCCCAAGTGCAAGGGCATGACCGAGACACGGACAGACGT GGAAGGCGATCACGGCTGCGATCACTGCCAAGAGAAGTTCACCACCGCCATGGGCCTGACACAGCACAAGAGACACAGACACA TCGTCCAGTACTGCAAAGAAAAAGAGGGCGAGATGACCGCCAGACGGAAGGGCGAAGTGGAAGCTGTGAAGTGGAGCGAGTG GGAAGAGTCCGAAGTGGCCAGACTGTCTGATGGACTGGCCGGCCTGAAGATGATCAACAGAAGAATCGCCGACAGCCTCGGCA CCGGCAAGACAGCTGAACAAGTGCGGCAGAAACGGCGGAGAATGCGGCCCGAGAAAGTCCGCTGCGACAAGCCTAAAGAGGC CAAGGACAAGTCCAACCTGATCAAGATGCTGAGCATCCCCAGCGCCACACCTACACCTCAGACAGGCCTGAAGGGCTTTCTGCTG GGAGAGCTGAATGGCGTGGCCACCAAAGGCGAGGTTCAGATCGGCGGAGTGACCCTGTCTCTGAGAGGCGTGGAACAGGATA GCGCCCTGCTGAACACAAGCGCCCTGGAACTGCAGAGACTGCTTGGAGGCAGAGCCGGAAGCGCCAATCCTCTGAGTCTGCAG CGGGAAAGAGAGACAACCCTGCCAAGCGAGCGGAGAAAGACCAAGCAGGGCGAGTATCGGCGGGTGCAGAAGATGTTCAGA AGCAACGAGAAGAAGATCGCCAAGTACATCCTGGACGGCAACGGCGACGGCGAAGCTGCTTCTCCTCCTCTGGAAATCGCCCTG GCCTTCAAGAGCAGATGGGAAGAAGTGGAAACCTTCCACGGCCTGGGCCAGTTCTACTCTAGAGGCGAAGCAGACGGCGTGGT GTTTCGGAGCCTGATCAGCATGAGCGAAGTGTGCGAGAACCTGGGCGCCATCAAGAACAATACTGCCGCCGGACCTGACGGCA TCACCAAACCTGCTCTGCTGGAATGGGATCCTACCGGCGCTAAACTGGCCGCCATCTTCAGCATCTGGCTGACCTCTGGAACCCT GCCTGGACCTTTCAAGAAGTGCCGGACCACACTGATCCCCAAGACCGACGATCCTATCCTGCTGACACAGGTGGCAGGCTGGCG GCCTCTGACAATTGGATCTGTGGTGCTGAGACTGTACAGCCGGATCCTGACACACCGGCTGGAAAGAGCCTGTCCTATCAACCCC AGACAGCGGGGCTTTATCAGCAGCCCTGGCTGCAGCGAGAATCTGATGATCCTCGGCGGACTGATCAAGCGGTCATGGGCCAA GGGCGAAAGACTGGCTGTGGTCCTGGTGGATTTCGCCAGAGCCTTCGATAGCGTGTCCCACAGCCACATCCTCGAGATCCTGAG ACAGAGAGGCCTGGACGAGCACATCATCGGCATCGTGGGCGACAGCTACACCGATGTGACCACCACCATCACCGTGTCTGGCGA GCAGAGCCCACCTATCGATATGAGAGTGGGCGTGAAACAGGGCGACCCTATGAGCCCTCTGCTGTTCAACCTGGCTCTGGACCC CATGATCGACACCCTGGAAAGATACGGACTGGGCTACAGAATGGGCGAGCAGCAGATTACCGCTCTGGCCTTCGCTGACGATCT GGTGCTGGTGTCCGATAGCTGGGAAGGCATGGCCTGCAACATCAGAATCCTGGAAGAGTTCTGCCGGCTGACCGGCCTGAGAA TCCAGCCTAGAAAGTGCCACGGCTTTCTGATCCAGAAGATTCAGCGGGCCAGATCCGTGAACCTGTGCAAGCCTTGGATCGTGT GCGGCGAGGAACTGCACATGGTCGGACCTGAGGAAAGCGTGTCCTACCTGGGCATGAAGGTGTCCCCATGGCACGGCATCATG GAACCCGATCCTGTGGAACGGCTGTGCAACTGGATCAGCTCTATCGGCAGAAGCCCTCTGAAGCCTTCTCAGAAAGTGCGGATG CTGAACGTGTACGCCGCTCCTAGAATGACCTACCAGGCCGATCATGGCGGCCTGGGACCTATCGTGCTGAATGTGCTGGATGGC ATGATCCGGAAGGCCGTGAAAGTGTGGCTGCATCTGCCTCTGTGTACCTGCGACGGCCTGCTGTACTCCAGATGTCAAGACGGT GGCCTGGGCATCGTGAAGCTGGCCTGTCAGATCCCTAGCATCCAGGCCAGACGGGTGTACAGACTGTGGCACAGCAAAGAAGC CATCACCAGAGTCGTGACCCGGCGGACAGTTGAGGCCGAAGAGTATAGAGGCATGTGGCTCAGAGCCGGCGGATCTGAAGCA GGACTTCCTCCACTGGAAGATAGAGAAGAGGGCGCCGTGCAGTGTACCGATACAGCTGGCTCTGTGAAGCCCAAGAATCCTGTG ATCCCCGACTGGCGGAGAGCCGAGTTTCTGAAGTGGCAGAATCTGACAGCCCAAGGCGTGGGCGTGCAAGTGTTTGGCGGCGA CAAGAACTCCAACCACTGGATGGCTAACCCCGAGACACTGGGCAGCAAAGAGCGGCACTATATCGCCGGACTGCAGCTGAGAG CCAACGTGTACCCTACAAGAGAGGCCCTGTCTAGAGGCAGACCCGACCTGCCTAAAGTGTGCAGACAGTGTCTGGCCGGCACAG AGTCTTGTGCCCACATCCTGGGACAGTGCCCTGCCGTGAAGGACAGCAGAATTCGGAGACACCACAAGCTGTGCGATCTGCTGG CCTCTGAGGCTGAATCAGCCGGATGGACCGTGATCAGAGAGATGTGCTGCAGAACCAGAGCTGGCGCCCTTAGAAGGCCTGAC CTGGTGTTTGTGAAAACCGGCTTCGCCCTGGTGGTGGACGTGACCGTCAGATACGAGATGGCCTACGATACCCTGATGGGAGCC GCCGCTGAGAAGGTGGCCAGATACACACCCATCACACCCTACGTGGCCATGACACTGAAGGCTCGGAGAGTGAAGGTTTTCGGC TTCCCACTGGGAGCCAGAGGCAAATGGCCTGGCAGCAACGACAGACTGCTGAAGGCCATGGGAGTTGGCGGCGGAAGAAGAA AGCAGCTGGCCAAGCTGTTCTCCAGACGGGCCCTGCTGTATAGCCTGGACGTGCTGAGAGACTTCTACAGAGCCGAGGGCGAA ACCGGCGACCTGGATGATGAATCCGTGGACGACCACCTGTGA (SEQ ID NO: 32)

R2-1_TG	**5' UTR**	GTCTAGTTACAACTGGGCATCGCTGCAGAGATCGCACCTCCTCGTGGTCCCGCTGGTAGCCCTTCGAAGGGTGACTAAGTCGATC TCTGCCCCAGGTACGGAGCCGTTGGGACTCACCAGTCCAACGTAACTCCTGCCTAAATTCGGTGAAACAAATTCCTCGGTAAAAA GCCCC (SEQ ID NO: 33)

	CDS	ATGGCTTCTTGCCCGAAACCTGGCCCCCCGGTTTCAGCAGGGGCAATGAGTTTGGAAAGTGGACTGACCACCCACTCCGTTCTCG CCATCGAACGTGGTCCCAATTCGTTGGCAAATTCCGGATCAGACTTTGGGGGGGGGGGGTCTGGGGCTACCGTTACGCCTATTGA GGGTATCGGTCGGCACTCAGACCTCCCGCTCCGACTGGGTAGACCTGGTGTCCTGGAGCCACCCAGGACCCACGTCTAAGTCCC AGCAGGTTGACCTGGTGTCTTTATTTCCTAAACACCGGGTTGACCTGTTATCCAAAAACGACCAGGTAGACCTGGTGGCTCAATT TTTACCATCTAAATTTCCCCCCAATTTGGCAGAAAATGATTTGGCTTTGCTGGTGAACTTAGAGTTCTACAGATCGGATTTGCATG TGTATGAGTGTGTTCATTTTGCTGCACATTGGGAGGGATTAAGTGGTTTGCCTGAGGTGTATGAACAACTTGCACCACAACCGTG TGTGGGAGAAACTTTACATTCTAGCCTCCCACGAGACAGTGAACTGTTTGTGCCTGAAGAGGGGAGCAGCGAGAAGGAGAGCG AGGACGCGCCAAAAACATCTCCTCCGACGCCTGGGAAACATGGTTTGGAACAGACTGGGGAGGAAAAAGTGATGGTGACTGTT CCTGACAAAAATCCACCTTGTCCTTGCTGTGGTACCCGGGTAAACTCTGTGTTGAATCTGATTGAACATCTGAAAGTGTCACACG GGAAAAGGGGGGTTTGTTTTCGGTGTGCAAAATGTGGAAAGGAAAATAGTAACTATCACAGTGTTGTTTGTCATTTTCCAAAAT GCAGGGGTCCAGAGACGGAGAAAGCCCCAGCTGGGGAGTGGATTTGTGAGGTATGCAACAGAGATTTTACAACCAAAATTGGC CTGGGACAACACAAGAGATTGGCACACCCAGCAGTGAGAAATCAGGAAAGGATCGTTGCTTCCCAACCGAAAGAAACATCAAA TAGAGGTGCTCACAAAAGGTGCTGGACAAAGGAGGAGGAAGAATTACTAATAAGACTGGAGGCTCAGTTCGAGGGAAACAAA AATATTAATAAGCTTATTGCAGAACACATAACCACCAAAACAGCTAAGCAGATCAGTGACAAAAGGCGATTGCTGTCCAGAAAG CCAGCAGAGGAGCCACGTGAGGAGCCTGGAACGTGTCATCACACCAGGAGAGCAGCTGCGAGCCTGAGAACGGAGCCTGAGA TGAGTCATCACGCCCAGGCAGAGGACAGAGATAATGGACCTGGGAGACGCCCTCTGCCAGGCAGGGCAGCTGCCGGAGGGAG AACAATGGACGAGATAAGACGCCACCCTGATAAGGGCAACGGACAGCAGAGACCCACCAAGCAAAAATCAGAAGAACAGCTGC AGGCTTACTATAAAAAGACACTAGAGGAACGACTTTCAGCTGGGGCACTTAACACCTTCCCCCGAGCATTCAAGCAGGTAATGG AAGGCCGGGATATAAAGCTAGTAATCAATCAGACAGCGCAGGACTGCTTCGGATGCCTGGAATCCATAAGCCAAATAAGAACG GCAACCCGAGATAAAAAGGACACGGTGACCCGGGAGAAACACCCAAAGAAACCTTTTCAGAAGTGGATGAAGGACAGAGCAAT CAAAAAAGGTAATTATCTTCGGTTCCAGCGTTTATTTTATCTTGATAGAGGGAAACTGGCTAAAATCATTTTAGATGATATTGAAT GCTTGTCTTGTGACATACCACTCAGTGAAATTTATTCGGTTTTTAAAACAAGATGGGAAACAACTGGTAGCTTTAAAAGCCTTGG GGACTTTAAAACTTACGGGAAGGCTGACAACACTGCCTTCAGAGAATTAATTACGGCTAAAGAAATTGAGAAAAATGTGCAGGA AATGAGCAAAGGCTCGGCTCCCGGTCCAGACGGGATTACTCTTGGGGACGTCGTAAAGATGGATCCCGAGTTTTCCCGGACCAT GGAGATTTTCAATTTATGGTTAACAACTGGTAAAATCCCGGACATGGTGAGGGGGTGCAGAACCGTTTTGATTCCAAAATCATCA AAGCCGGATCGTTTGAAAGACATTAATAACTGGAGACCTATCACGATCGGTTCCATCTTGCTGAGACTGTTCTCCAGGATTGTAA CAGCTAGGCTGAGCAAAGCGTGCCCCCTGAACCCAAGGCAAAGAGGCTTTATCAGAGCGGCGGGATGCTCTGAAAACTTAAAA CTCCTGCAAACTATAATTTGGTCGGCCAAAAGAGAACACAGACCACTGGGTGTTGTATTCGTGGACATCGCCAAGGCTTTTGACA CCGTAAGCCACCAGCACATCATTCATGCTTTGCAGCAAAGAGAGGTGGATCCCCACATCGTCGGTCTGGTGAGCAATATGTACG AGAACATCAGTACGTATATCACCACAAAGAGGAACACACACACAGACAAAATCCAGATCCGGGTTGGAGTAAAGCAGGGTGAC CCGATGTCGCCCCTTTTATTTAACCTGGCAATGGACCCTCTATTATGCAAGCTGGAAGAGAGTGGCAAAGGATACCACCGAGGAC AGAGCAGCATCACAGCGATGGCATTTGCAGACGATCTGGTTTTGCTGAGCGACTCCTGGGAAAATATGAATACAAATATTAGCA TACTGGAGACCTTCTGCAATCTGACCGGTCTCAAAACACAGGGGCAAAAGTGCCACGGCTTTTACATCAAGCCGACAAAGGACT CTTACACCATCAATGACTGCGCTGCCTGGACTATCAACGGCACACCCCTGAACATGATCGACCCCGGCGAATCTGAGAAATACCT CGGCCTGCAGTTTGACCCGTGGATTGGAATAGCAAGGTCCGGTCTCTCCACAAAACTAGATTTTTGGCTTCAGCGGATCGATCAA GCACCACTTAAACCTCTGCAGAAAACTGATATTCTCAAAACATACACCATCCCTCGGCTGATCTACATAGCTGACCACTCAGAAGT GAAAACTGCACTACTCGAAACCCTTGACCAGAAGATCCGGACAGCGGTCAAGGAATGGCTTCACCTACCTCCGTGCACCTGCGAT GCCATCCTGTACTCGAGCACGAGAGACGGCGGTTTGGGCATCACCAAATTGGCAGGACTGATCCCCAGCGTGCAGGCCCGTAGA CTGCATCGGATCGCACAGTCATCTGACGATACGATGAAATGCTTCATGGAAAAAGAGAAAATGGAACAGCTGCATAAGAAATTG TGGATTCAAGCTGGAGGGGACAGAGAGAACATACCCTCGATTTGGGAAGCACCACCGTCGAGTGAACCACCAAACAACGTGAG CACAAATTCGGAATGGGAAGCACCGACCCAGAAAGATAAATTTCCAAAGCCTTGCAATTGGAGGAAAAACGAATTCAAAAAATG GACCAAATTGGCATCCCAAGGCCGCGGAATTGTAAATTTTGAAAGAGACAAAATTAGTAACCATTGGATCCAATACTACAGACG CATACCTCACAGGAAACTCCTCACTGCACTACAACTCAGGGCCAACGTTTACCCCACGAGAGAATTTCTAGCCAGGGGTAGACAA GACCAATACATCAAGGCGTGTAGGCACTGCGATGCGGACATTGAATCCTGCGCCCACATCATCGGCAACTGCCCAGTGACACAG GACGCCCGAATCAAGAGGCACAATTACATCTGCGAACTGCTTCTCGAGGAGGCGAAGAAGAAGGACTGGGTAGTGTTCAAGGA ACCGCACATAAGGGATTCCAACAAGGAACTGTACAAACCTGACCTGATATTTGTGAAGGATGCCCGTGCACTTGTCGTGGATGT GACAGTACGGTATGAAGCAGCCAAATCATCGCTGGAGGAAGCCGCTGCAGAGAAAGTGAGAAAGTACAAACACCTGGAAACG GAAGTAAGACATCTCACGAATGCAAAGGACGTTACTTTTGTGGGCTTTCCCCTAGGAGCGCGGGGGAAATGGCACCAAGATAAC TTTAAACTTTTGACTGAGCTTGGCCTCTCCAAATCGAGGCAAGTGAAAATGGCAGAGACTTTTTCCACAGTAGCGCTCTTTTCATC TGTGGACATTGTACATATGTTTGCCAGTAGGGCCAGAAAATCTATGGTTATGTAA (SEQ ID NO: 34)
	3' UTR	TTCAGGTTATTTAGATGCTTAGTTTTTGTACCTTTCTTGTTTTGTTTAGGATTTTGATAGTGTTAGTATTTTTATATTTTTGTACGATT GCATAATGTTCTTTTTTATACAGTTCTGTTTTAATAAAATAGACGATAGCTAGAGACGTTAGGGCAGCCACAAGCCAGTTAGGTA GCGGATAGTAGGTAGGAACAGACTTTTACTATTTCATAACGCGTCAATTACCACCTGATTTGGACCAATTCACGGGATTTGTCCA AGGTGGACGGGCCACCTTTACTTAACCCGGAAAAGGAACATATATAATTTATGTGTGTTCGATAAA (SEQ ID NO: 35)
	Codon optimized	ATGGCCAGCTGTCCTAAGCCTGGACCTCCTGTTTCTGCCGGCGCTATGTCTCTGGAAAGCGGCCTGACAACACACAGCGTGCTGG CCATTGAGAGAGGCCCTAACAGCCTGGCCAATAGCGGCAGCGATTTTGGAGGCGGAGGACTGGGACTGCCTCTGAGACTGCTG AGAGTGTCTGTGGGCACCCAGACCAGCAGAAGCGATTGGGTTGACCTGGTGTCTTGGAGTCACCCCGGACCTACCAGCAAGTCT CAGCAGGTTGACCTCGTCAGCCTGTTTCCTAAGCACAGAGTGGACCTGCTGTCCAAGAACGACCAGGTGGACCTGGTGGCCCAG TTCCTGCCTAGCAAGTTCCCTCCAAACCTGGCCGAGAACGATCTGGCCCTGCTCGTGAACCTGGAATTCTACAGATCCGACCTGC ACGTGTACGAGTGCGTGCACTTTGCCGCTCACTGGGAGGGACTGTCTGGACTGCCAGAGGTGTACGAACAGCTGGCTCCTCAGC CTTGTGTGGGCGAGACACTGCATAGCAGCCTGCCTAGAGACAGCGAGCTGTTTGTGCCTGAGGAAGGCAGCAGCGAGAAAGAG TCTGAGGACGCCCCTAAGACAAGCCCTCCTACACCTGGAAAGCACGGCCTGGAACAGACCGGCGAAGAGAAAGTGATGGTCAC CGTGCCTGACAAGAACCCTCCTTGTCCTTGCTGCGGCACCAGAGTGAACTCCGTGCTGAACCTGATCGAGCACCTGAAGGTGTCC CACGGCAAACGGGGCGTGTGCTTCAGATGTGCCAAGTGCGGCAAAGAGAACAGCAACTACCACAGCGTCGTGTGTCACTTCCCC AAGTGCAGAGGCCCCGAGACAGAAAAAGCTCCTGCCGGCGAGTGGATCTGCGAAGTGTGCAACAGAGACTTCACCACCAAGAT CGGCCTGGGCCAGCACAAGAGACTGGCTCATCCTGCCGTGCGGAATCAAGAGCGGATTGTGGCCAGCCAGCCTAAAGAGACAA GCAACAGAGGCGCCCACAAGCGGTGCTGGACCAAAGAGGAAGAGGAACTGCTGATCCGGCTGGAAGCCCAGTTCGAGGGCAA CAAGAACATCAACAAGCTGATCGCCGAGCACATCACCACAAAGACCGCCAAGCAGATCAGCGACAAGCGGAGGCTGCTGAGCA GAAAGCCTGCCGAGGAACCCAGAGAGGAACCCGGAACCTGTCACCACACAAGAAGGGCCGCTGCCAGCCTGAGAACAGAGCCT GAGATGTCTCATCACGCTCAGGCCGAGGACAGAGACAATGGCCCTGGAAGAAGGCCTCTGCCTGGTAGAGCTGCTGCTGGCGG CAGAACCATGGACGAGATTAGACGGCACCCCGACAAAGGCAACGGCCAGCAGAGGCCAACAAAGCAGAAGTCCGAGGAACAG

		CTGCAGGCCTACTACAAGAAAACACTGGAAGAGAGACTGAGCGCTGGCGCCCTGAACACCTTTCCTAGAGCCTTCAAGCAAGTG ATGGAAGGCCGGGACATCAAGCTGGTCATCAACCAGACAGCCCAGGACTGCTTCGGCTGCCTGGAATCCATCAGCCAGATCAGA ACCGCCACCAGAGACAAGAAAGACACCGTGACCAGAGAGAAGCACCCCAAGAAACCCTTCCAGAAATGGATGAAGGACCGCGC CATCAAGAAGGGAAACTACCTGCGGTTCCAGCGGCTGTTCTACCTGGACAGAGGCAAGCTGGCCAAGATCATCCTGGACGACAT CGAGTGCCTGAGCTGCGACATCCCTCTGAGCGAGATCTACAGCGTGTTCAAGACCAGATGGGAGACAACCGGCAGCTTCAAGAG CCTGGGCGACTTCAAGACATACGGCAAGGCCGACAACACCGCCTTCAGAGAGCTGATCACCGCCAAAGAAATCGAGAAGAACG TGCAAGAGATGAGCAAGGGCAGCGCCCCTGGACCTGATGGAATCACACTGGGCGACGTGGTCAAGATGGACCCCGAGTTTAGC CGGACCATGGAAATCTTCAACCTGTGGCTGACCACCGGCAAGATCCCCGATATGGTTCGAGGCTGCAGAACCGTGCTGATCCCC AAGAGCAGCAAGCCCGACAGACTGAAGGATATCAACAACTGGCGGCCCATCACCATCGGCAGCATTCTGCTGAGGCTGTTCAGC AGGATCGTGACCGCCAGACTGAGCAAGGCCTGTCCTCTGAACCCTCGGCAGAGAGGCTTTATCAGAGCCGCCGGATGTAGCGA GAACCTGAAGCTGCTGCAGACCATCATTTGGAGCGCCAAGAGAGAGCACAGACCCCTGGGCGTCGTGTTCGTGGATATCGCCAA GGCCTTCGACACCGTGTCTCACCAGCACATCATTCACGCCCTGCAGCAGAGAGAGGTGGACCCTCATATCGTGGGCCTCGTGTCC AATATGTACGAGAACATCAGCACCTACATTACCACCAAGCGGAACACCCACACCGATAAGATCCAGATTAGAGTGGGCGTGAAG CAGGGCGACCCTATGAGCCCTCTGCTGTTCAATCTGGCCATGGATCCACTGCTGTGCAAGCTGGAAGAGTCCGGCAAGGGCTAT CACAGAGGCCAGTCTAGCATCACAGCCATGGCCTTCGCCGACGATCTGGTGCTGCTGTCTGACAGCTGGGAGAACATGAACACC AACATCTCTATCCTGGAAACCTTCTGCAACCTGACCGGCCTGAAAACCCAGGGACAGAAGTGCCACGGCTTCTACATCAAGCCCA CCAAGGACAGCTACACCATCAACGATTGTGCCGCCTGGACCATCAATGGCACCCCTCTGAATATGATCGACCCCGGCGAGAGCG AGAAGTACCTGGGCCTGCAATTCGACCCCTGGATCGGAATTGCCAGATCCGGCCTGTCCACCAAGCTGGATTTCTGGCTGCAGC GGATCGATCAGGCCCCACTGAAGCCTCTGCAGAAAACCGACATCCTCAAGACCTACACAATCCCCAGGCTGATCTATATCGCCGA CCACAGCGAAGTGAAAACAGCCCTGCTGGAAACCCTGGACCAGAAAATCCGGACCGCCGTGAAAGAGTGGCTGCATCTGCCTCC ATGCACCTGTGACGCCATCCTGTACTCTAGCACCAGAGATGGCGGCCTGGGAATCACAAAACTGGCCGGACTGATCCCCTCCGT GCAGGCTAGAAGGCTGCACAGAATTGCCCAGAGCAGCGACGACACCATGAAGTGCTTTATGGAAAAAGAAAAGATGGAACAGC TCCACAAGAAGCTGTGGATCCAGGCTGGCGGCGACAGAGAGAACATCCCCTCTATTTGGGAAGCCCCTCCTAGCAGCGAGCCTC CTAACAACGTGTCCACAAACTCCGAGTGGGAAGCTCCCACACAGAAGGACAAGTTCCCCAAGCCTTGCAATTGGCGGAAGAACG AGTTCAAGAAGTGGACCAAGCTCGCCAGCCAAGGCAGGGGCATCGTGAACTTCGAGCGGGACAAGATCAGCAACCACTGGATT CAGTACTACCGGCGGATCCCTCACAGAAAGCTGCTGACAGCACTGCAGCTGCGGGCCAACGTGTACCCTACCAGAGAATTCCTG GCCAGAGGACGGCAGGACCAGTACATCAAAGCCTGCAGACACTGCGACGCCGATATCGAGTCTTGCGCCCACATCATCGGCAAC TGCCCCGTGACACAGGATGCCCGGATCAAGCGGCACAACTACATCTGCGAACTGCTGCTCGAGGAAGCCAAGAAAAAGGACTG GGTCGTGTTCAAAGAGCCCCACATCCGGGACAGCAACAAAGAGCTGTACAAGCCCGATCTGATCTTCGTGAAGGACGCTAGAGC CCTGGTGGTGGACGTGACCGTTAGATATGAGGCCGCCAAGTCTAGTCTGGAAGAGGCTGCCGCTGAGAAAGTGCGGAAGTACA AGCACCTCGAAACCGAAGTCCGGCACCTGACCAACGCCAAGGATGTGACCTTCGTGGGCTTTCCACTGGGCGCCAGAGGAAAGT GGCACCAGGACAACTTCAAACTGCTGACTGAGCTGGGCCTGAGCAAGTCCCGGCAAGTGAAGATGGCCGAGACATTCAGCACA GTGGCCCTGTTCAGCTCCGTGGACATCGTGCACATGTTCGCCTCCAGAGCCAGAAAGTCTATGGTCATGTGA (SEQ ID NO: 36)
R2-1_TGut	**5' UTR**	CTGGGGACCGTGGTTACAACCCGGGCTTAGCTGCAGAGACAGTACCTCCCCGTGGTTCCCGCCGGACCCCGTAACATCGGGTGA CTGAATCTGTCTCTGCCCCGGGAGTAGTTCCTCCTTGCCCTATTGACCAGCGGTCGCCGGCTGCTCAATAGTATTCTAGGCGTGAA ATATAGCGATAGTCCTAGTGGTTGTCTTACTGGGCCATAGCCCCTTGCTTCAGGGGTCATTCGCGAAGTCTCTCAGGAGAACTGG GGGTGGTGTTCTTCTGGGTATAGCTAAACCCCCTAGACTGTGTCCGATCC (SEQ ID NO: 37)
	CDS	ATGGGGTCCTGGATCGTGAATTTCGTTTCGGTGGCGACTCAGACGGGAGAATTCCCTGTGGATACGGCCAGGAGGGCACCTGTG CCGGTAACATCATACCCTGAGTCGGAATGCCACNTACCGTTGCCCCTGACATTTTGTAACTCGGATGTGACTATTTGGGGAGGGG TTCGCCCTGAACCGGTGGACTGCTTGGGWGATCTTCCRGAGGYGTATGATGCACTCCCAGGGGTGGCTGGGCCTCGGGAAYCG GTGGGTGGGAGCCCGCCGGGRGAAGGGGTCAGGTCGCCAGGGATTGCGTCRCCCTCTGGTACTGCGGTCCAACATGATTTTGG GAGTCCCATCCTCGTACCGGGAGCCGAAGCCGCCGAGGTNTCTACCCCGGTAGTGAAGGTTCCNCAAGACCATCCAGCATGTCC NTGCTGTGGTACGAGGGTGGTGAAAGTAACGGCGTTGTCAGAACATCTTAGGAGGGCCCACGGTCGGAAACGGGTCCTATTTC AGTGCTCCCGATGCGGGAGGATGAATGAGAAACATCATAGCATCGCGTGCCATTTCCCGAAGTGCCGGGGGCCCCCAGTTGAG GAGGGTCCCCTGGGTGCACCCGAGTGGTGCTGTGAGGAGTGCGGGCAGAAATTTAACACCAAAAGCGGCCTGTCTCAGCACAA AAGATCTGTGCACCCACTTACGAGGAATGTGGAACGGATAGAGGCAGCTCGTCCGAAAGGAAAAGGGAAGCGTGGTGCCCACA AAGGCTGTTGGACCGAGGCGGAGGTGGCCCAGCTGATTGAACTGGAGGGGAGATTCAAGAACCAGCGATTTATCAACAAGCTG ATCGCGGAGCATTTGCCATCGAAATCGGCGAAGCAGATCAGCGATAAGAGAAGGCAGCTGGCGGCAGCGACCAAGACATCGTC GCCCGAGAAGAGGGTAACGTCATCAACGAGTGGGGAGTCCTCCCCTGAAGTGGAAAAGGTGGAGGGTATCAAGAGAGAATAT AGAAGGCGTGTTGGAGAGTGGTTGTGCGCTGGGTCCCTGMAGGACCAGACNTCGTTCCAGAAGATCTTGGAAGATGTGGAGA GCGGCTCCGAGATTGTCACCGGTCCGCTGGAGGAACTGGCCTCCTTTGCGAGGGGGAAGCTCGCGGCAGCTAGAGTGCGACAT CATCGTAAGCACCCAGCTGAGGCCGTGCCTGCGCGAGAGGAGCAGAGGTGGATGAAGCGCAGGGTGGGTCGTCGGGGCTTGT ACCTCAGGTTCCAGCGGCTTTTTGCNTTGGATCGCAGRAAGCTTGCTGGGATCATCCTCGACGATGTCGAGTCCATCAAGTGCCC CCTTCCGATGGAAGAAGTCGCTGACGTCTTCAGGAGAAGGTGGGAGGAGGTCGCCCCCTTTACCGGCTCGGGCTCATTCCGAAG TTTGGGGAAGGCTGACAACGGTGCCTTCAAGCCCATGATCTCCGCYAAGGAGGTCATGAAGAACGTCRCGGAGATGTCTCGACG CTCCGCGCYGGGWCCCGAYGGCCTCTCCCTGCGGGATCTGATGAAGATTGATCCCCAGGGCAGCCGCATGGCTGAATTGTTCAA CCTGTGGCTGTTGGCAGGACGGGTCCCGGACCAAGTGAAGGCGGGCCGAACGGTCCTGATCCCNAAGTCGGCCGATCCCGGGA AGATCGGGAACATTGACAACTGGCGGCCCATCACCATCGGGTCCGTTATNCTCAGAATGTTCTCTCGGATTTTGAGCGCTAGACT GCGGCGAGCATGCCCCATTAATAGAAGGCAAAGGGGGTTTATAGCAGCCCCTGGCTGCTCGGAAAATCTGAAGCTTCTCCAGGC GCTCATCAAGAGCGCGAAACGAGATCATAGGACCCTTGGAGTCGTGTTTGTCGATTTGGCTAAGGCCTTCGACTCCGTGAACCAC CAGCATATCTTCCAAGTCCTGGTCCAGAAGGGTGTCGATGGGCATATTATCGACATCCTAAGAGACCTGTATACCAACGCTGGAA CGTATCTGGAGTCAGGTTCCCAGCGATCGGGATTTATTAAGATCCTCAGGGGAGTGAAACARGGGGACCCACTCTCTCCCATCCT GTTCAATCTTGCATTAGATCCTTTGCTGTGCCGCCTGGAAGATCGGGGCCTCGGTTATAAGTATGGAGACCAACAAATAWCATC GTTGGCATTTGCGGATGATCTCGCCCTGCTCAGCGACTCTTGGGAGGGCATGCAGCAGAGTATTCGGGTGGTAGAGGAATTTTG TCAACGGACCGGGCTGCGGGTTCAAGCGCCGAAATGCCACGGGTTTTTGATCAGGCCAACTAAAGAGTCATATACCATCAATGA CTGTGACCCGTGGACGATTGCAGATATGCAATTGGATATGATCGATCCGGGCAGTTCCGAGAAGTATCTTGGCCTAGGGATAGA CCCATGGATTGGTCTATCGAGACCGGAACTGTCCGAGGTGCTGACCCGCTGGGTGAAGAACATCGGGGGCGCCCCTTTGAAGCC ACTCCAGAAGGTGGACATCTTGAGGAGCTACGCCCTNCCAAGGCTGCTGTTCATTGCGGATCACGCAGGCCTGAGCGCCACCTG TTTGCATTCCCTGGACCTTTCGATAAGATCTGCCGTCAAGGGCTGGTTACATCTACCGCCTAGTACGTGTGACGCTATTATTTACG

		TCAGCTACAAGGACGGCGGGCTGGGTCTTCCCCGTCTGGCGAGCCTAATTCCAAATGTACAGGCTCGCAGGTTGGTGCGGATCG CCCAATCGGAGGATGATGTCATCAGGAGTGTGGTACTCCAGGAGGGTATCCAGGAGGAGATCCGGAAGGTCTGGATCTCGGCT GGGGGGCGACCGGAGAAGGTTCCATCTGTGACGGGGGAGTTCCCAGTGATGGAGGCTCAGGCGGCTGACGAGGCCCTATCCG AGTGGGAGAGGCGAGCTCCACGAACCATCTATCCCATTCCCTGTAAGTGGAGGAAGAGAGAAATGGAGAATTGGACCAATCTA AAATCGCAAGGCCACGGGATTCGGAATTTTGAAAATGACCGAATCAGTAATGATTGGCTCCTGCATTATGGCCGCATTCCCCACC GCAAACTAATAACAGCTATCCAGTTGCGGGCCAATGTCTATCCCACWCGGGAGTTTTTGGCCCGCGGCCTGGGCGAGGGCGCA CCCAGGGGATGTAGGCACTGTCCCGCGGAGTGGGAATCTTGTTCCCACATAATTGGCTACTGCCCGGCTGTCCAGGAGGCCAGG ATCAAAAGGCATAATGACATCTGTGGTGTGCTGGCTGAAGAGGCNAGAAAGCTGGGATGGGTGATATTTATAGAGCCCCATCTC AGAGATAACACCAATGAGCTCTTCAAGCCAGATTTGGTTTTGGTGAAGGGATCCTGTGCGAAGGTAGTGGATGTAACCATCCGC TACGAGAGTGGGTTAACCACCTTGAGTGACGCCGCGGCAGAAAAGGCTAGGAAGTATCAACATCTGGCAGGGGAGGTGCGGG CCCTAACATCGGCCACTACTGTAGACTTCCTGGGTTTCCCTATTGGCGCTAGAGGGAAGTGGTACGTTGGTAATAATGGACTCCT TTCCGACCTTGGGTTCTCCACTAGCCGTGTAGTGCGGATAGCGAGGGCCCTCTCTAAAAAGGCTCTCCTATCGTCCGTGGACATT ATACATATTTTTGCGTCTCGCGCTAGACAGGCCCAAACGTCCGAGTAG (SEQ ID NO: 38)
	3' UTR	GGGGCTTGGCATTTCTCATTGCCTGCTCCTGAAAGGATATGGGTCCTGCGTCGCGTGGTAGGCAGACCCATTCGTCCGAGTAGG GGGCTTGGCAGTNTCCATTGCCTGTGCCCGAAAGGACGTGGGTCATCTGGTCTGTCTGCCTACACCTCTCTAGACTTGTAACATC TAGTCTGTCAACAAGATCAAAATTCTTCACACAGACGACCGAGCTTGCTCAGTCTTCCTGTACCCGCAGAATTTTGCTCTTGCTCTC CTTTGGCTGTGTCCTGGACGTGGGACTATTCCATCTCGTCCCAAATGCCGCGTCCAATTATACCGGATTTGACAAAGCGGACGGC CCGCTTTATAAGCCGGAAAAGGTGCCTTGTAAAATTGCAAGGTTCATTAAATAG (SEQ ID NO: 39)
	Codon optimized	ATGGGCAGCTGGATCGTGAACTTCGTGTCCGTGGCTACCCAGACCGGCGAGTTCCCTGTGGATACAGCTAGAAGGGCTCCCGTG CCTGTGACAAGCTACCCTGAGAGCGAGTGCCATCTTCCTCTGCCTCTGACCTTCTGCAACAGCGACGTGACAATCTGGGGCGGAG TCAGACCTGAGCCTGTGGATTGTCTGGGCGATCTGCCCGAGGTGTACGATGCACTTCCTGGCGTTGCCGGACCTAGAGAGTCTG TTGGAGGAAGCCCTCCTGGCGAGGGCGTTAGATCTCCTGGAATCGCCTCTCCTAGCGGCACAGCCGTGCAGCACGATTTCGGAA GCCCTATTCTGGTGCCTGGCGCCGAAGCTGCCGAAGTGTCTACACCTGTGGTCAAGGTGCCCCAGGATCACCCTGCCTGTCCTTG TTGTGGCACCCGGGTCGTGAAAGTGACAGCCCTGTCTGAGCATCTGCGGAGAGCCCACGGAAGAAAGCGGGTGCTGTTCCAGT GTAGCAGATGCGGCCGGATGAACGAGAAGCACCACTCTATCGCCTGTCACTTCCCCAAGTGCAGAGGCCCTCCTGTGGAAGAAG GACCTCTGGGAGCACCTGAGTGGTGTTGCGAGGAATGCGGCCAGAAGTTCAACACCAAGAGCGGCCTGAGCCAGCACAAGAGA TCTGTGCACCCTCTGACCAGAAACGTGGAACGGATCGAAGCCGCCAGACCTAAAGGCAAGGGAAAGAGAGGCGCCCACAAAGG CTGTTGGACAGAGGCTGAAGTGGCCCAGCTGATTGAGCTGGAAGGCCGGTTCAAGAACCAGCGGTTCATCAACAAGCTGATCG CCGAACATCTGCCCAGCAAGAGCGCCAAGCAGATCAGCGACAAGCGGAGACAACTGGCCGCTGCCACAAAGACAAGCAGCCCC GAGAAGAGAGTGACCAGCAGCACATCTGGCGAGAGCAGCCCTGAGGTGGAAAAGGTGGAAGGCATCAAGCGCGAGTACAGGC GGAGAGTTGGAGAGTGGCTGTGTGCCGGCTCTCTGAAGGATCAGACCAGCTTCCAGAAAATTCTCGAGGACGTGGAAAGCGGC AGCGAGATCGTGACAGGCCCTCTGGAAGAACTGGCCTCCTTTGCCAGAGGCAAACTGGCTGCCGCCAGAGTGCGGCACCACAG AAAACATCCTGCTGAGGCCGTGCCTGCCAGAGAAGAACAGAGATGGATGAAGCGGAGAGTGGGCAGAAGAGGCCTGTACCTG AGATTCCAGAGACTGTTCGCCCTGGACAGAAGAAAGCTGGCCGGCATCATCCTGGACGACGTGGAATCCATCAAGTGCCCTCTG CCTATGGAAGAGGTGGCCGACGTTTTCCGGCGGAGATGGGAAGAAGTGGCTCCCTTTACCGGCAGCGGCTCCTTTAGATCTCTG GGCAAAGCCGACAACGGCGCCTTCAAGCCTATGATCAGCGCCAAAGAAGTGATGAAGAACGTCGCCGAGATGAGCAGAAGAA GCGCCCCTGGACCTGATGGCCTGTCTCTGAGAGATCTGATGAAGATCGACCCTCAGGGCAGCAGAATGGCCGAGCTGTTCAATC TGTGGCTGCTGGCCGGAAGAGTGCCCGACCAAGTGAAAGCCGGAAGAACCGTGCTGATCCCCAAGTCTGCCGATCCTGGCAAG ATCGGAAACATCGACAATTGGCGGCCCATCACCATCGGCTCCGTGATCCTGAGAATGTTCAGCCGGATCCTGAGCGCCAGACTG AGAAGGGCTTGCCCCATCAACAGACGGCAGCGGGGCTTTATTGCCGCTCCTGGCTGTAGCGAGAACCTGAAACTGCTGCAGGCC CTGATCAAGTCCGCCAAGAGAGATCACAGAACCCTGGGCGTCGTGTTCGTGGATCTGGCCAAGGCCTTCGACAGCGTGAACCAC CAGCACATTTTCCAGGTGCTGGTGCAGAAAGGCGTGGACGGCCACATCATCGACATCCTGAGGGACCTGTACACCAACGCCGGC ACCTACCTGGAATCTGGCAGTCAGAGAAGCGGCTTTATCAAGATCCTGCGGGGCGTGAAGCAGGGCGATCCTCTGTCTCCCATC CTGTTCAACCTGGCTCTGGACCCTCTGCTGTGCAGACTGGAAGATAGAGGCCTGGGCTATAAGTACGGCGACCAGCAGATTACC AGCCTGGCCTTCGCTGATGATCTGGCCCTGCTGAGCGATAGCTGGGAGGGAATGCAGCAGAGCATCAGAGTGGTGGAAGAGTT CTGTCAGCGGACCGGCCTGAGAGTGCAGGCCCCTAAATGTCACGGCTTTCTGATCAGGCCCACCAAAGAGAGCTACACCATCAA CGACTGCGACCCCTGGACAATCGCCGACATGCAGCTGGACATGATCGATCCAGGCAGCAGCGAGAAGTATCTCGGCCTGGGAAT CGACCCTTGGATCGGCCTGTCTAGACCAGAGCTGAGCGAGGTGCTGACCAGATGGGTCAAGAACATTGGCGGAGCCCCTCTGA AGCCCCTGCAGAAAGTGGACATCCTGCGGAGCTATGCCCTGCCTCGGCTGCTGTTTATTGCTGATCACGCCGGACTGTCCGCCAC ATGTCTGCATAGCCTGGATCTGTCCATCCGCAGCGCCGTGAAAGGATGGCTGCATCTGCCTCCAAGCACCTGTGACGCCATCATC TACGTGTCCTACAAGGATGGCGGACTGGGCCTGCCTAGACTGGCCTCTCTGATCCCTAATGTGCAGGCCAGACGGCTCGTCAGA ATCGCCCAGTCTGAGGACGATGTGATCAGATCCGTGGTGCTGCAAGAGGGCATCCAAGAGGAAATCCGGAAAGTCTGGATCTCT GCCGGCGGAAGGCCTGAGAAAGTGCCTTCTGTGACCGGGGAGTTTCCCGTGATGGAAGCCCAGGCTGCTGATGAGGCTCTGAG CGAGTGGGAAAGACGGGCCCCTAGAACAATCTACCCCATTCCTTGCAAGTGGCGGAAGCGCGAGATGGAAAACTGGACCAACC TGAAGTCCCAAGGCCACGGCATCCGGAACTTCGAGAACGACAGAATCAGCAACGACTGGCTGCTGCACTACGGCAGAATCCCTC ACCGGAAGCTGATCACCGCCATCCAGCTGAGAGCCAACGTGTACCCCACCAGAGAGTTTCTGGCTAGAGGACTCGGAGAGGGC GCTCCTAGAGGATGCAGACACTGTCCTGCCGAGTGGGAGAGCTGCAGCCACATTATCGGCTACTGTCCCGCCGTGCAAGAGGCC AGAATCAAGCGGCACAACGACATCTGTGGGGTGCTCGCCGAGGAAGCCAGAAAACTCGGCTGGGTCATCTTTATCGAGCCCCAC CTGAGAGACAATACCAACGAGCTGTTTAAGCCCGACCTGGTGCTGGTCAAGGGCAGCTGTGCTAAGGTGGTGGACGTGACCATC AGATACGAGTCCGGCCTGACCACACTGTCTGATGCCGCCGCTGAGAAGGCCAGAAAGTACCAACATCTGGCCGGCGAAGTGCG GGCCCTGACATCTGCAACCACCGTGGACTTTCTGGGCTTTCCCATTGGCGCTAGAGGCAAGTGGTACGTGGGCAACAATGGCCT GCTGTCCGATCTGGGCTTCAGCACCAGCAGAGTTGTGCGGATTGCTAGAGCCCTGAGCAAGAAGGCTCTGCTGAGCAGCGTGG ACATCATCCACATTTTCGCCTCTCGGGCCAGACAGGCCCAGACCTCTGAATGA (SEQ ID NO: 40)

R2-1_TSP	**5' UTR**	CTCCTGACTAACCTGATTTCGTCCGTGCGGCGGCGTTTTCTTTTCGCTCTCCGCTCGTCGAAATTTGCTGTAGTTGATTCGCTTTTCT TTGCGTTTTCTTCTACTTTCGCAGTTTTTTCTGCATTGCCACG (SEQ ID NO: 41)
	CDS	ATGTCAAACCGCCTTGCCAATACTGCTGCGGCTGGTGGGGTTCCAGAGAAAACCTCGGGAACTTTAGACATTCCTGGCCAACCCT CTTCATCCGGTGAAAAGCGTGCGATCTCTTACCCTGGTCCATTCGGTTGCAATTCGTGTTCGTTTACGAGTACGACTTGGCTCTCA TTGGAATTGCATTTTAAAAGCGTCCATAATATTCGTGACTTCGTCTTCCTCTGCTCTAAATGTAAAAAAAGCTGGCCATCGATCAA CTCCGTAGCTAGCCATTACCCTCGGTGCAAAGGTAGCGTCAAGGCTGCAGTTGTTCCTACATCTTTGGCGAATACGTGCACCACG

		TGCGGCTCAAGCTTCGGTACTTTCAGTGGTCTTCAACTCCATCGGAAAAGAGCACATCCGGACGTTTTTGCTGCTTCTTGTAGCAA AAAAACGAAGGCGCGTTGGTCTAACGACGAATTTACCCTTCTGGCGAGACTCGAAGCAGGTCTGGATCCAGCCTGTAAAAACAT TAACCAAGTACTAGCGGAAAGGTTAATGGAGTATAACATCACCAGAGGCGTAGAAATGATAAAAGGCCAACGTAGAAAAGATC AGTACAAAGCGCTCGTTCGTCAACTCCGGTCAAATTCTGAAACACAGCAATGTGTAGGTTTAGCCGGAAGTATGGATTCGAACGT ACCGGCCAACGATACATCGTCTTCCGTTGCATCAGAGGTCAGCATTACGTACCCTGAGTACGGGGCCGTGATGTCGTGCGACCTA ATTAAAGAAGCGACTGGTATGGCCATAGTTGACATCAACGAGTTGCAAAGCAACTTACGAAAAGCCTTCTTGTCCGGCCGCAAG CTTCCCATGAAGTTCCATGGAGCGCGTGAAACCGCCCAGAAGAAAATGGCCAACCCCCGTGTTGCGAAATTCAAGCGTTTCCAAC GGTTGTTTCGAAGCAACAGGAGGAAACTGGCCAGCCACATCTTCGACAAAGCCTCACTGGAGCAATTCGGTGGCAGCATCGATG AGGCATCTGACCATTTAGAAAAGTTCCTCTCCCGGCCAAGATTGGAGTCCGATTCTTATTCCGTGATAAGCGGTGATAAGTCAAT CGGAGTTGCACATCCAATTTTGGCCGAGGAGGTGGAATTGGAATTAAAAGCCTCCCGACCAACCGCTGTTGGTCCGGATGGAAT TGCACTGGAAGACATTAAAAAACTCAATACTTACGACATAGCCAGTCTTTTCAACCTCTGGCTAAAAGCTGGCGACCTACCCGCAT CGGTGAAAGCCAGTAGAACCATCTTTTTGCCCAAAAGCGACGGCACCACCGACATATCGAACTGTCGGCCAATCACAATCGCATC CGCCATGTATCGGCTGTTCAGCAGAATAATAACGCGACGTCTGGCAGCCAGGTTGGAATTGAACGTGCGGCAAAAAGCGTTCCG GCCTGAAATGAACGGCGTATTCGAGAACTCCGCCATTTTATACGCCCTCATCAAGGATGCTAAGGTCAGGTCAAGGGAAATTTGC GTAACTACGCTCGACCTTGCCAAGGCCTTTGACACGGTGCCCCACTCACGCATTTTACGAGCCCTGAGGAAAAATAATGTCGACC CGGAATCCGTCGACCTGATTTCGAAAATGTTAACGGGTACGACTTATGCAGAAATAAAAGGGCTCCAGGGCAAACTTATACCCAT TCGCAATGGAGTCAGGCAAGGTGACCCCTTGTCGCCCCTATTATTTAGTCTATTTATAGACGAGATAATAGGTCGCCTACAAGCC TGCGGCCCTGCCTACGATTTCCATGGCGAAAAAATTTGCATCCTGGCTTTCGCCGATGATCTGACGCTGGTGGCTGACAGCGCAG CTGGTATGAAGATCCTTCTAAAAGCGGCTTGTGACTTCCTGGAGGAATCTGGAATGTCACTTAATGCAGAGAAATGCCGCACTCT CTGTATTACAAGATCTCCCCGAAGCCGCAAGACTTTCGTCAACCCAGCTGCCAAATTCATCATCAGCGATTGGAAAACGGGTATC AGCTCAGAAATCCCCTCCCTGTGTGCGACGGACACCTTTCGTTTCCTGGGGCACACCTTCGATGGAGAAGGAAAGATCCACATCG ATACGGAGGAAATTCGATCCATGCTCAAATCGGTGAAGTCAGCTCCACTGAAACCGGAACAGAAGGTGGCTTTGATACGGTCAC ACCTTCTTCCCCGCCTTCAGTTCCTGTTTTCTACAGCTGAAGCTGACAGCCGGAAAGCCTGGTTGATCGATTCCATCATCAGGGGG TGTGTGAAGGAGATCTTGCACTCAGTGAAAGCTGGTATGTGCACTGATATCTTTTACATACCCTCTAGAGACGGTGGAATGGGAT TTACTTCCCTCGGGGAGTTTTCTCTTTTCAGCAGGCAGAAGGCACTCGCCAAGATGGCTGGATCGTCGGACCCCCTCTCGAAACG GGTTGCTGAATTCTTCATCGAAAGGTGGAACATCGCCCGTGACCCGAAAGTCATTGAAGCTGCTCGGCGCGTCTACCAGAAAAA ACGGTACCAACGCTTTTTCCAGACGTACCAGAGCGGTGGATGGAATGAATTTTCGGGAAACACTATTGGGAACGCCTGGTTGAC AAACGGCCGTGCCCGCGGAAGAAATTTCATAATGGCTGTGAAATTCCGTTCCAACACCGCAGCCACCCGGGCCGAAAACCTACG AGGCCGCCCCGGCACGAAAGAATGCCGGTTTTGCAAGAGTGCCACCGAAACTTTGGCACACATTTGCCAGAGGTGTCCGGCAAA TCACGGCTTGGTTATCCAGCGCCATGACGCAGTCGTAACATTCCTGGGGGAAGTGGCGCGGAAGGAAGGTTACCAGGTCATGAT AGAGCCTAAGGTGTCAACCCCGGTCGGCGCGCTCAAGCCCGACCTCCTACTCATCAAAGCCGACACTGCATTCATTGTGGATGTA GGCATTGCATGGGAAGGTGGACGCCCACTAAAGCTGGTCAACAAAATGAAATGTGACAAGTACAAAACTGCCATCCCGGCAATT TTGGAAACATTTCACGTTGGCCATGCTGAGACGTACGGCGTTATTCTGGGCAGCCGCGGATGCTGGCTCAAGAGCAACGACAAG GCGTTGGCATCAATTGGGCTCAATATCACACGGAAGATGAAAGAACACCTGAGCTGGTTGACGTTTGAAATTATATTTATAACTC AAATAAGCCGGATTTATAACTCATTCATGAAAAAATGA (SEQ ID NO: 42)
	3' UTR	GGTTTTTGTTTTCTTTTTTCCTTTTACCATTCTTGTTCCATTGTTGTTATTTGCTTTAATCCTGTATTTTACCGCCGGCAATTCCATTGT TATTATTACTGTTACTGTTATTATTGTTACTATTGTTTTTACTTTTACTTACTACTGTTATTATACTTTAATTCGTTAACTTACGTTATT GTTACCACTACTTACTTTGCTCTCTCGCAAACGTTCGTTGTTGTTTCTTTTGGACCAGGTTTAGAGAAATCGCACGCACAGCGGAA CTGGACCGCTTAAGCCAGAAATAGTAAAGTAACAA (SEQ ID NO: 43)
	Codon optimized	ATGAGCAACAGACTGGCCAATACTGCCGCTGCTGGCGGCGTGCCAGAGAAAACATCTGGCACCCTGGACATCCCTGGCCAGCCA TCTTCTAGCGGAGAGAAGAGAGCCATCAGCTACCCCGGACCTTTCGGCTGCAACAGCTGTAGCTTTACCAGCACCACCTGGCTGA GCCTGGAACTGCACTTCAAGAGCGTGCACAATATCCGGGACTTCGTGTTCCTGTGCAGCAAGTGCAAGAAGTCCTGGCCTAGCA TCAACAGCGTGGCCTCTCACTACCCCAGATGCAAGGGATCTGTGAAGGCCGCCGTGGTGCCTACATCTCTGGCCAACACCTGTAC CACCTGTGGCAGCAGCTTCGGCACCTTTTCTGGACTGCAGCTCCACCGGAAAAGGGCTCACCCTGATGTGTTTGCCGCCAGCTGC AGCAAGAAAACAAAGGCCAGATGGTCCAACGACGAGTTCACCCTGCTGGCCAGACTGGAAGCTGGACTGGATCCCGCCTGCAA GAACATCAATCAGGTGCTGGCCGAGCGGCTGATGGAGTACAATATCACCAGAGGCGTCGAGATGATCAAGGGCCAGAGAAGAA AGGACCAGTACAAGGCCCTTGTGCGGCAGCTGAGAAGCAACAGCGAGACACAGCAGTGTGTTGGCCTGGCCGGCAGCATGGAT TCTAACGTGCCAGCCAACGACACCAGCAGCTCTGTGGCCAGCGAAGTGTCCATCACATACCCTGAGTATGGCGCCGTGATGAGC TGCGACCTGATCAAAGAAGCCACCGGCATGGCCATCGTGGACATCAATGAGCTGCAGAGCAACCTGAGAAAGGCCTTCCTGAGC GGCAGAAAGCTGCCCATGAAGTTCCATGGCGCCAGAGAGACAGCCCAGAAAAAGATGGCCAATCCTAGAGTGGCCAAGTTCAA GCGGTTCCAGCGGCTGTTCCGGTCCAACAGAAGAAAGCTGGCTTCCCACATCTTCGACAAGGCCAGCCTCGAGCAGTTTGGCGG CTCTATCGATGAGGCCTCCGACCACCTGGAAAAGTTTCTGAGCAGACCCCGGCTGGAAAGCGACTCCTACTCTGTGATCAGCGGC GACAAGAGCATCGGCGTGGCCCATCCTATTCTGGCCGAGGAAGTGGAACTGGAACTGAAGGCCAGCAGACCTACAGCCGTGGG ACCTGATGGAATCGCCCTGGAAGATATCAAGAAGCTGAACACCTACGATATCGCCAGCCTGTTCAACCTGTGGCTGAAGGCAGG CGATCTGCCAGCCTCTGTGAAAGCCAGCCGGACCATCTTCCTGCCTAAGTCCGATGGCACCACCGACATCAGCAACTGCAGACCC ATCACAATCGCCAGCGCCATGTACCGGCTGTTCAGCCGGATCATCACCAGAAGGCTGGCCGCTAGACTCGAGCTGAATGTTCGG CAGAAGGCTTTCAGACCCGAGATGAACGGCGTGTTCGAGAACAGCGCCATCCTGTACGCCCTGATCAAGGACGCTAAAGTGCGG AGCCGCGAGATCTGCGTGACCACACTGGATCTGGCCAAGGCCTTCGATACCGTGCCTCACAGCAGAATCCTGAGAGCCCTGCGG AAGAACAACGTGGACCCTGAGTCCGTGGATCTGATCAGCAAGATGCTGACCGGCACCACCTACGCCGAGATCAAAGGACTGCA GGGCAAGCTGATCCCCATCAGAAACGGCGTCAGACAGGGCGATCCTCTGAGCCCTCTGCTGTTTTCCCTGTTCATCGACGAGATC ATCGGCCGGCTGCAGGCTTGTGGACCTGCCTATGATTTCCACGGCGAGAAGATCTGCATCCTGGCCTTCGCCGACGATCTGACAC TGGTGGCTGATTCTGCCGCCGGAATGAAGATCCTGCTGAAGGCTGCCTGCGACTTCCTGGAAGAGTCCGGCATGTCTCTGAACG CCGAGAAGTGCAGAACCCTGTGCATCACAAGAAGCCCCAGGTCCAGAAAGACCTTCGTGAACCCTGCCGCCAAGTTTATCATCA GCGACTGGAAAACCGGCATCAGCAGCGAGATCCCTAGCCTGTGTGCTACCGATACCTTCCGGTTTCTGGGCCACACCTTTGACGG CGAGGGCAAGATCCACATCGACACCGAAGAGATCCGGTCCATGCTGAAGTCCGTGAAGTCTGCCCCTCTGAAGCCCGAGCAGAA GGTGGCCCTGATTAGAAGCCATCTGCTGCCCAGGCTGCAGTTCCTGTTTTCTACAGCCGAGGCCGACTCTCGGAAGGCCTGGCTG ATCGACTCTATCATCCGGGGCTGCGTGAAAGAAATCCTGCACAGCGTGAAAGCCGGCATGTGTACCGACATCTTCTACATCCCCA GCCGCGACGGCGGCATGGGATTCACATCTCTCGGAGAGTTCTCCCTGTTCTCCAGACAGAAAGCCCTGGCCAAGATGGCCGGAA GCAGCGATCCACTGTCTAAGCGCGTGGCCGAGTTCTTCATCGAGCGGTGGAACATTGCCAGAGATCCCAAAGTGATCGAGGCCG CCAGACGGGTGTACCAGAAGAAGAGATACCAGCGGTTCTTCCAGACCTACCAGAGCGGCGGCTGGAATGAGTTCAGCGGCAAC ACAATCGGCAACGCTTGGCTGACCAACGGCAGAGCCAGAGGCAGAAACTTCATCATGGCCGTGAAGTTTCGGAGCAACACCGCC

		GCCACAAGAGCCGAGAATCTGAGAGGCAGACCCGGCACCAAAGAGTGCAGATTCTGCAAGAGCGCCACCGAGACACTGGCCCA CATCTGTCAGAGATGCCCTGCCAATCACGGCCTGGTCATCCAGAGACACGATGCCGTGGTCACCTTCCTGGGAGAAGTGGCCAG AAAAGAGGGCTACCAAGTGATGATCGAGCCCAAGGTGTCCACACCAGTGGGAGCCCTGAAACCTGACCTGCTGCTGATTAAGGC CGACACCGCCTTCATCGTGGATGTGGGCATTGCTTGGGAAGGCGGCAGACCACTGAAGCTGGTCAACAAGATGAAGTGCGACA AGTACAAGACCGCCATTCCTGCCATCCTGGAAACCTTCCACGTGGGACACGCCGAGACATACGGCGTGATCCTGGGATCTAGAG GCTGCTGGCTGAAGTCTAACGATAAGGCCCTGGCCTCCATCGGCCTGAACATCACCCGGAAGATGAAGGAACACCTGAGCTGGC TGACCTTTGAGATCATCTTCATCACCCAGATCTCCCGGATCTACAACAGCTTTATGAAGAAGTGA (SEQ ID NO: 44)

R2-1_ZA	**5' UTR**	CGACTTGAGAAGGTCTGGTTACAACTGGGCATAGCTGCAGAGATCGCGCCTCCTCGTGGCCCCGCTGGTAAGCCCTTAACAGGG TGACTAAGTCGATCTCTGCCCCAGTCCAGGAGCCGCTGGGTTTCACCAGCCCAGCGATTCCTTCCAAATTCGGTGAAACAAATTC CTCGGTAAAAGCCGCGTGGCTTATTGC (SEQ ID NO: 45)
	CDS	CTGAAACCTGGCCCCCCGGTTTCAGACAGGGGCAAAGAGTTCGGAAGTGGACTGACCACCCACCCCGAACCCGAGAGCGAATCT GGTCATGACCCAACTGTCCCAAATCCTGGTCCGTCTCTTGGAGCGGGGGAAGGTGCACAGCCACTACCCTTACTCAGGGTATCG GTGGGCACCCAAACCTGTGAAGAGGACTTTATAACATCTAGACCAACCAAATTACCCGGAATTGAATCAGAATTAGGCCCGCTG GTGAAGTTTTCTTTAGAGGTTTACAGGTCAGATCTTAAGGGGGATGTGCAATTTGAGGGGATTCATTTTCCAGATAATTGGGGG GTACTGGAGGGGTTTCCTGAGGTGTACGAACAACTGGCACCACAGCCAAACGGGGGAGACGAGTTAAATCATAGTCTCCCAGG GGACAGGGAGGGGGATGTACTTGAGAAGGATAGCAGCGAAAAGGAGAAGGAGGCTGCACCAGAGGCATTGCCCTCAGTGCAA AGGGCCCGCAGTGAACAGTTGCCAGATAACATCGTAAAGGTGACTGTTCCCGACAAAAATCCACCATGTCCCTGCTGTGGTGTCC GCTTAAACTCAGTGTTAGCTCTGATTGAACATCTGAAGGGCTCACACGGGAGGAGGAGGGTGTGCTTTAGGTGTGCCAAATGTG GGAGGGAGAATTTTAACCACCATAGTACTGTTTGTCATTACGCAAAGTGCAAAGGTCCACAGATTGAAAGGCCACCAGTGGGAG AGTGGATCTGTGAGGTATGCGGAAGGGACTTCACGACCAAAATTGGCCTGGGACAACACAAAAGACATATGCATGCAATGGTG AGAAACCAGGAAAGGATCGATGCTTCCCAACCGAAAGAGACATCAAATCGAGGAGCCCACAAGAGGTGCTGGACGAAGGAGG AGGAAGAACTGCTCATGAAGTTGGAGGTACAGTTTGAGAATCACAAAAACATCAATAAGCTTATCGCAGAGCAATTAACAACTA AAACAGCTAAACAAATTAGTGATAAAAGGAGAATGCTGCTCAAAAAAGGTAGGGGGACAACTGGTAATTTGGAAACAGAGCCT GGGATGAGTCATCAATCGCAGGCAAAAGTTAAGGACAATGGACTGGGTGGGGACCATCTGCCGGGAGGACCAGTTGTCGATAA GGGAACAATAGGGAAGCCAGGACAACATCTTGACACAGATAACAGCCATCAAATAACTGCTGGCAAGAAGAAAGGGGGAGGG CTGCAGGCTCGTTATAGAAGGAGAATAATGAAACGATTAGCGGCCGGGACAATTAACATCTTCCCCAAAGTGTTTAAAGAACTG ATTAACGACCAAGAGGCGAGACCGCTAATCAATCAAACAACAGAAGACTGCTTTGGCCTCTTGGACTCTGCATGCCAAATTAGAA CGGCACTCCGGGAGAAGGGCAAATCTCAGGAGGAACGACCAAGAAAACAGTATCAGAAGTGGATGAAGAAGAGAGCGATTAA AAGGGGGGACTATCTCCGCTTCCAGCGATTATTCCATCTAGACAGGGGGAAACTGGCGAGAATTATCTTGGACAACACTGAGAG CTTGTCTTGCGATATATCACCCAGTGAAATTTATTCGGTATTCAAGGCCAGATGGGAAACACCTGGACACTTCAACGGCCTTGGG GACTTTGAAATTAAAGGGAAGGCCAACAACAAAGCCTTCAGGGACTTCATCACGGCTAAAGAAATTGAAAAGAACGTGCGGGA AATGAGTAAGGGTTCGGCGCCAGGTCCAGATGGGATCGCCCTTGGGGACATCAAGAAGATGGATCCCGGGTATTCCCGGACCG CCGAGCTATTCAACTTGTGGCTGACAGCTGGTGACATCCCGGACATGGTGAGGGGGTGCAGGACTGTTTTGATCCCGAAATCGA CGACACCGGAGCGCCTAAAGGACATCAACAACTGGAGACCCATCACGATTGGTTCCATCTTGCTAAGGCTGTTCTCCAGGATCAT AACGGCGAGGATGACTAAGGCGTGCCCCCTCAACCCGAGACAGAGAGGCTTCATCAGTGCGCCGGGATGCTCTGAGAACCTGA AACTCCTGCAATCTATAATTCGGACTGCCAAAAATGAGCACAAGCCGCTGGGTGTTATTTTCGTGGACATTGCTAAGGCTTTTGA CACCGTGAGCCACCAACACATCATACACGTTTTACAGCAACGGAGGGTTGACCCCCACATTGTTGGACTGGTGAACAATATGTAC AAGGACATCAGTACGTATGTCACCACAAAGAAGAACACACACACGGACAAAATCCAGATCCGGGTTGGAGTGAAGCAGGGTGA CCCACTATCACCCCTTCTATTCAACTTGGCAATGGACCCCCTGTTGTGTAAGCTGGAAGAAAGTGGCAAAGGATTCCATCGAGGA CAGAGCTCAATAACCGCGATGGCGTTCGCCGACGATCTGGTCTTGTTAAGCGACTCCTGGGAGAACATGAAAGAGAACATCAAA ATACTGGAGACCTTTTGCAATCTCACCGGTCTCAAAACACAGGGTCAGAAGTGCCACGGCTTTTACATCAAGCCTACAAAGGACT CTTACACCATCAACAACTGCCCTGCATGGACCATCAACGGCACACCCCTGAACATGATCAACCCCGGGGAGTCAGAGAAATACCT CGGCCTGCAGATCGACCCATGGACTGGAGTAGCAAAATACGATCTCTCCACAAAATTGAAAATATGGCTCGAAAGCATTGACCG AGCTCCACTTAAACCTCTGCAAAAATTAGACATCCTCAAAACATACACCATTCCTCGACTGACCTACCTGGCTGACCATTCAGAGA TGAAAGCAGGGGCTCTGGAAGCACTCGACCAGCAGATTCGAACAGCGGTCAAAGACTGGCTGCACCTGCCCTCGTGCACCTGTG ATGCCATCTTGTACGTGAGCACGAGGGACGGCGGTTTGGGTGTTACCAAGTTGGCGGGACTGATTCCAAGTGTGCAAGCCCGG AGGCTGCATCGCATTGCGCAGTCGCCGGACGAGACGATGAAGGACTTCCTAGAGAAGGCGCAGATGGAGAAGATGTATGAGA AGTTATGGGTTCAAGCTGGAGGCAAAAAGAAGGGGATGCCGTCAATTTGGGAGGCCCTACCGATGACTGTACCACCCACTAATA CAGGTAATCTTTCGGAGTGGGAAGCACCGAACCCCAAAAGTAAGTACCCAAAACCTTGTGATTGGAGAAGGAAAGAGCTTAAA AAGTGGACAAAATTGGAGTCCCAAGGTCGTGGAGTCAAAAATTTTAGGAATGATACAATTAGTAACGATTGGATCCAATATTAT AGACGCATACCTCACAGGAAACTCCTCACTGCCATACAACTCAGGGCCAATGTATACCCCACAAGGGAATTTCTCGCGCGGGGG AGGGGTGATAACTATGTTAAGTTTTGTAGGCACTGTGAAGCGGACCTTGAAACCTGTGGCCATATCATCGGCTTTTGCCCAGTAA CGAAGGACGCCCGAATCAAGAGGCACAATCGCATATGCGACAGGCTTTGCGAGGAGGCAGCTAAGAGGGAATGGGTGGTCTTC AAGGAGCCGCACTTGAGGGATGCCACCACGGAACTGTTTAAACCGGATGTGATATTCGTGAAAGAGGACCGTGCACTGGTTGT GGATGTGACAGTACGATATGAATCAGCCAAGACAACGCTGGAGGCAGCTGCTATGGAGAAAGTGGACAAGTACAAACATCTGG AGGCAGAAGTGAAGGAACTCACCAACGCAAAGGACGTTGTTTTTATGGGGTTCCCCCTTGGAGCGCGAGGGAAATTCTACAAAG GGAACTTTAACTTGCTAGAGACTCTTGGCCTCCCAAAAACGAGGCAATTGAGTGTGGCAAAGACTCTATCCACGTACGCGCTCAT GTCATCTGTGGACATTGTGCATATGTTTGCCAGTAGATCTAGGAAACCAAATGTCTAG (SEQ ID NO: 46)
	3' UTR	GTAGTCACATTGCACTTTCTGTAACTTGCACTGGGTGTGGGATGTGGGCCTGGGGTGTGGGTTATGGGGTATATATGTGGGATA TTCTGGTGGGAATGTCCATTCACTGTATGCCTATCTTTTTAATAAAAAGACGGTAGCTAGGTTCGCGAAGCAGCCACAAGCCAAT AGCCAGTTAGGTAGCTCATAGTGGGTAGGTGACAGGAACCTTTGACTCAGAACGCGTCCATTAACATCTAGAACGGACCAAACT TCGGACATGCACCGATTAACCGGATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGAAAGGGAACATATATAGTTATAT GTGTTCGTAATA (SEQ ID NO: 47)
	Codon optimized	CTGAAACCTGGACCTCCTGTGTCCGACCGGGGCAAAGAGTTTGGCTCTGGCCTGACAACACACCCCGAGCCTGAGTCTGAGTCT GGACACGATCCCACCGTGCCTAATCCTGGACCATCTCTTGGAGCTGGCGAAGGCGCTCAACCTCTGCCTCTGCTGAGAGTGTCTG TGGGCACCCAGACCTGCGAAGAGGACTTCATCACCAGCAGACCCACCAAGCTGCCTGGCATCGAGTCTGAACTGGGCCCTCTGG TCAAGTTCAGCCTGGAAGTGTACAGAAGCGACCTGAAGGGCGACGTGCAGTTCGAGGGCATCCACTTTCCAGACAACTGGGGC GTGCTGGAAGGCTTCCCTGAGGTGTACGAACAGCTGGCCCCTCAACCTAACGGCGGCGACGAACTGAATCACAGCCTGCCTGGC GATAGAGAAGGGGATGTGCTGGAAAAGGACAGCAGCGAGAAAGAGAAAGAGGCCGCTCCTGAGGCTCTGCCCTCTGTGCAAA

		GAGCCAGATCTGAGCAGCTGCCCGACAACATCGTGAAAGTGACCGTGCCTGACAAGAACCCTCCTTGTCCTTGTTGTGGCGTGC GGCTGAATTCTGTGCTGGCCCTGATTGAGCACCTGAAAGGCTCTCACGGACGGCGGAGAGTGTGCTTCAGATGTGCCAAGTGCG GCAGAGAGAACTTCAACCACCACAGCACCGTGTGCCACTACGCCAAGTGTAAAGGCCCTCAGATCGAGAGGCCTCCTGTCGGAG AGTGGATCTGCGAAGTGTGCGGCAGGGATTTCACCACCAAGATCGGACTGGGACAGCACAAGAGACACATGCACGCCATGGTC CGAAATCAAGAGCGGATCGACGCCAGCCAGCCTAAAGAGACAAGCAACAGAGGCGCCCACAAGCGGTGCTGGACCAAAGAGG AAGAGGAACTGCTGATGAAGCTGGAAGTGCAGTTTGAGAACCACAAGAACATCAACAAGCTGATCGCCGAGCAGCTGACCACA AAGACCGCCAAGCAGATCAGCGACAAGCGGCGGATGCTGCTGAAGAAAGGCAGAGGCACCACCGGCAACCTGGAAACAGAGC CTGGAATGAGCCACCAGAGCCAGGCCAAAGTGAAGGACAATGGCCTCGGCGGCGATCATCTTCCTGGTGGACCTGTGGTGGAC AAGGGCACAATCGGAAAGCCTGGCCAGCACCTGGACACCGACAATTCCCACCAGATCACCGCCGGCAAGAAAAAAGGCGGAGG ACTGCAGGCCAGATACCGGCGGAGAATCATGAAGAGACTGGCCGCTGGCACCATCAACATCTTCCCCAAGGTGTTCAAAGAGCT GATCAACGACCAAGAGGCCAGGCCACTGATCAACCAGACCACCGAGGACTGTTTCGGCCTGCTGGATAGCGCCTGCCAGATCAG AACAGCCCTGAGAGAGAAGGGCAAGAGCCAAGAGGAACGGCCCAGAAAGCAGTACCAGAAATGGATGAAGAAGCGGGCCATC AAGCGGGGCGACTACCTGAGATTCCAGCGGCTGTTCCACCTGGATAGAGGCAAGCTGGCCCGGATCATCCTGGACAATACCGA GAGCCTGAGCTGCGACATCAGCCCCAGCGAGATCTACAGCGTGTTCAAGGCCAGATGGGAGACACCCGGCCACTTTAATGGCCT GGGCGACTTCGAGATCAAAGGCAAGGCCAACAACAAGGCCTTCCGGGACTTTATCACCGCCAAAGAAATCGAGAAGAACGTGC GCGAGATGAGCAAGGGATCTGCCCCTGGACCTGATGGAATCGCCCTGGGAGACATCAAGAAGATGGACCCCGGCTACAGCAGA ACCGCCGAGCTGTTTAACCTGTGGCTGACAGCCGGCGACATCCCCGATATGGTTCGAGGCTGTAGAACCGTGCTGATCCCCAAG AGCACCACACCTGAGCGGCTGAAGGATATCAACAACTGGCGGCCCATCACCATCGGCAGCATCCTGCTGAGACTGTTCAGCAGA ATCATCACAGCCCGGATGACCAAGGCCTGTCCTCTGAACCCTAGACAGCGGGGCTTTATTAGCGCCCCTGGCTGTAGCGAGAAC CTGAAGCTGCTGCAGAGCATCATCCGCACCGCCAAGAACGAGCACAAGCCTCTGGGCGTGATCTTCGTGGATATCGCCAAGGCC TTTGACACCGTGTCTCACCAGCACATTATCCACGTGCTCCAGCAGCGCAGAGTGGACCCTCATATTGTGGGCCTCGTGAACAACA TGTACAAGGACATCAGCACCTACGTGACCACGAAGAAGAACACCCACACCGATAAGATCCAGATTCGCGTGGGCGTGAAGCAG GGCGATCCTTTGTCTCCCCTGCTGTTCAATCTGGCCATGGATCCTCTGCTGTGCAAACTGGAAGAGAGCGGCAAGGGCTTCCACA GAGGCCAGTCTAGCATTACCGCCATGGCCTTCGCTGACGACCTGGTGCTGCTGAGCGATAGCTGGGAAAACATGAAGGAAAACA TCAAGATCCTGGAAACCTTCTGCAACCTGACCGGCCTGAAAACCCAGGGCCAGAAGTGCCACGGCTTCTACATCAAGCCCACCAA GGACAGCTACACCATTAACAACTGCCCCGCCTGGACTATCAACGGCACCCCTCTGAACATGATCAACCCCGGCGAGAGCGAGAA GTACCTGGGCCTGCAAATCGATCCTTGGACCGGCGTGGCCAAATACGACCTGAGCACAAAGCTGAAGATTTGGCTCGAGAGCAT CGACAGAGCCCCACTGAAGCCTCTGCAGAAGCTGGACATCCTCAAGACCTACACAATCCCCAGACTGACCTACCTGGCCGACCAC AGCGAAATGAAGGCTGGCGCTCTGGAAGCTCTGGACCAGCAGATTAGGACCGCCGTGAAGGACTGGCTGCATCTGCCTAGCTG TACCTGCGACGCCATCCTGTACGTGTCCACCAGAGATGGTGGCCTGGGAGTGACAAAACTGGCCGGACTGATCCCTAGCGTGCA GGCTAGAAGGCTGCACAGAATCGCCCAGTCTCCAGACGAGACAATGAAGGACTTCCTGGAAAAAGCCCAGATGGAAAAGATGT ACGAGAAGCTGTGGGTGCAAGCCGGCGGAAAGAAAAAGGGCATGCCTAGCATCTGGGAAGCACTGCCCATGACAGTGCCACCT ACCAACACCGGCAATCTGAGCGAATGGGAAGCCCCTAATCCTAAGAGCAAGTACCCCAAGCCTTGCGACTGGCGGAGAAAAGA ACTGAAGAAGTGGACCAAGCTGGAAAGCCAAGGCCGGGGAGTGAAGAACTTCCGGAACGACACCATCAGCAACGATTGGATCC AGTACTACAGACGGATCCCTCACAGAAAGCTGCTGACCGCCATTCAGCTGCGGGCCAACGTGTACCCTACCAGAGAATTCCTGGC CAGAGGCAGGGGCGACAACTACGTGAAGTTCTGCAGACACTGCGAGGCCGATCTGGAAACATGCGGCCACATCATCGGCTTTT GCCCCGTGACCAAGGACGCCAGAATCAAGCGGCACAACAGAATCTGCGACCGGCTGTGTGAAGAGGCTGCCAAGCGAGAGTG GGTCGTGTTTAAAGAGCCTCACCTGAGGGACGCCACCACAGAGCTGTTCAAGCCCGACGTGATCTTTGTGAAAGAGGATAGAGC CCTGGTCGTGGACGTGACCGTCAGATACGAGAGCGCCAAGACCACACTGGAAGCCGCCGCTATGGAAAAAGTGGACAAGTACA AACACCTCGAGGCCGAAGTGAAAGAACTCACCAACGCCAAGGACGTGGTGTTCATGGGCTTTCCACTGGGCGCCAGAGGGAAG TTCTACAAGGGCAACTTTAACCTGCTGGAAACCCTGGGACTGCCCAAGACCAGACAGCTGAGTGTGGCCAAGACACTGTCCACA TACGCCCTGATGTCTAGCGTGGACATCGTGCACATGTTCGCCAGCAGATCCAGAAAGCCCAACGTGTGA (SEQ ID NO: 48)

R2-2_PM	**5' UTR**	CGGTGCGTTCCCTTGGGTAAGGAACACGAGTCTTAGTGGCCTTGACCTCCACGTGGTCCCGCTGGTAACATCATCTCTTGATGAT GGCTAACAAGGCTAATGCACCCATTCCATCTCCTATCTCGCATGGAGGCCGCTATGCGTGATTACTAGAGGCCACAAACAAAACT T (SEQ ID NO: 49)
	CDS	ATGACTGACAAACTTAAGTTCTCAAGCCAGTTGGCACGAGGCCTGGCAAAACAACGTGCTATGGATGGCGCTCGGGTTGGCGAT CCACCCATTACAGTTAGACCCACAGAAACCGATCTGTGCAACACTGAGGGTTCATGGGGACGCCGTCCTATGAAACTCTTGTTTG TCTCGGTGTCAACCCAGACACAGAATGAAGATGCCCTCTGGGCATCTGATGTTGCTAAACCTATGGCGTCTAGGTCGGCGCTAAA AATGACGAGTATACCTTCCATGACCTTCCATAACTCGTCCTTGGAAAAGGAAGAGGAGATGAACTACGATTTTTACGAACAGATT AAAAGTCTAGTTGAGTCGGATGACTCTTCAGATGACTTTACAGAGGATGATGAGGATGTGGAGGAGTCCTTCCTCGACATATCG GCTGAGGAACCCGTGTTGGGAAAGTTTCCCATTGACACCAAGGGAACTATCACGGTTGTACTGCCTAGTCTGGAGTATATATGC GTTATCTGTAAACAACACATGGGCAAAGCATCCGAACTTGTTGCACATTTTAACATCAAACACAGAGACATTCCTCTGGTGTTTAA GTGCGCTAAATGTGACAAAACCAACTCAAACCACCGATCGATTGCCTGTCATGCCCCCAAATGTGGGGGAATAAAGTTAACTGA GGAAAGCTTACCAATGGTCTGTGAATGTTGCCAGGCACGCTTTGCGACTCTAAGTGGCCTTTCGCAGCACAAGAGGCATGCTCAC CCAGTCACCCGTAATGAGGAAAGGATTAAAGATGGTATAAAGGGTACCTCGCAGAGAGGGGTACACCGTAGCTGCTGGTCTTT GAAGGAAGTAGAACAGCTGGCCCTTCTAGAGTTGCAGTTTCAGGGGAAAAAGAATATCAATAAGATCATTGCTGAAGCGCTTGG GACTAAGACCAACAAGCAAGTCTCTGACAAGAGACGGGACCTAAGTAAAAAGACAGGGGCCCCCATGTCAGACAGCTTACATTT TTCTTCTAGGCCTCTTGAGACATTGTCTCCCCCACCAAATGTAACAACGGGGACTTCATCCATACTCGCTCAAGCAGCTGAGCGGC TTACGAATGAGAATTCTGGGACCCTGGAAAAGCCTGCAATGGAGGCAATAAAGGCTTGGCTTAACGGCGAGGGCCAACATGAT GCCCTCGTAGAAACTGCCACAGCATTGATGCTTTGTCCGATGAGATTGGTGAAAAACAAAGGCAAACGTTCAAAACCCGAGAAC GACATTATTAAACCCAGGATATTACCCACACGATCTTGGATGAAGAAGAGAGCGGAAAAACGAGGAAGCTTCATGAAGCACCA GAAGCTCTTCTTTAAGAACCGCTCTCTTCTTGCGTCCTTAGTCCTGGATGGCACTGAACGTCATGAATGCCGAATCCCGAACGCAG ATGTATATCGTTTTTACTGCGAAAAATGGGAGAAGGTGTTGCCATTCAATGGCCTGGGCCAATTTAAGTCATCAGGTGTTGCAAA TAACGAATACTTTGAGCCCCTAATTTCGGTGGAGGAAGTTCAGACTGCCATACGGGCCATTAAACCAACGTCAGCAGCTGGGCC AGATGGCCTAACAAGGGCTGCAATCTGTGCTGCCGACCCCGAGGGTCGGACACTGACAGCCCTATTTAATGCATGGATGATTAC AGGAATTATTCCCAAAGAGTTGAAAAAGAATAGGACGATTCTTATTCCTAAGGTTATGGACGATGAAAAGCTGAAAGAATTGGG GAACTGGAGACCAATAACGATTGGTTCAATGATTCTGAGATTATTTTCCAGAATAATGACTGCACGTCTTGCTCGTGCTTGTCCCT TAAACCCAAGGCAGCGTGGTTTTATAGCGGCATCTGGCTGCTCTGAAAATCTTAAGGTGCTACAGGACCTTATGAGACACGCTAA GAAATTGCACAGGCCGTTGGCTGTCATGTTCATCGACATAGCGAAAGCTTTTGACTCGGTTTCGCATGCTCATATTTTATGGGTGT

		TAAGGCACAAGAAAGTAGATGAACACGTGGTGGGCATCATCCAGAACGCCTACGATCGGTGTACGACCTCGTTCAAAAGCAATG GCGAGTCGACTCGAGAAATTAGCATACGTGTTGGTGTCAAACAGGGTGACCCCATGTCACCCCTGCTCTTCAATCTTGCCATGGA CCCTTTGATATGCACCCTAGAGTCACACGGAGTTGGGTACTCCATTGATACCGACCACGTGACAGCTCTTGCGTTTGCTGATGATT TGGTGTTGGTGAGCGAATCTTGGGTTGGTATGGCCGCCAATCTAGCGATCTTGGAATCATTTTGTGGGCTATCGGGATTGGAGG TTCAGGCCAGAAAGTGCCAGGGCTTCATGATAAGCCCAACCAAAGATTCATATACGGTGAACAACTGCGACCCATGGACTATCA AAAATAAAGATGTCCATATGATCCAACCTGATGAATCAACGAAATACCTTGGTCTAAAAATTTGCCCTTGGACTGGCATTATACG GTCGGATCTACATGTTCAACTAAAGACACGGATCTCGAAAATCGATGAGGCGCCTCTGAAACCGACTCAGAAGGTCGAACTCCTC AATGCCTACGCCCTACCCAGATTATTGTACCCTGCTGACCACTCGGACTGCAAGCAATCAACTCTCCGTGTGTTGGACCAAGAAAT AATAAAGGCGGTAAAAGGATGGCTCCATCTTCCCGCGTCAACCTGTGACGGGCTGTTGTACGCCAGAGCCCGAGACGGAGGCCT TGCCATCTTGAAACTGGAAAATGCAATTCCTTCGGTTCAAGTTAGAAGGCTGCAACGTATTGCAAACTCCTCTGACGCTATCGCTC GAAACATTGCGTCCTCGCAGGGTGTGGAGGAAGAGTACCGAAGTCTGTGGGTACGGGCAGGGGGTGACAGCGAAGCAATCCC AACGTTCTTTCTCAGGGGTTCGGAATCAAAAGAGCCCGTGTATCCGAGACCCTGCGATTGGAGGAAACGCGAATCTCGGAGACG GTGCGAAAAGCCGGTTCAAGGAAGGGGCATTGTAAACTTTGCGCAAGATAGAATCAGTAATGCATGGTTGGGGCCACGGTGCG GCTTTAAACAATGCTTCTTTATCGCAGCATTACAATTAAGGGCAAATATTTACCCAACAAGAGAAAGCATAAACAGAGGCAGAGA TGGTGCCTCACGGTCCTGCAGGAAATGCTCTGCCAGGCTGGAGTCTCTCTCGCACATTCTTGGTCAATGTCCCGCAGTACAAAAA TTCAAGGATTGCGCGACGCAATAA (SEQ ID NO: 50)
	3' UTR	GATCAGCGACATTCTAGCTGACGAAGCGGCGAGACTGGGCTGGTGGGTGTACAAAGAGCCACGGTTCACATCTGAAGCCGGAG AGCTAAGGAAACCTGCCCTTGTGTTTGCCAAAGGTGAGGAAGCGCTTGTTATTGATGTCACCGTCCGGTTTGAGCTCTCGAGGAA AACCTCATCAGAGGCTGCCTCGCACCAAGTTGCGTACTACACCCCCCCTTGTGATCAAGTCAAAGTGCTGACGAAGGCAAGCAAT GTCACATTCTTTGGATTCCAGGTTGGGGCAAGAGGGAAAGTGGCCCCTTGAGAATAATGAGGTGCTAACCTCCCTGGGCCTGAC CAAACCCAGAACACATCACTGGCCAAGATGATTTCCCGCAGCACGTTGCTTTTCTCTCTCGATACCCGAGAATGTTCTGCGGAACT ACGCAGTCTATGAACAGTCACAGACAACCTCTGATCCAAG (SEQ ID NO: 51)
	Codon optimized	ATGACCGACAAGCTGAAGTTCAGTTCCCAGCTGGCCAGAGGCCTGGCCAAACAGAGAGCTATGGACGGCGCTAGAGTGGGCGA TCCTCCTATTACAGTGCGGCCCACCGAGACAGACCTGTGCAATACCGAAGGCTCTTGGGGCCGCAGACCTATGAAGCTGCTGTTC GTGTCCGTGTCCACACAGACCCAGAACGAGGATGCCCTGTGGGCCTCTGATGTGGCCAAACCTATGGCCTCTAGAAGCGCCCTG AAGATGACCAGCATTCCCAGCATGACCTTCCACAACAGCAGCCTCGAGAAAGAGGAAGAGATGAACTACGACTTCTACGAGCAG ATCAAGAGCCTGGTGGAAAGCGACGACAGCAGCGACGACTTCACCGAGGACGACGAGGATGTGGAAGAGAGCTTCCTGGACA TCAGCGCCGAGGAACCTGTGCTGGGCAAGTTCCCCATCGATACCAAGGGCACCATCACCGTGGTGCTGCCTAGCCTCGAGTACA TCTGCGTGATCTGCAAGCAGCACATGGGCAAAGCCTCTGAGCTGGTGGCCCACTTCAACATCAAGCACCGGGACATCCCTCTGGT GTTCAAGTGCGCCAAGTGCGACAAGACCAACAGCAACCACCGGTCTATCGCCTGTCACGCCCCTAAGTGTGGCGGCATCAAGCT GACCGAGGAATCCCTGCCTATGGTCTGCGAGTGCTGCCAGGCCAGATTTGCCACACTGTCTGGCCTGTCTCAGCACAAGAGACA CGCTCACCCCGTGACCAGAAACGAGGAAAGAATCAAGGACGGCATCAAAGGCACCAGCCAGAGAGGCGTGCACAGAAGCTGTT GGAGCCTGAAAGAGGTGGAACAGCTGGCCCTGCTCGAGCTGCAGTTTCAGGGCAAGAAGAACATCAACAAGATCATTGCCGAG GCTCTGGGCACCAAGACAAACAAACAGGTGTCCGACAAGCGGCGGGACCTGAGCAAAAAAACAGGCGCCCCTATGAGCGACAG CCTGCACTTTAGCAGCAGACCCCTGGAAACACTGAGCCCTCCACCTAATGTGACCACCGGCACCAGCTCTATTCTGGCCCAGGCT GCCGAGAGACTGACCAACGAGAATAGCGGCACCCTGGAAAAGCCCGCCATGGAAGCCATTAAGGCCTGGCTGAATGGCGAGG GACAGCACGATGCCCTGGTCGAAACAGCCACAGCTCTGATGCTGTGCCCCATGCGGCTGGTCAAGAACAAGGGCAAGAGAAGC AAGCCCGAGAACGACATCATCAAGCCCCGGATCCTGCCTACCAGATCCTGGATGAAGAAGCGGGCCGAGAAGCGGGGCAGCTT TATGAAGCACCAGAAACTGTTCTTCAAGAATCGGAGCCTGCTGGCTAGCCTGGTGCTGGATGGAACCGAGAGACACGAGTGCA GAATCCCCAACGCCGACGTGTACCGGTTCTACTGCGAGAAGTGGGAGAAAGTGCTGCCCTTCAACGGCCTGGGACAGTTCAAGT CTAGCGGCGTGGCCAACAACGAGTACTTCGAGCCTCTGATCAGCGTGGAAGAGGTGCAGACCGCCATCAGAGCCATCAAGCCTA CATCTGCCGCCGGACCTGATGGCCTGACAAGAGCCGCTATTTGTGCCGCCGATCCTGAGGGCAGAACACTGACAGCCCTGTTCA ACGCCTGGATGATCACCGGCATCATCCCCAAAGAGCTGAAGAAGAACCGGACCATTCTGATCCCTAAAGTGATGGACGATGAGA AGCTGAAAGAACTCGGCAACTGGCGGCCCATCACCATCGGCTCTATGATCCTGCGGCTGTTCAGCCGGATCATGACCGCCAGAC TGGCTAGAGCCTGTCCTCTGAACCCTCGGCAGAGAGGCTTTATCGCCGCCTCTGGCTGTAGCGAGAACCTGAAGGTTCTGCAGG ATCTGATGCGGCACGCCAAGAAGCTGCATAGACCTCTGGCCGTGATGTTCATCGATATCGCCAAGGCCTTCGACTCCGTGTCTCA CGCCCATATCCTGTGGGTGCTGCGGCACAAGAAAGTGGACGAGCACGTCGTGGGAATCATCCAGAACGCCTACGACAGATGCA CCACCAGCTTCAAGAGCAACGGCGAGAGCACCCGCGAGATCTCTATCAGAGTGGGAGTGAAGCAGGGCGACCCAATGAGCCCA CTGCTGTTCAACCTGGCCATGGATCCTCTGATCTGCACACTGGAATCTCACGGCGTGGGCTACAGCATCGACACCGATCATGTGA CTGCCCTGGCCTTCGCCGATGATCTGGTTCTGGTGTCTGAGAGCTGGGTCGGAATGGCCGCCAATCTGGCCATCCTGGAAAGCTT TTGTGGCCTGAGCGGCCTTGAGGTGCAGGCCAGAAAATGCCAGGGCTTCATGATCAGCCCCACCAAGGACAGCTACACCGTGAA CAACTGCGACCCTTGGACCATCAAAAACAAGGACGTGCACATGATCCAGCCTGACGAGTCCACCAAGTACCTGGGCCTGAAGAT CTGCCCATGGACAGGCATCATTCGGAGCGACCTGCATGTGCAGCTGAAAACCCGGATCAGCAAGATCGACGAGGCCCCTCTGAA GCCCACACAGAAGGTTGAGCTGCTGAACGCTTACGCCCTGCCTAGACTGCTGTACCCTGCCGATCACAGCGATTGCAAGCAGAG CACCCTGAGAGTGCTGGACCAAGAGATCATCAAGGCCGTGAAAGGCTGGCTGCATCTGCCTGCCTCTACATGTGACGGCCTGCT GTATGCCAGAGCTAGAGATGGCGGACTGGCCATTCTGAAGCTGGAAAACGCTATCCCCAGCGTGCAAGTGCGGCGGCTGCAGA GAATCGCCAATAGCTCCGATGCCATTGCCAGAAATATCGCCAGCAGCCAGGGCGTCGAAGAGGAATATCGCAGCCTGTGGGTTC GAGCCGGCGGAGATAGTGAAGCTATCCCTACATTCTTCCTGAGAGGCAGCGAGAGCAAAGAACCTGTGTACCCCAGACCTTGCG ACTGGCGGAAGCGGGAATCTCGGAGAAGATGCGAGAAGCCCGTGCAAGGCAGAGGCATCGTGAACTTCGCCCAGGACCGGAT CTCTAACGCCTGGCTGGGACCTAGATGCGGCTTCAAGCAGTGCTTCTTTATCGCTGCCCTGCAGCTGAGAGCCAACATCTACCCT ACCAGAGAGAGCATCAACCGGGGCAGAGATGGGGCCAGCAGATCCTGTAGGAAGTGCTCTGCCAGGCTGGAAAGCCTGAGCC ACATTCTGGGACAATGCCCTGCCGTGCAGAAGTTCAAGGACTGCGCCACACAGTGA (SEQ ID NO: 52)

R2-2_SMed	**5' UTR**	AGTCATAGGGTGAACTGCAATTCTGACACGATGACCGAGCTGTGTCAGTTTGCAGCTAGTCGCTAAAGACTCGATCAGTCCGCCA AGTGAGGTGGCCGGGTATCTGCAGCACTAGAGCCACTGGTATCAAGAGCAGAGATACGCGAGTGGAAGTTGAGTACGACTACC TTCACGGGGTCCTCCTGATAACCACAGTGGACTGTGGGAACTAAATGTGTGCTCAGCGTTCCCTACTTTCTCGTAGGGTAAAGGG TATGATAACCCAGAGAATATCCCATGGGAGATATCCATGGAAAAAGCACCACGTTAGACAATCCGATGGTCTAACTCGGCTCCG AGGGGCTAACTATCCCAAAGGGCTTAAAGAAAAGAA (SEQ ID NO: 53)
	CDS	CTGGTTACAATCAAAAATCTATTTGAAGAGTCAGGTGCCACCGCGCCTGCACCAGTTCCCCTAGAAGTTGCAGTTGAGGTACACC AGTCATCGAGTGTTCCCGAGATAACCGACGAATCTACGACTACTCAGGAAGGAAGCTATTCAGAACCACCGATACACCGATGTG

		AGAACTGTGGAAGAGAATTCAGAACAAGAGCAGGAGTTCAACAGCACCGGAGAAAAGCTCACACCAACGAGTTTATGGAGGA GAAAGAGAAGGCAGCTCCAACCAAGAAGCTTCGATGGACAGACGAAGAAAAAGAGATCCTCATCGAAAGCGAGATAAAGATCA TCAAGGAAGGATCCCTTAAAGAACAACACGAAATAAACAAGATTCTAGCTTCAAGAATGCCTGGAAGATCACAAGACGGAATAG CCAAAATTCGCCAGAAACAAGAGCACAAGGCTGAAATACAGAGAAGGCTACATGGGACCGTCACTACCAACGAAACCAGAGGA AACAGAACCAGCGAAATTACCGAACCAATAAGAAGCTTACCAATAAACACCAAAACCTGGAGCGAAGACGAAATGAAAAGAAT GCTAGCCGAAGAAGTGAAGTTGAGAACCAAGAATGAAAAGGATATCAACAAGAAGCTAGCAGAAATATTCCCGAACAGAACAA TGGGGTCCATAAAGAGCAAAAGGACGAAAGATAAGGACTACCAGGATTTGGTAAAGCTAACAATGCAAACAATCAGCGAGAAC CCAGACAATGAAACAGACTTCAATACCAGCAACACGGAAAACAACAGCACTGATGCAGAGAAAGAAGTGAAGAACTACCTCAAT ATGCTACTACTGACCATCAACGAGGAAGAATGGTTGACATCCACTCTGAAGGAAGCAGCAACACTAGCACTACAAGGAAAGAAA ACTGAAGCATCTGAAAAGCTTAACGAGTATGCAAGCAAAACGCTGTTCCCTGGATTGAAAATAACTAATCAGACGAGGAAACGA GAAAAGAAAATATCAAAAAGAGAAACCAGAAGGCAAGAATACGCTGAAATACAGAAACTCTACAAAAAGAATATTTCAAGCGC AGCCGAAAAAGCAATCAATGGGAAATGGTCCATAAAACCCGAAGAAGAATACCACAACAATAAGGATCTCATTAAAGCATGGA AACCAATACTAGAGGCACCTCCATTCAGTGACTGTAGGCCCATCGAAAACATCAAGGAAATGGACTACGCTTTAATGGAGATCA GCACGGCAGAAATCTTCCTCGCAATAAGAGCCATGGGGAAGACTGCACCAGGACCTGATGGCATTAAATATTCAAAGCTCAAGA AAAATATCCAATCAATGGCAATATTATTCAACACATGCCTACTAACGAGCTTCCTGCCACTCCCATTGAAGATAGCAAGGACAATC CTGATACCTAAACAAGAGAATCCAGGTATCCTTGACTATCGACCACTAACAATAGCCTCAGTGGTGACAAGAGTGTTCCACAGTA TCCTTGCGAAGAAGCTCGACAACAATGCCCAATTAAGTCAACGACAAAAAGGATTTCGAAAATGTGATGGAGTTGCGGAAAATA TAGTAATACTCGAAACTATATTAACCAACAGCAGAAGTGAAAAGAGACCGCTCTGTATGGCCTTCGTTGACTTAAGAAAAGCATT CGATTCTGTGGGACATGAGTCTATCATCAGAGGAGCAAAAAGAGTTGGAGTGCCACCAATGTTGCTCGAGTACATTTCGTCAAG TTACCAGAATGCGTCTACTAACTTGTTCGGCGAAATACTCAACTCGAGAAGAGGAGTCAGGCAAGGCGACCCTCTGAGCCCTATT CTCTTCAATTTTGTTATCGATGAAGCTCTAGAAAACCTCAACAGGAATATTGGATATCTACTGAAGGAAGAAAAAGTGAGTTGCC TAGCTTTCGCGGACGACATAGTCCTGATAGCTGAGACAAAAGGAGGCCTAGAGAATCATATCGAGAAACTATTAGAGAAGCTGA ATGGGGCCGGTCTCGAGTTGAACGCCTCGAAATGCGCAACACTGATGGTGATGAAAAACGGAAAGGAAAAATCAACGTATATA TCAACAAAAGCAATCAAAATCAAAGAAAATGACATTCCAACAATGAAAGCCACAGAAACGTACAAATATCTCGGATTGCAAATG GGTTTCAAAGCTAGAGAACAGAATGCTAATGAGGTTATTACAGAAGGACTGGAGAATATAACAAGAGCACCACTGAAGCCCCA GCAGAGGATACATATCCTACGAGACTTCCTTATACCAAGATTAATACACAAATTGGTATTAGGAAGAGTGGCCAAGAAGTCATTA AAAAGAATCGACCAGAATATAAGAAAGAAAGTGAGGAATTGGCTACATCTCCCTAAAGACACGACAGCAGCATTCATACACGCT GATGCAGGAGATGGAGGGCTTGGAGTACCAGCGTTAGAACACACAATTCCTCTACTGAAAAGAGAAAGAATAACTAATCTAAG AAAATCCAATGATCCAGTTACCAAAGAATGCCTGAGAATGGAGTACACCAAACAAGTACTGGGAAAATGGAGTAGACCAACTAA AATTGGAGAAACTCTGGCTACCAACAAAAGCCAACTAAAAGAAGCATTCAGAAAACAGATGTTAATAACGCTAGATGGAAAAGG GCTAAAAGATCACCACGAAACGCCCACTATCCACAAATGGATCAGAAGAGGAGAGAACATGACCGGCAAACAGTTTATCACAGC AGTTAAAATAAGAGGAAACCTTGTGGCAACTAAGTCAAGAAATAGCAGAGGGAGACCCGAACAAGAGAAACTATGTGAAGCTC AATGTGGACGACCAGATAGCCTGGGACACATATTACAAGGTTGCTGGAGAACACATGGTATGCGAGTGGAAAGGCATAACAAC ATTTGTCGCAGAATTAAAGCAATAATGAAAGGAAAGGAAAGCGAAGTAGTCGAAGAACCAAGACTACAAACGAATGAAGGTCT CAGAAAGCCTGACTTACTGATCTGTCACAAAGGTAAAATAATAATATGCGATGCACAAGTAGTGGCAGATAGCTCGAACTGCAG TCTTGAAAGCGAAAACCAAAGAAAGATAGATTACTACAAGAAAGATTCAGTAGTATCAGAAGCAAGAAAACTTATCGGACGTGT CGACGAAGATATAATTATAATGGCAGTGACCTTTAACTGGAGAGGAGCCATCTCAAAAACATCAATAAGAGATTTGGACATGCTT CTAGATATAAAATCAAAAGAAGTAATTAAAATGTCAAGGAAAATAATCAGAGATAATAGCATCATGGTGGAGATGCACAGAAAC CGGACTGAGAAAAGGAGATAG (SEQ ID NO: 54)
	3' UTR	AAGGGAAACAAAGGAAAAACGAAATGACTGGAAACTATGAAGGATATAGCTGAAAGCCGCAAGGAAGGCTAAGTCCTGAAAC CGATCTACATCTTCGATCCCAAGAGGAACTGTGGGTTAAGCTTGAGCCGACGGAAAAAGCGAATGCATGTTAGACGACGAGGTA CAGTCACCTCCTCGTGGTATTTGGCGGGCAATGCTCACTAAATTAACTGTGAGTAGCTGAGAACTGTATGTGTATCATGAAAAAA AAA (SEQ ID NO: 55)
	Codon optimized	CTGGTCACCATCAAGAACCTGTTCGAGGAAAGCGGCGCCACCGCTCCTGCTCCAGTTCCTCTTGAAGTGGCCGTGGAAGTGCACC AGAGCAGCTCTGTGCCTGAGATCACCGACGAGAGCACCACCACACAAGAGGGCAGCTACAGCGAGCCTCCTATCCACAGATGCG AGAACTGCGGCAGAGAGTTCAGAACCAGAGCTGGCGTGCAGCAGCACAGAAGAAAGGCCCACACCAACGAGTTCATGGAAGA GAAAGAGAAGGCCGCTCCTACCAAGAAACTGCGGTGGACCGACGAGGAAAAAGAGATCCTGATCGAGAGCGAGATCAAGATC ATCAAAGAGGGCTCCCTGAAAGAGCAGCACGAGATCAACAAGATCCTGGCCAGCAGAATGCCCGGCAGAAGCCAGGACGGAAT CGCCAAGATCAGACAGAAGCAAGAGCACAAGGCCGAGATCCAGAGAAGGCTGCACGGCACCGTGACCACCAATGAGACAAGA GGCAACCGGACCTCCGAGATCACTGAGCCCATCAGAAGCCTGCCTATCAACACCAAGACTTGGAGCGAGGACGAGATGAAGCG GATGCTGGCCGAAGAAGTGAAGCTGCGGACCAAGAACGAGAAGGACATTAACAAGAAGCTCGCCGAGATCTTCCCCAACAGAA CCATGGGCAGCATCAAGAGCAAGAGGACCAAGGACAAGGACTACCAGGACCTGGTCAAGCTGACCATGCAGACCATCAGCGAG AACCCCGACAACGAGACAGACTTCAACACCAGCAACACCGAGAACAACAGCACCGACGCCGAGAAAGAAGTCAAGAACTACCT GAACATGCTGCTGCTGACCATCAACGAGGAAGAGTGGCTGACCAGCACACTGAAAGAGGCCGCTACACTGGCTCTGCAGGGCA AGAAAACAGAGGCCAGCGAGAAGCTGAACGAGTACGCCAGCAAGACACTGTTCCCCGGCCTGAAGATCACAAACCAGACCAGA AAGCGCGAGAAGAAGATCAGCAAGAGAGAGACACGGCGGCAAGAGTACGCCGAGATTCAGAAGCTGTACAAGAAGAACATCT CCTCTGCCGCCGAGAAGGCCATCAACGGCAAGTGGTCCATCAAGCCCGAGGAAGAATACCACAACAACAAGGATCTGATCAAG GCCTGGAAGCCCATCCTCGAGGCCCCTCCATTCAGCGACTGTAGACCCATCGAGAATATCAAAGAGATGGACTACGCCCTGATG GAAATCTCTACCGCCGAAATCTTTCTGGCCATCCGCGCCATGGGAAAGACAGCCCCTGGACCTGATGGCATCAAGTACAGCAAG CTGAAGAAAAACATCCAGAGCATGGCCATCCTGTTCAATACCTGCCTGCTGACCTCCTTCCTGCCTCTGCCACTGAAGATCGCCCG GACCATTCTGATCCCCAAGCAAGAGAACCCTGGCATCCTGGACTACAGACCCCTGACAATCGCCAGCGTGGTCACCAGAGTGTTC CACAGCATTCTGGCCAAGAAGCTGGACAACAACGCCCAGCTGAGCCAGCGGCAGAAAGGCTTCAGAAAGTGTGATGGCGTGGC CGAGAACATCGTGATCCTGGAAACCATCCTGACCAACAGCAGAAGCGAGAAGAGGCCCCTGTGCATGGCCTTCGTGGATCTGAG AAAGGCCTTCGACTCTGTGGGCCACGAGAGCATCATTAGAGGCGCCAAGAGAGTGGGCGTCCCACCTATGCTGCTCGAGTACAT CAGCTCCAGCTACCAGAACGCCAGCACCAATCTGTTCGGCGAGATTCTCAACTCTCGGAGAGGCGTCAGACAGGGCGATCCTCT GAGCCCCATTCTGTTCAACTTCGTGATCGACGAGGCCCTGGAAAACCTGAACCGGAACATCGGCTACCTGCTGAAAGAAGAGAA GGTTTCCTGCCTGGCCTTCGCCGACGACATCGTGCTGATCGCCGAAACAAAAGGCGGCCTGGAAAATCACATTGAGAAGCTGCT GGAAAAGCTCAATGGCGCCGGACTGGAACTGAACGCCTCCAAGTGTGCCACACTGATGGTCATGAAGAACGGCAAAGAGAAGT CCACCTACATCAGCACCAAGGCCATTAAGATCAAAGAAAACGACATCCCCACCATGAAGGCCACCGAGACATACAAGTACCTGG GCCTGCAGATGGGCTTTAAGGCCAGAGAGCAGAACGCTAACGAAGTGATCACCGAGGGCCTCGAAAACATCACACGGGCCCCT CTGAAGCCACAGCAGAGAATCCACATCCTGCGGGACTTTCTGATTCCCCGGCTGATCCACAAGCTGGTGCTGGGCAGAGTGGCC

		AAAAAGAGCCTGAAGAGAATCGACCAGAACATCCGGAAGAAAGTGCGGAACTGGCTGCATCTGCCCAAGGATACCACCGCCGC CTTTATTCATGCCGATGCTGGCGACGGTGGACTGGGAGTTCCTGCTCTGGAACACACAATCCCTCTGCTGAAGAGAGAGCGGAT CACCAACCTGCGCAAGAGCAACGACCCCGTGACCAAAGAATGCCTGCGGATGGAGTACACCAAACAGGTGCTCGGAAAGTGGT CCCGGCCTACAAAGATCGGAGAGACACTGGCCACCAACAAGTCTCAGCTCAAAGAGGCCTTTCGGAAGCAGATGCTGATCACCC TGGATGGCAAGGGCCTGAAGGACCACCACGAGACACCTACCATCCACAAGTGGATTCGGAGGGGCGAGAACATGACCGGCAA GCAGTTTATCACCGCCGTGAAGATCCGGGGCAACCTGGTGGCCACAAAGTCCAGAAACTCCAGAGGCAGACCCGAGCAAGAAA AGCTGTGCGAGGCTCAGTGCGGCAGACCTGATTCTCTGGGCCACATTCTGCAAGGCTGTTGGAGAACCCACGGCATGAGAGTG GAACGGCACAACAATATCTGCCGGCGCATCAAAGCCATCATGAAGGGCAAAGAAAGCGAGGTGGTGGAAGAACCCCGGCTGCA GACAAATGAGGGCCTGAGAAAGCCCGACCTGCTGATCTGTCACAAGGGCAAGATCATTATCTGCGACGCCCAGGTGGTGGCCG ACAGCTCTAACTGTAGCCTGGAATCCGAGAACCAGCGGAAGATCGACTACTACAAAAAGGACAGCGTGGTGTCTGAGGCCCGG AAGCTGATCGGTAGAGTGGACGAGGACATCATCATCATGGCCGTGACCTTCAATTGGAGGGGCGCCATCTCCAAGACCAGCATC AGAGATCTGGATATGCTGCTGGACATCAAGTCCAAAGAAGTGATTAAGATGAGCCGGAAGATCATCCGGGACAACAGCATCAT GGTGGAAATGCACCGGAACCGGACCGAGAAGCGGAGATAA (SEQ ID NO: 56)
R2NS-1_CSi	**5' UTR**	TAGTCGGCGAGCTGAACCACCTCCTCGTGGTGCCGA (SEQ ID NO: 57)
	CDS	CTGGGTAGCCTGGACGCCAGTCTGGGTGAGCTAAGAGTTCAGCAACTCCAGACAGGGCTAACCACCCTGTTTGGTTTCAATGTG CTGGTTACTTTCGACAATGTACACTACAAAACATCTGGCGCCTCCGCACCAGTTCCAACCAGTACACAGGAGAGACTCGTGGGGC TCACGTGTGAGGAATGTGGCAAGTGGTGTAAATCGAAAGCTGGCTTGGTAGCCCACCATCGAGTTCACGACAATGATAGTGTTG GTACGAACATGGTCGCCCAATTGGCCTGCGCTGATTGTTCCCGCCTTTTTCCGACGAAGATTGGCCTGAGTCAACATCGCCGGCA CGCACACCCCACCCAGCATAATGCAGATAAGCTCAGCCGAGTGAAGCATTCCGGTGCCCGCTGGTCCCAACAAGAGTCACAATC GCTCTTGCGCCTGGCTAACAATTTATATCCGTCTTGTGAAACGCAAACCGCGCTGTTCGCGAGACTGGAGCAGTATTTTCCTGGTC GATCGGCTATCAGCATCAAAACCAGGTTACGGGTGCTTAACTGGCAAGCACAACAGGACGAATCATCATCTGGTGGACCTGACC AAACCATCGGTCAAATAGCGGCCTACTCGTCAGAAGCCGATGACTATAGCGTCTGGTTTAAACAAACCGTGGATTGCGCCGTGTC ACTCTTAGAATCCCATGCTGACAGTTCACTTGCTAGTGTTGACCTTCTGGCATTTGCCCGAGGGTTGCAGTCTGGCATCATGACAC CGGAACAAGTCCTCTCGCTTCTGGACCTCCACGCCTCCAGGACATTTCCACACACCTGGAAAACCGTATCCCGACGCCGTCGTCA GTTAGCCCATCGGATGCCGGTCAACCGAAAGCAGATCCGCCGAGCCAACTACGCCGCCATCCAGACCCTATACCACCAAAGACG GAAAGATGCAGCGTCTGCTGTGCTCGATGGGTCATGGAAAGATCTGTACAAAGGCAACTGCGGCCTACCTCCAGACGCCGAGCA GTACTGGAAACAGGTGCTCTCAGCCCCAAAACACGTGGACAGCCGGCCAAGTCGCGTAGTAGTACCATCCGATTGGAGCTTGAT CGAGCCCATCACAGGAGAGGAAGTCGGCCGTACGGTCCGTTCAATGGGCAATTCGTCCCCAGGCCTGGACAAGCTCACACCCAG GATGCTCCGCCGGTTCAATGCGAACGTACTTGCTGGGTATTTCAACTTACTCTTACTATCTGGGGGTTGTCCCCCACACCTGTGTC GTGCCCGTATTACTCTAGTTCCGAAGGTCCCTAATCCGACTTCACCGGATCAACTGAGACCGATCTCCGTATCATCCATTCTTGTCC GATGCTTTCACAAGGTGCTTGCTGATCGCTGGAGTCGTAGATTGCAACTGCCTTCACTCCAGTTTGCATTTCTACATCGGGACGGC TGTTTGGAAGCTACATCGCTGTTACATGCCCTGCTTCGTCACTCTTCAGCGACTGCTTCTAACCTCAGCTTGGCGTTCGTCGATATC TCGAAAGCATTTGATTCGGTTTCGCATGACACAATCGTCAGATCTGCCGAGGCGTTTGGAGCACCGTCACCTCTAGTTCGATACA TCGCACAGTCCTATGAGAATGCTGTAGCAGTTTTCCCCAGTTCCGAGGTTCACTGCCACAGAGGTGTGAGGCAAGGAGATCCACT CTCCCCATTATTGTTCATCATGGCAATGGATGAGGTTCTTGGATTGTCGATGCCGCAACTGGGATACCAGTTCCATGACACCCTAG TAGATGGTTTTGCCTTCGCTGACGACTGGGTCGTGTGTGCAGAAAGTCAAGCTCGCCTCAAAGAGAAGCTTGAAGCTGCCGCGG TTGAGTTGGGAAGAGCTGGTATGAAGATAAACGCCCGGAAGACGAAAGCGATGGTGATCTGTGGGGACAGGAAACATCGGGC GACAGCAGTCTCAGTCGAACCATTTTGCTTCGCCGAGGAACTCATCACTCCTCTGGGTCCGACAGACACAGTAACATATTTGGGC ATCCCCTTTACTTTCAAGGGGAAAGGAGTCTTCAACCATCGACAGCATCTGCTCAAACTGCTCGACGAGGTGACGCGTGCCCCGT TGAAGCCGCATCAAAGAATGGAGATCACGAGGAACTACCTGATACCGAGGCTGACATATTCGCTCGTGCTCGGCCAAGTCCACC GAAACACGCTCAAAAGGTTGGACAACTACATCAGGCAATCTATTCGCGGCTGGCTACGTCTACCGAAGGATACCCCGATCAGCT ACATCCATGCCGGTAAACAACATGGGGGACTTGGTATCCCAAGTCTGAGTGCAACAATCCCGATGCAGCGGAGAGTGCGTATGG TAAAGCTGCTTTCTACTCAGTGCCGTGTACTACGCAATGTGGTCAACGATTCCGCATTTGGCAAGGTTGTTCGGGATCTCAGCCTT CCGATCCGTGTCCATGGTTCATGCGTAAACACCAAGGAAGAACTGGTGGCTGCTTGGGGTGATAGCCTGCACAACAGCGTTGAC GGTCGCGGTTTACGAGAGTTGGTCGCTTCGCCTCTTTCTAATCGATGGCTTGTATTTCCAGAAAGAGTGTTTTCTCGGATATTTAT CCGCGGTATCCAGCTCCGATGCAACCTGCTCAGGACGAGAGTTAGAAGCGCTCGACATGGTCACGGTGGCCAGACGATCTTATG TCGTGGAAACTGTGGCCAACCAGAGAGTTTGGTGCATATTTTGCAGTCCTGCTGGATCACGCATGACGCCAGATGTGCTCGTCAT AATCGGGTTGCAAGGGAACTCGCAAAACGCCTCCGTCGCCTGGGATACACCGTCTTTGAGGAGTTGAGAGCACCAACTTCGACG TCCTTCATCAAGCCTGACCTGATCGCCGTTCGGGAGCGCCGAGCGACTGTAATAGACGTCAGCATAGTCTCGGATGGGCGCGGA GTGACTGTGTGGAATGAGAAAAAGCAGAAGTATGGTGCTGATGAATTTTCCCTCGCCATAATCTCAGCCCTACGTGCTATTGGTT GTGATGTGGACTTTTTGGTCCACCAACCGATGATCATCTCTTATCGAGGGATCTGCTTTCCTCAATCCGCCAAAGCTGTTATCGGA CTGGGACTTTCTAAAGTCACAGTCAGTGACTTGTGTTTGCTCGCCATTGTGGGTTCTCTGCGTACGTACGACACTTTTATGCGTGG CACATGGCGTTGA (SEQ ID NO: 58)
	3' UTR	ATGTACATTCTTGCCATTGAATCTCACACCAGACCCAGTATGACGGACACTTTGTGCCTGATGTGCGAGTCTTGACTGTGCTAGCG CTTACCGCGCCTTGAAGAGCATTCAGCATTGTTTGTCCTTTCTTCGGTTGTTAGACTTTACACGCATGTTTCCTTACCAAAATTCTTA CGTACGTTGGGATTCATCCTATCTGACTGGAACTGTTGGTTGCATGACTTCGAATGAGACATTTCTTTCTTTTATCTCATATTCTCC TGTAACCCTTTCGCATTCATCGCTTGCATTCACTTTTTATGTCTGTGACCATGCTCTTCAAAAATAAACGATA (SEQ ID NO: 59)
	Codon optimized	CTGGGATCTCTGGATGCCTCTCTGGGAGAGCTGAGAGTGCAGCAGCTGCAGACAGGCCTGACCACACTGTTCGGCTTCAACGTG CTGGTCACCTTCGACAACGTGCACTACAAGACCAGCGGCGCCTCTGCTCCTGTGCCTACAAGCACCCAAGAAAGACTCGTGGGCC TGACCTGCGAGGAATGTGGCAAGTGGTGCAAGAGCAAGGCCGGACTGGTGGCCCACCACAGAGTGCACGATAATGATAGCGT GGGCACCAACATGGTGGCTCAGCTGGCTTGTGCCGACTGCAGCAGACTGTTCCCTACCAAGATCGGCCTGAGCCAGCACAGAAG GCACGCCCATCCTACACAGCACAACGCCGACAAGCTGAGCAGAGTGAAACACAGCGGAGCCCGGTGGTCCCAGCAAGAGTCTC AATCTCTGCTGCGGCTGGCCAACAATCTGTACCCCAGCTGCGAAACCCAGACAGCCCTGTTCGCTCGGCTGGAACAGTACTTCCC TGGCAGAAGCGCCATCAGCATCAAGACCAGACTGCGGGTGCTGAACTGGCAGGCTCAGCAGGATGAATCTAGCTCTGGCGGCC CTGATCAGACCATCGGACAGATCGCCGCCTACAGCTCTGAGGCCGATGATTACAGCGTGTGGTTCAAGCAGACCGTGGACTGCG CCGTGTCTCTGCTGGAATCTCACGCCGATTCTAGCCTGGCCTCCGTGGATCTGCTGGCCTTTGCTAGAGGACTGCAGAGCGGCAT CATGACCCCTGAACAGGTGCTCAGCCTGCTGGATCTGCATGCCAGCAGAACCTTTCCACACACCTGGAAAACCGTGTCCAGACGG CGTAGACAGCTGGCCCATAGAATGCCCGTGAACCGGAAGCAGATCAGACGGGCCAATTACGCCGCCATCCAGACACTGTACCAC

		CAGAGAAGAAAGGACGCCGCCTCTGCCGTGCTGGATGGCTCTTGGAAGGATCTGTACAAGGGCAACTGCGGCCTGCCTCCTGAT GCCGAGCAGTACTGGAAGCAGGTTCTGAGCGCCCCTAAGCACGTGGACAGCAGACCTTCTAGAGTGGTGGTGCCCAGCGACTG GTCCCTGATCGAACCTATCACAGGCGAGGAAGTGGGCAGAACCGTCAGATCCATGGGCAATAGCAGCCCTGGCCTGGATAAGCT GACCCCTCGGATGCTGAGAAGATTCAACGCCAATGTGCTGGCCGGCTACTTCAACCTGCTGCTGCTTTCTGGCGGCTGCCCTCCT CATCTGTGCAGAGCCAGAATCACCCTGGTGCCTAAGGTGCCCAATCCTACAAGCCCCGATCAGCTGAGGCCTATCAGCGTGTCCT CTATCCTCGTGCGGTGCTTCCACAAGGTGCTGGCTGACAGATGGTCCAGAAGGCTGCAGCTTCCCAGCCTGCAGTTCGCCTTCCT GCACAGAGATGGATGCCTGGAAGCCACAAGCCTGCTGCATGCCCTGCTGAGACACTCTTCTGCCACCGCCAGCAATCTGTCCCTG GCTTTCGTGGACATCAGCAAGGCCTTCGATAGCGTGTCCCACGACACAATCGTGCGCTCTGCCGAAGCTTTTGGCGCCCCTTCTCC TCTTGTGCGGTATATCGCCCAGAGCTACGAGAACGCCGTGGCCGTGTTTCCATCTAGCGAGGTGCACTGTCATAGAGGCGTCAG ACAGGGCGATCCTCTGAGCCCTCTGCTGTTCATTATGGCCATGGACGAGGTGCTGGGCCTGAGCATGCCTCAGCTCGGCTACCA GTTTCACGATACCCTGGTGGACGGCTTCGCCTTCGCTGATGATTGGGTTGTGTGCGCCGAGAGCCAGGCCAGACTGAAAGAGAA ACTGGAAGCTGCCGCCGTGGAACTGGGCAGAGCCGGCATGAAGATCAATGCCAGAAAGACCAAGGCCATGGTCATCTGCGGCG ACAGAAAGCACAGAGCCACAGCCGTGTCCGTGGAACCTTTCTGCTTTGCCGAGGAACTGATCACCCCTCTGGGCCCTACCGATAC CGTGACCTATCTGGGCATCCCCTTCACCTTCAAAGGCAAGGGCGTGTTCAACCACCGGCAGCATCTGCTGAAGCTGCTGGACGAA GTGACACGGGCCCCTCTGAAACCTCACCAGCGGATGGAAATCACCCGGAACTATCTGATCCCCAGACTGACCTACAGCCTGGTGC TGGGACAAGTGCACCGGAACACCCTGAAGAGACTGGACAACTACATCCGGCAGAGCATCAGAGGCTGGCTGAGACTGCCTAAG GACACCCCTATCAGCTACATCCACGCCGGCAAACAGCATGGCGGACTGGGAATCCCTAGCCTGAGCGCCACAATTCCCATGCAG AGAAGAGTGCGGATGGTCAAGCTGCTGAGCACACAGTGCAGAGTGCTGCGGAACGTGGTCAACGATAGCGCCTTTGGCAAGGT CGTGCGGGACCTGTCTCTGCCCATTAGAGTGCATGGCAGCTGTGTGAACACCAAAGAGGAACTGGTTGCCGCCTGGGGCGACA GCCTGCACAATTCTGTTGATGGCAGAGGCCTGCGCGAGCTGGTTGCTAGCCCTCTGTCTAACAGATGGCTGGTGTTCCCCGAGCG GGTGTTCAGCCGGATCTTTATCAGAGGAATCCAGCTGCGGTGCAATCTGCTGAGAACCAGAGTCAGATCCGCCAGACACGGACA TGGCGGCCAGACCATTCTGTGTAGAGGCAATTGCGGACAGCCCGAGTCTCTGGTGCACATCCTGCAGTCTTGCTGGATCACCCAC GACGCCAGATGCGCCAGGCATAACAGAGTGGCCAGAGAGCTGGCCAAGCGGCTGAGAAGGCTGGGCTACACCGTGTTCGAAG AACTGAGAGCCCCTACCTCCACCAGCTTCATCAAGCCCGATCTGATCGCCGTGCGCGAGAGAAGGGCTACAGTGATCGATGTGT CCATCGTGTCTGACGGCAGGGGCGTGACAGTGTGGAACGAGAAGAAGCAGAAGTACGGCGCCGACGAGTTCAGCCTGGCCATC ATTTCTGCCCTGAGAGCCATCGGCTGCGACGTGGACTTTCTGGTGCATCAGCCCATGATCATCAGCTACCGGGGCATCTGCTTTCC CCAGTCTGCCAAGGCTGTGATCGGACTGGGCCTGTCCAAAGTGACCGTGTCCGATCTGTGCCTGCTGGCCATCGTCGGAAGCCT GAGAACCTACGACACCTTCATGAGAGGCACCTGGCGGTGA (SEQ ID NO: 60)
R2Sm-A	**5' UTR**	ATGTTTTAATTTATTTTTGAACTACTACTGTCTGAGTGCTTCTTACAACCTGAAGGCTCAGAAACTACCCACTTTTTGCTGTTTATCC ACAACAACAGTTGTGAATCTATTCTCCAAATATTCCTTGTGCTTTTGTCAACATTATTCTATACCAACTGTACCACCTACTTCTTCAT CTCACGTTTTAATTCTGGTCTAATTTTCTCATCATTAGTCACGGAGAGGGCCTATGAACGGTCCGTGACGCGAAATTCAATCCACG AATTCGTCCTCTTCTGCTAGTGGTCCCCGAAATACGGTTCCTCTGGCCTGTCAGTTGTGTTAAAACTATATAATAACG (SEQ ID NO: 61)
	CDS	ATGCCGGTCTCAACCGGCGCAGAAACTGACATAACCTCTTCTTTGCCTATTCCTGCATCCTCAATCGTCTCGCCAAACTACACACTC CCTGATTCCTCTTCAACCTGCCTTATATGTTTCGCTATCTTCCCCACCCACAACATACTCCTCTCCCATGCCACTGCAATCCACCATA TTTCTTGTCCTCCTACTCCAGTGCAAGACGGTTCTCAGCAGATGTCTTGTGTTCTTTGCGCCGCCGCTTTTTCATCTAACAGGGGAC TAACACAACACATTCGCCACCGGCACATCTCCGAATATAACGAACTAATCAGACAACGAATTGCAGTGCAGCCGACGTCTCGCAT ATGGTCACCATTCGATGATGCTTCTCTACTATCAATCGCTAACCATGAAGCCCATAGATTCCCCACGAAGAATGACTTATACCAAC ACATCAGCACTGTATTAACACGCAGGACGGCAGAGGCCGTCAAACGCCGACTCCTCCACCTACAGTGGTCCAGATCACCCACAG CGATTACTACCTCTTCGAATAATCACACAACCACAGACATCCCCAATACCGAGGCCCGATATATTTTTCCGGTAGACCTAGACGAA CATCCACCATTGTCTGATGCCACAACCCCCGACGCATCGACACATCCACTCCCAGAACTCCTTGTCATCTTGACACCGCTTCCATCC CCGACTAGACTACAAAACATATCCGAATCACAGACCTCCCATGAATCCAATAGGAACTCAATGCATACACCGCCAACGTATGCCT GCGATTCGGATGAGTCACTAGGGGTTACTCCCTCATCAACTATCCCCTCATGCTTCCACAGTTATCGGGACCCCCTAGCTGAACAA AGAAGCAAACTCCTGAGGGCATCCGCCAGCCTACTACAAAGCAGTTGTACTCGCATACGGTCCTCCAGCCTGCTCGCCTTCCTCC AAAACGCATCCACATTAATGGACGAGGAACACGTGTCCACCTTCCTCAATAGTCATGGAGAATTCGTCTTCCCTAGAACATGGAC CCCATCCCGACCCAAACACCCCTCCCACGCCCCAGCTAATGTTTCTAGGAAGAAAAGGAGGAAAATAGAGTACGCACACATCCA GACACTCTTCCACCACCGTCCCAAAGATGCCGCCAACACCGTTCTAGACGGTCGGTGGAGAAACCCCTATGTCGCAAACCATTCA ATGATTCCAGACTTCGACTGCTTCTGGACAACAGTCTTTACTAAAACAAATTCCCCAGACAGCCGGGAGATTACTCCAATCATCCC TATGACTCCCTCTCTCATTGACCCGATCCTCCCCTCTGACGTCACATGGGCGCTGAAAGAAATGCATGGCACGGCCGGTGGGATT GATCGTCTAACATCGTACGATCTGATGAGATTCGGGAAGAATGGTCTTGCTGGATATCTCAACATGCTACTCGCTCTTGCATACCT TCCCACTAATCTCTCAACAGCACGGGTAACTTTCGTCCCCAAGTCATCAAGTCCTGTGTCACCTGAGGACTTCCGTCCCATCAGTG TCGCTCCAGTAGCCACTAGGTGCCTGCACAAAATTCTAGCAAAGAGATGGATGCCGCTCTTTCCACAGGAACGACTTCAGTTCGC TTTCCTAAACCGAGATGGATGCTTTGAAGCAGTTAATCTTCTGCACTCGGTCATACGGCACGTCCACACCCGCCATGCAGGAGCA TCCTTCGCCCTGCTCGACATATCACGGGCCTTTGACACTGTATCACATGACTCCATCATCAGAGCGGCGAAAAGATATGGGGCAC CTGAACTGTTATGCCGCTACCTCAATAACTATTACCGACGTTCAACCAGCTGCGTCAACCGCACTGAATTGCATCCTACGTGTGGG GTGAAGCAAGGAGACCCCCTGTCGCCACTCCTCTTCATCATGGTTCTCGACGAATTACTGGAAGGTCTAGATCCAATGACCCACC TAACAGTTGATGGAGAGAGCTTGAACTACATAGCTTATGCTGACGATCTCGTAGTTTTCGCTCCAAATGCAGAACTCCTTCAACG GAAACTCGATCGGATCTCCCTACTTCTACACGAGGCTGGATGGTCGATTAACCCTGAAAAAAGCCGGACCCTGGACCTAATCTCT GGTGGCCATTCCAAAATCACAGCGCTCTCTCAGACAGAATTCACCATCGCGGGGATGCGTATACCACCGCTTTCCGCCGCCGACA CCTTCGACTATCTGGGTATCAAATCCAACTTCAAGGGCCGATGCCCAGTGGCCCATATTGACTTATTGAACAACTACCTCACGGAA ATATCGTGCGCTCCACTTAAGCCGCAGCAGCGCATGAAGATCTTGAAAGATAATCTACTCCCTCGACTCCTCTACCCCCTGACTCT AGGAATAGTACACCTGAAAACCCTGAAGTCAATGGACCGAAATATCCACACGGCCATAAGGAAATGGTTGCGGCTACCCTCCGA CACCCCGCTAGCATATTTTCACTCACCCGTCGCTGCCGGAGGCCTAGGGATCCTCCATCTGTCCTCATCGGTTCCATTCCACCGTC GAAAACGTCTAGAAACCCTCCTATCTTCACCGAACCGCCTACTGCACAAGTTGCCAACTTCCCCAACACTAGCTTCTTATTCACACC TTAGTCAACTGCCAGTTCGAATTGGGCACGAGACCGTAACGTCTAGAGAAGAGGCTTCCAACAGCTGGGTGAGACGATTACATT CGTCCTGCGACGGGAAGGGACTACTCCTAGCACCACTAAGCACCGAGTCCCATGCATGGCTGCGCTACCCCCAGTCTATTTTTCC AAGTGTTTACATCAACGCCGTTAAATTACGAGGTGGCTTACTATCCACCAAAGTCAGGAGATCTCGCGGAGGTAGAGTGACGAA TGGCCTGAACTGTCGAGGCGGTTGCGCCCATCATGAAACAATCCACCACATTCTGCAACATTGCGCGCTCACCCATGACATCAGA TGCAAACGCCATAACGAACTATGCAACCTTGTGGCAAAGAAACTGCGTAGGCAAAAAATCCATTTCTTACAGGAGCCCTGCATTC

		CTCTAGAAAAAACTTACTGCAAACCTGATTTTATAATTATACGTGATTCAATTGCTTATGTTCTAGACGTCACTGTATCGGACGAC GGAAACACCCACGCCAGCCGCCTGTTAAAAATATCAAAATACGGCAATGAGCGAACCGTCGCATCGATCAAGCGATTCCTCACAT CCAGTGGATATATCATTACCAGTGTTCGACAAACACCAGTCCTTACATTCAGAGGTATTCTGGAGAGAGCAAGTTCACAATCCCT ACGACGCCTATGTTTTTCGTCCCGTGACCTCGGTGACCTTTGCCTGAGTGCGATTCAAGGCTCAATTAAAATATATAATACCTATA TGAGAGGAACCCAACGGCTGAACGAATAG (SEQ ID NO: 62)
	3' UTR	CCCCCTTCACTCTTAGACATTCCCCCACTGTTGTTGCTTATCTTCATGTTTTTGTGTTAATTGACTGCTCTCTTCTGGGTTGATGTCT GATTGTCTCTCTCTCTTTCCATATTGCTTGCTCTCCCCGCTTACTTCCAATAGTTGTCATATTATGTCTTTGTTTACTTGCCATGTCTA ACGACAATTACTTTATCTACCTTAGTTGGTCCTCTTGGTTTGGTTGCCTTCATGTGTTCATGGCGGAATCTGATGTTTATAATGACT ATTCCTACTACCACCATTACAACTATTATTATTATCACTATTATTAACATTATTATTACTTCTACAATTAGTATTATGGCTACTCCTTT CAGCACACCAATAAAATCTCAATCAAACATCTCACTTATTAAACTCTCTATTTCCCCTTCGTTATAAACTTACAATTCAGTTTAACCG AATATCTCTCTTTTACAAATCTTAAGTATGTAATTTTGTGCCAAGCCCATTTGGGTCTGTACAATTTGATACTTAAAAATAAATGTT AT (SEQ ID NO: 63)
	Codon optimized	ATGCCAGTGTCTACAGGCGCCGAGACAGACATCACAAGCAGCCTGCCTATTCCTGCCAGCAGCATCGTGTCCCCAAACTACACCC TGCCTGACAGCAGCAGCACCTGTCTGATCTGCTTCGCCATCTTTCCCACACACAACATCCTGCTGAGCCACGCCACAGCCATCCAC CACATCAGCTGTCCTCCAACACCTGTGCAGGATGGCAGCCAGCAGATGAGCTGTGTGCTGTGTGCCGCCGCTTTCAGCAGCAAC AGAGGACTGACCCAGCACATCCGGCACAGACACATCAGCGAGTACAACGAGCTGATCCGGCAGAGAATCGCCGTGCAGCCCAC CAGCAGAATCTGGTCCCCATTCGATGATGCCAGCCTGCTGTCTATCGCCAACCACGAGGCCCACAGATTCCCCACCAAGAACGAC CTGTATCAGCACATCTCCACCGTGCTGACCAGACGGACAGCCGAGGCTGTGAAAAGACGGCTGCTGCATCTGCAGTGGTCTAGA AGCCCTACCGCCATCACCACCAGCTCCAACAACCACACCACCACAGACATCCCCAACACAGAGGCCCGGTACATCTTCCCCGTGG ACCTGGATGAACACCCTCCTCTGTCCGATGCCACCACACCAGACGCCTCTACACACCCTCTGCCTGAGCTGCTGGTCATCCTGACA CCTCTGCCTTCTCCAACCAGACTGCAGAACATCTCCGAGAGCCAGACCAGCCACGAGAGCAACAGAAACAGCATGCACACCCCTC CAACCTACGCCTGCGATTCCGATGAGTCTCTGGGCGTGACACCCAGCAGCACAATCCCTAGCTGCTTCCACAGCTACAGGGACCC TCTGGCCGAGCAGAGAAGCAAACTGCTGAGAGCCTCTGCCTCTCTGCTGCAGAGCAGCTGCACCAGAATCAGAAGCTCTAGCCT GCTGGCCTTCCTGCAGAACGCCAGCACACTGATGGACGAGGAACACGTGTCCACCTTTCTGAACAGCCACGGCGAGTTCGTGTT CCCCAGAACCTGGACACCCTCCAGACCTAAGCACCCTTCTCATGCCCCTGCCAACGTGTCCAGAAAGAAGCGGCGGAAGATCGA GTACGCCCACATCCAGACACTGTTCCACCACCGGCCTAAGGACGCCGCCAATACTGTGCTGGATGGAAGATGGCGGAACCCCTA CGTGGCCAACCACAGCATGATCCCCGACTTCGACTGCTTCTGGACCACCGTGTTCACCAAGACAAACAGCCCCGACTCCAGAGAG ATCACCCCTATCATCCCCATGACTCCCAGCCTGATCGACCCCATCCTGCCTTCCGATGTGACATGGGCCCTGAAAGAGATGCACG GAACAGCCGGCGGAATCGACAGACTGACCAGCTACGACCTGATGAGATTCGGCAAGAATGGCCTGGCCGGCTACCTGAATATG CTGCTCGCTCTGGCCTACCTGCCTACCAATCTGAGCACCGCCAGAGTGACCTTCGTGCCCAAGTCTAGCAGCCCCGTGTCTCCCGA GGACTTCAGACCTATTTCTGTGGCCCCTGTGGCCACCAGATGCCTGCACAAGATTCTGGCCAAGCGGTGGATGCCTCTGTTCCCT CAAGAGAGACTGCAGTTCGCCTTCCTCAACAGAGATGGCTGCTTCGAGGCCGTGAACCTGCTGCACTCTGTGATCAGGCACGTG CACACAAGACATGCCGGCGCTAGCTTTGCCCTGCTGGATATCTCCAGAGCCTTCGACACCGTGTCTCACGACAGCATCATCAGAG CCGCCAAGAGATATGGCGCCCCAGAGCTGCTGTGCAGATACCTGAACAACTACTACCGGCGGAGCACCAGCTGCGTGAACAGA ACAGAACTGCACCCTACCTGCGGCGTGAAGCAAGGCGATCCTTTGAGCCCTCTGCTGTTCATCATGGTGCTGGACGAACTGCTGG AAGGACTGGACCCCATGACACACCTGACAGTGGATGGCGAGAGCCTGAACTATATCGCCTACGCCGACGACCTGGTGGTGTTCG CCCCTAATGCTGAACTGCTGCAGCGGAAGCTGGACCGGATTAGTCTGCTGTTGCATGAGGCCGGCTGGTCCATCAATCCCGAGA AGTCTAGAACCCTGGACCTGATCTCTGGCGGCCACTCCAAGATCACAGCCCTGAGCCAGACAGAGTTCACCATTGCCGGCATGC GGATCCCTCCACTGTCTGCCGCCGATACCTTTGACTACCTGGGCATCAAGAGCAACTTCAAGGGCAGATGCCCCGTGGCTCACAT CGACCTGCTGAACAATTACCTGACCGAGATCAGCTGCGCCCCTCTGAAGCCTCAGCAGCGGATGAAGATCCTGAAGGACAATCT GCTGCCCCGGCTGCTGTACCCTCTGACACTGGGAATCGTGCACCTGAAAACCCTGAAGTCCATGGATCGGAACATCCACACCGCC ATCCGGAAGTGGCTGAGACTGCCTAGCGATACCCCACTGGCCTACTTCCATTCTCCTGTGGCTGCTGGCGGCCTGGGCATTCTGC ATCTGTCTAGCTCCGTGCCTTTCCACAGACGGAAGCGGCTGGAAACACTGCTGTCAAGCCCCAACAGACTGCTGCACAAGCTGCC TACAAGCCCCACACTGGCCAGCTACTCTCACCTGTCTCAGCTGCCTGTGCGGATCGGACACGAGACAGTGACCTCTAGAGAGGA AGCCAGCAACTCCTGGGTCCGAAGGCTGCACAGCTCCTGTGATGGAAAGGGACTGCTGCTTGCCCCTCTGTCCACAGAATCTCAC GCCTGGCTGAGATACCCTCAGAGCATCTTCCCTAGCGTGTACATCAACGCCGTGAAGCTGAGAGGCGGACTGCTGAGCACAAAA GTGCGGAGATCTAGAGGCGGCAGAGTGACAAACGGCCTGAATTGCAGAGGCGGCTGTGCCCACCACGAGACAATTCACCACAT CCTGCAGCACTGCGCCCTGACACACGACATCAGATGCAAGCGGCACAACGAACTGTGCAACCTGGTGGCTAAGAAGCTGCGGA GACAGAAGATCCACTTCCTGCAAGAGCCCTGCATTCCCCTGGAAAAGACCTACTGCAAGCCCGACTTCATCATCATCCGGGACAG CATTGCCTACGTCCTGGACGTGACCGTGTCCGACGATGGAAATACCCACGCCTCCAGGCTGCTGAAGATCTCTAAGTACGGCAAC GAGCGGACCGTGGCCAGCATCAAGCGGTTTCTGACAAGCAGCGGCTACATCATCACCAGCGTGCGCCAAACTCCTGTGCTGACC TTTCGGGGCATCCTGGAAAGAGCCAGCTCTCAGTCTCTGCGGAGGCTGTGCTTCAGCTCCAGAGATCTGGGCGATCTGTGCCTG AGCGCCATCCAGGGCTCCATCAAGATCTACAACACCTACATGCGGGGCACCCAGCGGCTGAATGAATGA (SEQ ID NO: 64)

R2bm	**5' UTR**	
	CDS	
	3' UTR	
	Codon optimized	ATGATGGCCAGCACAGCCCTGTCTCTGATGGGCAGATGCAACCCTGATGGCTGCACCAGAGGCAAGCACGTGACAGCCGCTCCT ATGGATGGACCTAGAGGCCCTTCTTCTCTGGCCGGCACATTTGGATGGGGCCTTGCTATTCCTGCCGGCGAGCCTTGTGGCAGA GTGTGTTCTCCTGCCACCGTGGGATTCTTCCCAGTGGCCAAGAAGTCCAACAAAGAGAACAGACCCGAGGCCAGCGGCCTGCCT CTGGAATCTGAAAGAACCGGCGACAACCCTACAGTGCGGGGATCTGCTGGTGCCGATCCTGTTGGACAAGATGCCCCTGGATGG ACATGCCAGTTCTGCGAGAGAACCTTCAGCACCAACAGAGGCCTGGGCGTGCACAAGAGAAGGGCCCATCCTGTGGAAACAAA CACCGACGCTGCCCCTATGATGGTCAAGAGAAGATGGCACGGCGAGGAAATCGACCTGCTGGCCAGAACAGAAGCCAGACTGC TGGCTGAGAGGGGCCAATGTTCTGGCGGCGATCTGTTTGGAGCCCTGCCTGGCTTTGGCAGAACCCTGGAAGCCATCAAGGGAC AGCGCAGAAGAGAGCCCTATAGAGCCCTGGTGCAGGCCCACCTGGCCAGATTTGGATCTCAGCCTGGACCTAGCAGCGGCGGC TGTTCTGCCGAACCTGATTTTCGGAGAGCCTCTGGCGCTGAAGAGGCCGGCGAAGAAAGATGTGCCGAAGATGCCGCCGCTTAC GATCCTTCTGCTGTGGGCCAGATGAGCCCCGATGCTGCTAGAGTGCTGAGCGAACTGCTTGAAGGCGCCGGACGTAGAAGGGC TTGTAGAGCCATGAGGCCTAAGACCGCCGGCAGACGGAATGACCTGCACGACGATAGAACAGCCAGCGCTCACAAGACCAGCA

		GACAGAAGCGGAGAGCCGAGTACGCCAGAGTGCAAGAGCTGTACAAGAAGTGCAGAAGCCGGGCTGCCGCCGAAGTGATTGA TGGTGCTTGTGGTGGCGTGGGCCACAGCCTGGAAGAGATGGAAACCTACTGGCGCCCCATCCTGGAAAGAGTGTCTGACGCTCC TGGGCCTACACCTGAAGCTCTGCATGCTCTGGGCAGAGCCGAATGGCATGGCGGCAACAGAGATTACACCCAGCTGTGGAAGC CCATCAGCGTGGAAGAAATCAAGGCCAGCAGATTCGACTGGCGGACAAGCCCTGGACCTGACGGCATTAGAAGCGGACAATGG CGAGCCGTGCCTGTGCACCTGAAGGCCGAGATGTTCAACGCCTGGATGGCCAGAGGCGAGATCCCCGAGATTCTGAGACAGTG CAGAACCGTGTTCGTGCCCAAGGTGGAAAGACCTGGTGGCCCTGGCGAGTACAGACCCATCTCTATCGCCAGCATTCCTCTGCG GCACTTCCACTCCATCCTGGCTAGAAGGCTGCTGGCTTGCTGCCCTCCTGATGCCAGACAGAGAGGCTTCATCTGCGCCGATGGC ACCCTGGAAAATTCCGCCGTGCTGGATGCAGTGCTGGGCGACAGCAGAAAGAAACTGCGCGAATGTCACGTGGCCGTCCTGGA TTTCGCCAAGGCCTTCGATACCGTGTCTCACGAGGCTCTGGTGGAACTGCTGAGACTGAGGGGAATGCCCGAGCAGTTCTGTGG CTATATCGCCCACCTGTACGACACCGCCTCTACCACACTGGCCGTGAACAACGAGATGAGCAGCCCCGTGAAAGTCGGAAGAGG CGTTAGACAGGGCGACCCTCTGAGCCCTATCCTGTTCAACGTGGTCATGGACCTGATCCTGGCCAGCCTGCCTGAGAGAGTGGG CTATAGACTGGAAATGGAACTGGTGTCTGCCCTGGCCTACGCCGATGATCTGGTTCTGCTCGCCGGAAGCAAAGTGGGCATGCA AGAGTCTATCAGCGCCGTGGATTGCGTGGGCAGACAGATGGGCCTGCGCCTGAATTGCAGAAAGTCTGCCGTGCTGAGCATGA TCCCCGACGGCCACAGAAAGAAGCACCACTACCTGACCGAGCGGACCTTCAACATCGGCGGCAAGCCACTGAGACAGGTGTCCT GCGTTGAGCGGTGGCGGTATCTGGGAGTCGATTTTGAGGCCTCCGGCTGCGTGACACTGGAACACTCTATTAGCAGCGCCCTGA ACAACATCAGCAGAGCCCCTCTGAAGCCCCAGCAGAGGCTGGAAATTCTGAGAGCCCATCTGATCCCTCGGTTCCAGCATGGCTT CGTGCTGGGCAATATCAGCGACGACCGGCTGAGAATGCTGGACGTGCAGATCAGAAAGGCCGTCGGCCAGTGGCTGAGACTTC CTGCCGATGTGCCTAAGGCCTACTATCATGCCGCTGTGCAGGATGGCGGACTGGCCATTCCTAGCGTGCGGGCCACAATTCCCG ATCTGATCGTGCGGAGATTCGGCGGCCTTGATAGCTCTCCTTGGAGCGTGGCAAGAGCCGCTGCCAAGAGCGACAAGATCCGG AAGAAACTGAGATGGGCTTGGAAGCAGCTGCGGCGGTTCAGCAGAGTGGATTCCACAACACAGAGGCCCTCCGTGCGGCTGTT TTGGAGAGAACATCTGCACGCCTCCGTGGACGGCAGAGAGCTGAGAGAGAGCACCAGAACACCCACCAGCACCAAGTGGATCA GAGAGAGATGCGCCCAGATCACCGGCAGAGACTTCGTGCAGTTTGTGCACACCCACATCAACGCCCTGCCTAGCAGAATCAGAG GCAGCAGGGGTAGAAGAGGCGGCGGAGAGTCAAGCCTGACATGTAGAGCCGGCTGCAAAGTGCGCGAGACAACAGCCCATAT CCTGCAGCAGTGTCACAGAACACACGGCGGCAGAATCCTGCGGCACAACAAGATCGTGTCCTTCGTGGCCAAGGCCATGGAAG AGAACAAGTGGACCGTGGAACTGGAACCCAGACTGAGAACAAGCGTGGGCCTGAGAAAGCCCGACATCATTGCCAGCAGAGAT GGCGTGGGAGTGATCGTGGATGTGCAGGTTGTGTCCGGGCAGAGATCCCTGGATGAGCTGCATAGAGAGAAGCGGAACAAAT ACGGCAACCACGGCGAGCTGGTCGAACTGGTTGCTGGTAGACTGGGCCTGCCAAAGGCCGAGTGTGTCAGAGCCACAAGCTGC ACCATCTCTTGGAGAGGCGTGTGGTCCCTGACCAGCTACAAAGAGCTGCGGAGCATCATCGGACTGAGAGAGCCCACACTGCAG ATCGTGCCTATTCTGGCCCTGAGAGGCTCCCACATGAACTGGACCCGGTTCAACCAGATGACCAGCGTGATGGGAGGCGGCGTG GGATAA (SEQ ID NO: 65)

[0087] A non-LTR retrotransposon may comprise multiple retrotransposon polypeptides or polynucleotides encoding same. In some embodiments, the retrotransposon polypeptides may form a complex. For example, a non-LTR retrotransposon is a dimer, e.g., comprising two retrotransposon polypeptides forming a dimer. The dimer subunits may be connected or form a tandem fusion. A Cas protein or polypeptide may be associate with (e.g., connected to) one or more subunits of such complex. In some examples, the non-LTR retrotransposon is a dimer of two retrotransposon polypeptides; one of the retrotransposon polypeptides comprises nuclease or nickase activity and is connected with a Cas protein or polypeptide.

[0088] The retrotransposon polypeptides may be enzymes or variants thereof. In some examples, a retrotransposon polypeptide may be a reverse transcriptase, a nuclease, a nickase, a transposase, nucleic acid polymerase, ligase, or a combination thereof. In one example, a retrotransposon polypeptide is a reverse transcriptase. In another example, a retrotransposon polypeptide is a nuclease. In another example, a retrotransposon polypeptide is nickase. In a particular example, a non-LTR retrotransposon comprises a first retrotransposon polypeptide and a second retrotransposon polypeptide, wherein the second retrotransposon polypeptide comprises nuclease or nickase activity. In certain cases, a retrotransposon polypeptide may comprise an inactive enzyme. For example, a retrotransposon polypeptide may comprise a

nuclease domain that is inactivated. Such inactivated domain may serve as a nucleic acid binding domain.

[0089] The retrotransposon polypeptides may comprise one or more modifications to, for example, enhance specificity or efficiency of donor polynucleotide recognition, target-primed template recognition (TPTR), and/or reduce or eliminate homing function. The retrotransposon polypeptides may also comprise one or more truncations or excisions to remove domains or regions of wild-type protein to arrive at a minimal polypeptide that retain donor polynucleotide recognition and TPTR. In some example embodiments, the native endonuclease activity may be mutated to eliminate endonuclease activity.

[0090] In certain example embodiments, the modifications or truncations of the non-LTR retrotransposon peptide may be in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain.

[0091] A non-LTR retrotransposon may comprise polynucleotide encoding one or more retrotransposon RNA molecules. The polynucleotide may comprise one or more regulatory elements. The regulatory elements may be promoters. The regulatory elements and promoters on the polynucleotides include those described throughout this application. For example, the polynucleotide may comprise a pol2 promoter, a pol3 promoter, or a T7 promoter.

[0092] In some cases, the polynucleotide encodes a retrotransposon RNA with at least a portion of its sequence complementary to a target sequence. For example, the 3' end of the retrotransposon RNA may be complementary to a target sequence. The RNA may be complementary to a portion of a nicked target sequence. In some embodiments, a retrotransposon RNA may comprise one or more donor polynucleotides. In certain cases, a retrotransposon RNA may encode one or more donor polynucleotides.

[0093] A retrotransposon RNA may be capable of binding to a retrotransposon polypeptide. Such retrotransposon RNA may comprise one or more elements for binding to the retrotransposon polypeptide. Examples of binding elements include hairpin structures, pseudoknots (e.g., a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem), stem loops, and bulges (e.g., unpaired stretches of nucleotides located within one strand of a nucleic acid duplex). In certain examples, the retrotransposon RNA comprises one or more hairpin structures. In some examples, the retrotransposon RNA comprises one or more pseudoknots. In certain examples, a retrotransposon RNA comprises a sequence encoding a donor polynucleotide and one or more binding elements for forming a complex with the

retrotransposon polypeptide. The binding elements may be located on the 5' end, the 3' end, or a location in between.

[0094] In some embodiments, a retrotransposon RNA comprises a region capable of hybridizing with an overhang of a target polynucleotide at the target site. The overhang may be a stretch of single-stranded DNA. The overhang may function as a primer for reverse transcription of at least a portion of the retrotransposon RNA to a cDNA. In some cases, a region of the cDNA may be capable of hybridizing a second overhang of the target polynucleotide. The second overhang may function as a primer for the synthesis of a second strand to generate a double-stranded cDNA. The cDNA may comprise a donor polynucleotide sequence. The two overhangs may be from different strands of the target polynucleotide.

Donor Construct

[0095] The systems may comprise one or more donor constructs comprising one or more donor polynucleotide sequences for insertion into a target polynucleotide. The donor construct comprises one or more binding elements. Examples of binding elements include hairpin structures, pseudoknots (e.g., a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem), stem loops, and bulges (e.g., unpaired stretches of nucleotides located within one strand of a nucleic acid duplex). In certain examples, the retrotransposon RNA comprises one or more hairpin structures. In some examples, the retrotransposon RNA comprises one or more pseudoknots. In certain examples, a retrotransposon RNA comprises a sequence encoding a donor polynucleotide and one or more binding elements for interacting to the retrotransposon polypeptide.

[0096] In certain example embodiments, the donor construct comprises a 5' binding element and a 3' binding element with a donor polynucleotide sequence located between the 5' and 3' prime binding element.

[0097] A donor polynucleotide may be any type of polynucleotides, including, but not limited to, a gene, a gene fragment, a non-coding polynucleotide, a regulatory polynucleotide, a synthetic polynucleotide, etc.

[0098] A target polynucleotide may comprise a protospacer adjacent motif (PAM) sequence. An example of the PAM sequence is AT.

[0099] The donor construct may further comprise one or more processing element. The processing element is an element that may be added to ensure accurate processing and incorporation of the donor polynucleotide sequence by the fusion proteins disclosed herein. Example processing elements include, but are not limited to, LRNA processing elements (e.g.

GGCTCGTTGGGAGGTCCCGGGTTGAAATCCCGGACGAGCCCG (SEQ ID NO: 66)), human 28s processing elements (e.g. TAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT CCCACTGTCCCTACCTACTATCCAGCGAAACCACAGCCAAGGGAA (SEQ ID NO: 67)), and natural retrotransposon processing elements such as R2 processing elements from *Bombyx mori* (e.g. tagccaaatgcctcgtcatctaattagtgacgcgcatgaatggattaacgagattcccactgtccctatctactatctagcgaaaccacag ccaagggaacgggcttgggagaatcagcggggaa (SEQ ID NO: 68)).

[0100] The donor construct may comprise one or more homology sequence. A homology sequence is a sequence that shares or complete or partial homology with a target sequence at the site the targeted site of insertion. The homology sequence may be located on the 5' end, '3 end, or on both the 5' and 3' end of the donor construct. In certain example embodiments, the homology sequence is only located on the 5' end of the donor construct. In certain example embodiments, the homology sequence is located only on the 3' end of the donor construct. In certain example embodiments, the location of the homology sequence may depend on whether the site-specific nuclease is being directed to create a nick or cut 5' or 3' of the targeted insertion site, e.g. a 5' homology sequence on the donor construct may be used when the site specific nuclease creates a nick or cut 5' of the targeted insertion site and a 3' homology sequence may be used when the site-specific nuclease is configured to create a nick or cut 3' of the targeted insertion site. In certain example embodiments, the homology sequence is included on both the 5' and 3' ends of the donor construct regardless of whether the site-specific nuclease creates a nick or cut 5' or 3' of the targeted insertion site. In certain example embodiments the donor construct may comprise in a 5' to 3', a binding element, and the donor sequence In certain example embodiments the donor construct may comprise in a 5' to 3' direction a homology sequence, a binding element, and the donor sequence. In certain example embodiments the donor construct may comprise in a 5' to 3' direction a homology sequence, a first binding element, the donor sequence, and second binding element. In certain example embodiments, the donor construct may comprise in a 5' to 3' direction a first homology sequence, a first binding element, the donor sequence, and a second homology sequence. In certain example embodiments the donor construct may comprise, in a 5' to 3' direction, a first homology sequence, a first binding element, the donor sequence, a second binding element, and a second homology sequence. In certain example embodiments, the donor construct may comprise, in a 5' to 3' direction, the donor sequence and a binding element. In certain example embodiments, the donor construct may comprise, in a 5' to 3' direction, the donor sequence, a

binding element, and a homology sequence. A processing element may be further incorporated 3’ of the donor sequence in any of the above donor construct configurations.

[0101] The homology sequence may have at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 bases of homology to the target DNA. In certain example embodiments, the homology sequence may have between 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 base pairs of homology to the target sequence. In embodiments, with a homology sequence on both the 5’ and 3’ end of the donor construct, the size of the homology may be the same or different on each end. In some examples, the homology sequence comprises from 1 to 30, from 4 to 10, or from 10 to 25 nucleotides. For example, the homology sequence comprises from 4 to 10 nucleotides. For example, the homology sequence comprises from 10 to 25 nucleotides. For example, the homology sequence comprises 1 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides.

[0102] The donor polynucleotides may be inserted to the upstream or downstream of the PAM sequence of a target polynucleotide. For example, the donor polynucleotide may be inserted at a position between 10 bases and 200 bases, e.g., between 20 bases and 150 bases, between 30 bases and 100 bases, between 45 bases and 70 bases, between 45 bases and 60 bases, between 55 bases and 70 bases, between 49 bases and 56 bases or between 60 bases and 66 bases, from a PAM sequence on the target polynucleotide. In some cases, the insertion is at a position upstream of the PAM sequence. In some cases, the insertion is at a position downstream of the PAM sequence. In some cases, the insertion is at a position from 49 to 56 bases or base pairs downstream from a PAM sequence. In some cases, the insertion is at a position from 60 to 66 bases or base pairs downstream from a PAM sequence.

[0103] In a strand of a polynucleotide, anything towards the 5' end of a reference point is "upstream" of that point, and anything towards the 3’ end of a reference point is “downstream” of that point. A location upstream of a PAM sequence refers to a location at the 5’ side of the PAM sequence on the PAM-containing strand of the target sequence. A location downstream of a PAM sequence refers to a location at the 3’ side of the PAM sequence on the PAM-containing strand of the target sequence.

[0104] The compositions and systems herein may be used to insert a donor polynucleotide with desired orientation. For example, appropriate homology sequence may be selected to control the orientation of insertion on the 5’ or 3’ strand of the target sequence.

[0105] The donor polynucleotide comprises a homology sequence of a region of the target sequence. The homology sequence may share at least 50%, at least 60%, at least 70%, at least

80%, at least 90%, at least 95%, or 100% sequence identity with the region of the target sequence. In an example, the homology sequence shares 100% sequence identity with the region of the target sequence.

[0106] In some embodiments, the donor polynucleotide may be inserted to the strand on the target sequence that contains the PAM (e.g., the PAM sequence of the site-specific nuclease such as Cas). In such cases, the donor polynucleotide may comprise a homology sequence of a region on the PAM containing strand of the target sequence. Such region may comprise the PAM sequence. The region may be at the 3' side of the cleavage site of the site-specific nuclease. In some examples, the homology sequence may comprise from 4 to 10, or from 10 to 25 nucleotides in length. An example of such homology sequence may be of the "h1" region shown in FIG. 36.

[0107] In some embodiments, the donor polynucleotide may be inserted to the strand on the target sequence that binds to the guide, e.g., the strand that contains a guide-binding sequence. In such cases, the donor polynucleotide may comprise a homology sequence of a region that comprises at least a portion of the guide-binding sequence. In some cases, the region may comprise the entire guide-binding sequence. Such region may further comprise a sequence at the 3' side of the guide-binding sequence. For example, the region may comprise from 5 to 15 nucleotides, e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 nucleotides from the 3' side of the guide-binding sequence. In some cases, the region may be adjacent to the R-loop of the guide. For example, in the cases where the guide forms a RNA-DNA duplex with the guide-binding sequence, the region comprises a sequence at the 3' side from the RNA-DNA duplex, e.g., from 5 to from 5 to 15 nucleotides, e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 nucleotides from the 3' side from the RNA-DNA duplex. An example of such homology sequence may be of the "h2" region shown in FIG. 36.

[0108] In some examples, the homology sequence is of a region on the target sequence at 3' side of a PAM-containing strand. In certain examples, the homology sequence is of a region on the target sequence 10 nucleotides from 3' side of a RNA-DNA duplex formed by a guide molecule and a target sequence. For example, the guide molecule forms a RNA-DNA duplex with the target sequence, and the homology sequence is of a region on the target sequence 5 to 15 nucleotides from 3' side of the RNA-DNA duplex. In some embodiments, the donor polynucleotide is inserted to a region on the target sequence that is 3' side of a PAM-containing strand. In some cases, the donor polynucleotide is inserted to a region on the target sequence that is 3' side of a sequence complementary to the guide molecule.

[0109] The donor polynucleotide may be used for editing the target polynucleotide. In some cases, the donor polynucleotide comprises one or more mutations to be introduced into the target polynucleotide. Examples of such mutations include substitutions, deletions, insertions, or a combination thereof. The mutations may cause a shift in an open reading frame on the target polynucleotide. In some cases, the donor polynucleotide alters a stop codon in the target polynucleotide. For example, the donor polynucleotide may correct a premature stop codon. The correction may be achieved by deleting the stop codon or introduces one or more mutations to the stop codon. In other example embodiments, the donor polynucleotide addresses loss of function mutations, deletions, or translocations that may occur, for example, in certain disease contexts by inserting or restoring a functional copy of a gene, or functional fragment thereof, or a functional regulatory sequence or functional fragment of a regulatory sequence. A functional fragment refers to less than the entire copy of a gene by providing sufficient nucleotide sequence to restore the functionality of a wild type gene or non-coding regulatory sequence (e.g. sequences encoding long non-coding RNA). In certain example embodiments, the systems disclosed herein may be used to replace a single allele of a defective gene or defective fragment thereof. In another example embodiment, the systems disclosed herein may be used to replace both alleles of a defective gene or defective gene fragment. A "defective gene" or "defective gene fragment" is a gene or portion of a gene that when expressed fails to generate a functioning protein or non-coding RNA with functionality of a the corresponding wild-type gene. In certain example embodiments, these defective genes may be associated with one or more disease phenotypes. In certain example embodiments, the defective gene or gene fragment is not replaced but the systems described herein are used to insert donor polynucleotides that encode gene or gene fragments that compensate for or override defective gene expression such that cell phenotypes associated with defective gene expression are eliminated or changed to a different or desired cellular phenotype.

[0110] In certain embodiments, the donor may include, but not be limited to, genes or gene fragments, encoding proteins or RNA transcripts to be expressed, regulatory elements, repair templates, and the like. According to the invention, the donor polynucleotides may comprise left end and right end sequence elements that function with transposition components that mediate insertion.

[0111] In certain cases, the donor polynucleotide manipulates a splicing site on the target polynucleotide. In some examples, the donor polynucleotide disrupts a splicing site. The disruption may be achieved by inserting the polynucleotide to a splicing site and/or introducing

one or more mutations to the splicing site. In certain examples, the donor polynucleotide may restore a splicing site. For example, the polynucleotide may comprise a splicing site sequence.

[0112] The donor polynucleotide to be inserted may has a size from 5 bases to 50 kb in length, e.g., from 50 to 40kb, from 100 and 30 kb, from 100 bases to 300 bases, from 200 bases to 400 bases, from 300 bases to 500 bases, from 400 bases to 600 bases, from 500 bases to 700 bases, from 600 bases to 800 bases, from 700 bases to 900 bases, from 800 bases to 1000 bases, from 900 bases to from 1100 bases, from 1000 bases to 1200 bases, from 1100 bases to 1300 bases, from 1200 bases to 1400 bases, from 1300 bases to 1500 bases, from 1400 bases to 1600 bases, from 1500 bases to 1700 bases, from 600 bases to 1800 bases, from 1700 bases to 1900 bases, from 1800 bases to 2000 bases, from 1900 bases to 2100 bases, from 2000 bases to 2200 bases, from 2100 bases to 2300 bases, from 2200 bases to 2400 bases, from 2300 bases to 2500 bases, from 2400 bases to 2600 bases, from 2500 bases to 2700 bases, from 2600 bases to 2800 bases, from 2700 bases to 2900 bases, from 2800 bases to 3000 bases, from 2900 bases to 3100 bases, from 3000 bases to 3200 bases, from 3100 bases to 3300 bases, from 3200 bases to 3400 bases, from 3300 bases to 3500 bases, from 3400 bases to 3600 bases, from 3500 bases to 3700 bases, from 3600 bases to 3800 bases, from 3700 bases to 3900 bases, from 3800 bases to 4000 bases, from 3900 bases to 4100 bases, from 4000 bases to 4200 bases, from 4100 bases to 4300 bases, from 4200 bases to 4400 bases, from 4300 bases to 4500 bases, from 4400 bases to 4600 bases, from 4500 bases to 4700 bases, from 4600 bases to 4800 bases, from 4700 bases to 4900 bases, or from 4800 bases to 5000 bases in length.

CRISPR-Cas systems

[0113] The retrotransposon, e.g., retrotransposon polypeptide(s) may be associated with one or more components of a CRISPR-Cas system, e.g., a Cas protein or polypeptide. The complex of Cas and retrotransposon may be directed to or recruited to a region of a target polynucleotide by sequence-specific binding of a CRISPR-Cas complex. In certain example embodiments, the retrotransposon (e.g., retrotransposon polypeptide(s)) may be connected to, fused or tethered (e.g. by a linker) to, or otherwise form a complex with one or more components in a CRISPR-Cas system, e.g., Cas protein, guide molecule etc.).

[0114] The systems herein may comprise one or more components of a CRISPR-Cas system. The one or more components of the CRISPR-Cas system may serve as the nucleotide-binding component in the systems. The nucleotide-binding molecule may be a Cas protein or polypeptide (used interchangeably with CRISPR protein, CRISPR enzyme, Cas effector, CRISPR-Cas protein, CRISPR-Cas enzyme), a fragment thereof, or a mutated form thereof. The Cas protein may have reduced or no nuclease activity. For example, the Cas protein may

be an inactive or dead Cas protein (dCas). The dead Cas protein may comprise one or more mutations or truncations. In some examples, the DNA binding domain comprises one or more Class 1 (e.g., Type I, Type III, Type VI) or Class 2 (e.g., Type II, Type V, or Type VI) CRISPR-Cas proteins. In certain embodiments, the sequence-specific nucleotide binding domains directs a transposon to a target site comprising a target sequence and the transposase directs insertion of a donor polynucleotide sequence at the target site. In certain example embodiments, the transposon component includes, associates with, or forms a complex with a CRISPR-Cas complex. In one example embodiment, the CRISPR-Cas component directs the transposon component and/or transposase(s) to a target insertion site where the transposon component directs insertion of the donor polynucleotide into a target nucleic acid sequence.

[0115] In general, a CRISPR-Cas or CRISPR system as used in herein and in documents, such as WO 2014/093622 (PCT/US2013/074667), refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). See, e.g., Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems", Molecular Cell, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008.

[0116] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the effector protein complex as disclosed herein to the target locus of interest. In some embodiments, the PAM may be a 5' PAM (i.e., located upstream of the 5' end of the protospacer). In other embodiments, the PAM may be a 3' PAM (i.e., located downstream of the 5' end of the protospacer). The term "PAM" may be used interchangeably with the term "PFS" or "protospacer flanking site" or "protospacer flanking sequence".

[0117] In a preferred embodiment, the CRISPR effector protein may recognize a 3' PAM. In certain embodiments, the CRISPR effector protein may recognize a 3' PAM which is 5'H, wherein H is A, C or U.

[0118] In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. The term "target RNA" refers to a RNA polynucleotide being or comprising the target sequence. In other words, the target RNA may be a RNA polynucleotide or a part of a RNA polynucleotide to which a part of the gRNA, i.e. the guide sequence, is designed to have complementarity and to which the effector function mediated by the complex comprising CRISPR effector protein and a gRNA is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

Cas proteins and polypeptides

[0119] The CRISPR-Cas systems herein may comprise a Cas protein and a guide molecule. In some embodiments, the system comprises one or more Cas proteins. The Cas proteins may be Type II or V Cas proteins, e.g., Cas proteins of Type II or V CRISPR-Cas systems.

[0120] A CRISPR-Cas system or CRISPR system refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system).

[0121] In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50

bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0122] Examples of Cas proteins include those of Class 1 (e.g., Type I, Type III, and Type IV) and Class 2 (e.g., Type II, Type V, and Type VI) Cas proteins, e.g., Cas9, Cas12 (e.g., Cas12a, Cas12b, Cas12c, Cas12d), Cas13 (e.g., Cas13a, Cas13b, Cas13c, Cas13d,), CasX, CasY, Cas14, variants thereof (e.g., mutated forms, truncated forms), homologs thereof, and orthologs thereof.

[0123] The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related.

Class 2 Cas proteins

[0124] In certain example embodiments, the Cas protein is the Cas protein of a Class 2 CRISPR-Cas system (i.e., a Class 2 Cas protein). A Class 2 CRISPR-Cas system may be of a subtype, e.g., Type II-A, Type II-B, Type II-C, Type V-A, Type V-B, Type V-C, or Type V-U, OBJCRISPR-Cas system. OBJIn certain example embodiments, the Cas protein is Cas9, Cas12a, Cas12b, Cas12c, or Cas12d. In some embodiments, Cas9 may be SpCas9, SaCas9, StCas9 and other Cas9 orthologs. Cas 12 may be Cas12a, Cas12b, and Cas12c, including FnCas12a, or homology or orthologs thereof. The definition and exemplary members of the CRISPR-Cas system include those described in Kira S. Makarova and Eugene V. Koonin, Annotation and Classification of CRISPR-Cas systems, Methods Mol Biol. 2015; 1311: 47–75; and Sergey Shmakov et al., Diversity and evolution of class 2 CRISPR–Cas systems, Nat Rev Microbiol. 2017 Mar; 15(3): 169–182.

[0125] In some examples, the Cas protein comprises at least one RuvC and at least one HNH domain. In some examples, the Cas comprises at least one RuvC domain but does not comprise an HNH domain.

Class 2 Type II Cas

[0126] In some embodiments, the Cas protein may be a Cas protein of a Class 2, Type II CRISPR-Cas system (a Type II Cas protein). In some embodiments, the Cas protein may be a

class 2 Type II Cas protein, e.g., Cas9. By "Cas9 (CRISPR associated protein 9)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_269215 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). "Cas9 function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein. By "Cas 9 nucleic acid molecule" is meant a polynucleotide encoding a Cas9 polypeptide or fragment thereof. An exemplary Cas9 nucleic acid molecule sequence is provided at NCBI Accession No. NC_002737. In some embodiments, disclosed herein are inhibitors of Cas9, e.g., naturally occurring Cas9 in S. pyogenes (SpCas9) or S. aureus (SaCas9), or variants thereof. Cas9 recognizes foreign DNA using Protospacer Adjacent Motif (PAM) sequence and the base pairing of the target DNA by the guide RNA (gRNA). The relative ease of inducing targeted strand breaks at any genomic loci by Cas9 has enabled efficient genome editing in multiple cell types and organisms. Cas9 derivatives can also be used as transcriptional activators/repressors.

[0127] In some examples, the Cas9 may be in a mutated form. Examples of Cas9 mutations include D10A, E762A, H840A, N854A, N863A and D986A in respect of SpCas9. In one example, the Cas9 is Cas9D10A. In another example, the Cas9 is Cas9H840A.

Class 2 Type V Cas

[0128] In certain embodiments, the Cas protein may be a Cas protein of a Class 2, Type V CRISPR-Cas system (a Type V Cas protein). Examples of class 2 Type V Cas proteins include Cas12a (Cpf1), Cas12b (C2c1), Cas12c (C2c3), or Cas12k.

[0129] In some examples, the Cas protein is Cpf1. By "Cpfl (CRISPR associated protein Cpfl)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to GenBank Accession No. AJI61006. 1 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). "Cpfl function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein. By "Cpfl nucleic acid molecule" is meant a polynucleotide encoding a Cpfl polypeptide or fragment thereof. An exemplary Cpfl nucleic acid molecule sequence is provided at GenBank Accession No. CP009633, nucleotides 652838

- 656740. Cpf1(CRISPR-associated protein Cpf1, subtype PREFRAN) is a large protein (about 1300 amino acids) that contains a RuvC-like nuclease domain homologous to the corresponding domain of Cas9 along with a counterpart to the characteristic arginine-rich cluster of Cas9. However, Cpf1 lacks the HNH nuclease domain that is present in all Cas9 proteins, and the RuvC-like domain is contiguous in the Cpf1 sequence, in contrast to Cas9 where it contains long inserts including the HNH domain. Accordingly, in particular embodiments, the CRISPR-Cas enzyme comprises only a RuvC-like nuclease domain.

[0130] The Cpf1 gene is found in several diverse bacterial genomes, typically in the same locus with cas1, cas2, and cas4 genes and a CRISPR cassette (for example, FNFX1_1431-FNFX1_1428 of Francisella cf . novicida Fx1). Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the Cpf1 protein contains a readily identifiable C-terminal region that is homologous to the transposon ORF-B and includes an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). However, unlike Cas9, Cpf1 is also present in several genomes without a CRISPR-Cas context and its relatively high similarity with ORF-B suggests that it might be a transposon component. It was suggested that if this was a genuine CRISPR-Cas system and Cpf1 is a functional analog of Cas9 it would be a novel CRISPR-Cas type, namely type V (See Annotation and Classification of CRISPR-Cas Systems. Makarova KS, Koonin EV. Methods Mol Biol. 2015;1311:47-75). However, as described herein, Cpf1 is denoted to be in subtype V-A to distinguish it from C2c1p which does not have an identical domain structure and is hence denoted to be in subtype V-B.

[0131] In some examples, the Cas protein is Cc2c1. The C2c1 gene is found in several diverse bacterial genomes, typically in the same locus with cas1, cas2, and cas4 genes and a CRISPR cassette. Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the C2c1 protein contains an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). C2c1 (Cas12b) is derived from a C2c1 locus denoted as subtype V-B. Herein such effector proteins are also referred to as “C2c1p”, e.g., a C2c1 protein (and such effector protein or C2c1 protein or protein derived from a C2c1 locus is also called “CRISPR enzyme”). Presently, the subtype V-B loci encompasses cas1-Cas4 fusion, cas2, a distinct gene denoted C2c1 and a CRISPR array. C2c1 (CRISPR-associated protein C2c1) is a large protein (about 1100 - 1300 amino acids) that contains a RuvC-like nuclease domain homologous to the corresponding domain of Cas9 along with a counterpart to the characteristic arginine-rich cluster of Cas9. However, C2c1 lacks the HNH nuclease domain that is present in all Cas9 proteins, and the RuvC-like domain

is contiguous in the C2c1 sequence, in contrast to Cas9 where it contains long inserts including the HNH domain. Accordingly, in particular embodiments, the CRISPR-Cas enzyme comprises only a RuvC-like nuclease domain.

[0132] C2c1 proteins are RNA guided nucleases. Its cleavage relies on a tracr RNA to recruit a guide RNA comprising a guide sequence and a direct repeat, where the guide sequence hybridizes with the target nucleotide sequence to form a DNA/RNA heteroduplex. Based on current studies, C2c1 nuclease activity also requires relies on recognition of PAM sequence. C2c1 PAM sequences may be T-rich sequences. In some embodiments, the PAM sequence is 5' TTN 3' or 5' ATTN 3', wherein N is any nucleotide. In a particular embodiment, the PAM sequence is 5' TTC 3'. In a particular embodiment, the PAM is in the sequence of Plasmodium falciparum. C2c1 creates a staggered cut at the target locus, with a 5' overhang, or a "sticky end" at the PAM distal side of the target sequence. In some embodiments, the 5' overhang is 7 nt. See Lewis and Ke, Mol Cell. 2017 Feb 2;65(3):377-379.

Nickases

[0133] In some embodiments, the Cas protein or polypeptide may be a nickase. The Cas proteins with nickase activity may be a mutated form of a wildtype Cas protein. Mutations can also be made at neighboring residues at amino acids that participate in the nuclease activity. In some embodiments, only the RuvC domain is inactivated, and in other embodiments, another putative nuclease domain is inactivated, wherein the effector protein complex functions as a nickase and cleaves only one DNA strand. In some embodiments, two Cas variants (each a different nickase) are used to increase specificity, two nickase variants are used to cleave DNA at a target (where both nickases cleave a DNA strand, while minimizing or eliminating off-target modifications where only one DNA strand is cleaved and subsequently repaired). In preferred embodiments the Cas protein cleaves sequences associated with or at a target locus of interest as a homodimer comprising two Cas protein molecules. In a preferred embodiment the homodimer may comprise two Cas protein molecules comprising a different mutation in their respective RuvC domains.

[0134] The Cas protein may be mutated with respect to a corresponding wild-type enzyme such that the mutated Cas protein lacks the ability to cleave one or both DNA strands of a target locus containing a target sequence. In particular embodiments, one or more catalytic domains of the Cas protein are mutated to produce a mutated Cas protein which cleaves only one DNA strand of a target sequence.

[0135] In certain embodiments of the methods provided herein the Cas protein is a mutated Cas protein which cleaves only one DNA strand, i.e. a nickase. More particularly, in the context

of the present invention, the nickase ensures cleavage within the non-target sequence, i.e. the sequence which is on the opposite DNA strand of the target sequence and which is 3' of the PAM sequence. By means of further guidance, and without limitation, an arginine-to-alanine substitution (R911A) in the Nuc domain of C2c1 from Alicyclobacillus acidoterrestris converts C2c1 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). It will be understood by the skilled person that where the enzyme is not AacC2c1, a mutation may be made at a residue in a corresponding position.

[0136] In certain embodiments, the Cas protein may be a C2c1 nickase which comprises a mutation in the Nuc domain. In some embodiments, the C2c1 nickase comprises a mutation corresponding to amino acid positions R911, R1000, or R1015 in Alicyclobacillus acidoterrestris C2c1. In some embodiments, the C2c1 nickase comprises a mutation corresponding to R911A, R1000A, or R1015A in Alicyclobacillus acidoterrestris C2c1. In some embodiments, the C2c1 nickase comprises a mutation corresponding to R894A in Bacillus sp. V3-13 C2c1. In certain embodiments, the C2c1 protein recognizes PAMs with increased or decreased specificity as compared with an unmutated or unmodified form of the protein. In some embodiments, the C2c1 protein recognizes altered PAMs as compared with an unmutated or unmodified form of the protein.

[0137] In some embodiments, to minimize the level of toxicity and off-target effect, a Cas nickase can be used with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as described herein.

[0138] In some examples, the system may comprise two or more nickases, in particular a dual or double nickase approach. In some aspects and embodiments, a single type Cas nickase may be delivered, for example a modified Cas or a modified Cas nickase as described herein. This results in the target DNA being bound by two Cas nickases. In addition, it is also envisaged that different orthologs may be used, e.g., a Cas nickase on one strand (e.g., the coding strand) of the DNA and an ortholog on the non-coding or opposite DNA strand. The ortholog can be, but is not limited to, a Cas nickase. It may be advantageous to use two different orthologs that require different PAMs and may also have different guide requirements, thus allowing a greater deal of control for the user. In certain embodiments, DNA cleavage will involve at least four types of nickases, wherein each type is guided to a different sequence of target DNA, wherein each pair introduces a first nick into one DNA strand and the second introduces a nick into the second DNA strand. In such methods, at least two pairs of single stranded breaks are introduced into the target DNA wherein upon introduction of first and

second pairs of single-strand breaks, target sequences between the first and second pairs of single-strand breaks are excised. In certain embodiments, one or both of the orthologs is controllable, i.e. inducible.

Dead Cas

[0139] In certain embodiments, the Cas protein is a catalytically inactive or dead Cas protein (dCas). For example, the Cas protein or polypeptide may lack nuclease activity. In some embodiments, the dCas comprises mutations in the nuclease domain. In some embodiments, the dCas effector protein has been truncated. In some cases, the dead Cas proteins may be fused with one or more functional domains.

IscB and TnpB

[0140] In certain example embodiments, the RNA-guide protein may be an IscB or TnpB protein.

dCas - Functional Domain

[0141] The Cas protein or its variant (e.g., dCas) may be associated (e.g., fused) to one or more functional domains. The association can be by direct linkage of the Cas protein to the functional domain, or by association with the crRNA. In a non-limiting example, the crRNA comprises an added or inserted sequence that can be associated with a functional domain of interest, including, for example, an aptamer or a nucleotide that binds to a nucleic acid binding adapter protein. The functional domain may be a functional heterologous domain.

[0142] The functional domain may cleave a DNA sequence or modify transcription or translation of a gene. Examples of functional domains include domains that have methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g., light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that Fok1 is provided, multiple Fok1 functional domains may be provided to allow for a functional dimer and that gRNAs are designed to provide proper spacing for functional use (Fok1).

[0143] In some cases, the functional domains may be heterologous functional domains. For example, the one or more heterologous functional domains may comprise one or more nuclear localization signal (NLS) domains. The one or more heterologous functional domains may comprise at least two or more NLS domains. The one or more NLS domain(s) may be positioned at or near or in proximity to a terminus of the Cas protein and if two or more NLSs, each of the two may be positioned at or near or in proximity to a terminus of the Cas protein.

The one or more heterologous functional domains may comprise one or more transcriptional activation domains. In a preferred embodiment the transcriptional activation domain may comprise VP64. The one or more heterologous functional domains may comprise one or more transcriptional repression domains. In a preferred embodiment the transcriptional repression domain comprises a KRAB domain or a SID domain (e.g. SID4X). The one or more heterologous functional domains may comprise one or more nuclease domains. In a preferred embodiment a nuclease domain comprises Fok1. Other examples of functional domains include translational initiator, translational activator, translational repressor, nucleases, in particular ribonucleases, a spliceosome, beads, a light inducible/controllable domain or a chemically inducible/controllable domain.

[0144] The positioning of the one or more functional domain on Cas or dCas protein is one which allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, if the functional domain is a transcription activator (e.g., VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor may be positioned to affect the transcription of the target, and a nuclease (e.g., Fok1) will be advantageously positioned to cleave or partially cleave the target. This may include positions other than the N- / C- terminus of the Cas protein.

[0145] The Cas or dCas protein may be associated with the one or more functional domains through one or more adaptor proteins. The adaptor protein may utilize known linkers to attach such functional domains.

[0146] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 $((GGGGS)_3$ (SEQ ID NO: 69)) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the guide RNAs and the functional domain (activator or repressor), or between the nucleic acid-targeting effector protein and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

The term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker. Suitable linkers for use in the methods of the present invention

are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the Cas protein and the nucleotide deaminase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS (SEQ ID NO: 70) or GSG can be used. GGS, GSG, GGGS or GGGGS linkers can be used in repeats of 3 (such as $(GGS)_3$ (SEQ ID NO: 71), $(GGGGS)_3$ (SEQ ID NO:69)) or 5, 6, 7, 9 or even 12 or more, to provide suitable lengths. In some cases, the linker may be (GGGGS)3-15, For example, in some cases, the linker may be (GGGGS)3-11, e.g., GGGGS (SEQ ID NO: 72), $(GGGGS)_2$ (SEQ ID NO: 73), $(GGGGS)_3$ (SEQ ID NO: 69), $(GGGGS)_4$ (SEQ ID NO: 74), $(GGGGS)_5$ (SEQ ID NO: 75), $(GGGGS)_6$ (SEQ ID NO: 76), $(GGGGS)_7$ (SEQ ID NO: 77), $(GGGGS)_8$ (SEQ ID NO: 78), $(GGGGS)_9$ (SEQ ID NO: 79), $(GGGGS)_{10}$ (SEQ ID NO: 80), or $(GGGGS)_{11}$ (SEQ ID NO: 81). In particular embodiments, linkers such as $(GGGGS)_3$ (SEQ ID NO: 69) are preferably used herein. $(GGGGS)_6$ (SEQ ID NO: 76, $(GGGGS)_9$ (SEQ ID NO: 79) or $(GGGGS)_{12}$ (SEQ ID NO: 82) may preferably be used as alternatives. Other preferred alternatives are $(GGGGS)_1$ (SEQ ID NO: 72), $(GGGGS)_2$ (SEQ ID NO: 73), $(GGGGS)_4$ (SEQ ID NO:74), $(GGGGS)_5$ (SEQ ID NO: 75), $(GGGGS)_7$ (SEQ ID NO: 77), $(GGGGS)_8$ (SEQ ID NO: 78), $(GGGGS)_{10}$ (SEQ ID NO: 80), or $(GGGGS)_{11}$ (SEQ ID NO: 81). In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 83) is used as a linker. In yet an additional embodiment, the linker is an XTEN linker. In particular embodiments, the CRISPR-cas protein is a CRISPR-Cas protein and is linked to the deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 83) linker. In further particular embodiments, the CRISPR-Cas protein is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an

LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 83) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID NO: 84)).

[0147] The skilled person will understand that modifications to the guide which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

Guide Molecules

[0148] The system herein may comprise one or more guide molecules. The guide molecule(s) may be component(s) of the CRISPR-Cas system herein. As used herein, the term "guide sequence" and "guide molecule" in the context of a CRISPR-Cas system, comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. The guide sequences made using the methods disclosed herein may be a full-length guide sequence, a truncated guide sequence, a full-length sgRNA sequence, a truncated sgRNA sequence, or an E+F sgRNA sequence. In some embodiments, the degree of complementarity of the guide sequence to a given target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In certain example embodiments, the guide molecule comprises a guide sequence that may be designed to have at least one mismatch with the target sequence, such that a RNA duplex formed between the guide sequence and the target sequence. Accordingly, the degree of complementarity is preferably less than 99%. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less. In particular embodiments, the guide sequence is designed to have a stretch of two or more adjacent mismatching nucleotides, such that the degree of complementarity over the entire guide sequence is further reduced. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less, more particularly, about 92% or less, more particularly about 88% or less, more particularly about 84% or less, more particularly about 80% or less, more particularly about 76% or less, more particularly about 72% or less, depending on whether the stretch of two or more mismatching

nucleotides encompasses 2, 3, 4, 5, 6 or 7 nucleotides, etc. In some embodiments, aside from the stretch of one or more mismatching nucleotides, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence.

[0149] The guide molecule may direct the fusion proteins of the present invention to a target sequence that is 5' to or 3' the targeted insertion site. In the case of paired nickase embodiments, one guide molecule be configured to bind to a target sequence on the sense strand of the target polypeptide and a second guide may be configured to bind to the anti-sense strand of the target polynucleotide.

[0150] In certain embodiments, the guide sequence or spacer length of the guide molecules is from 15 to 50 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt,

from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer. In certain example embodiment, the guide sequence is 15, 16, 17,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 40, 41, 42, 43, 44, 45, 46, 47 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 nt.

[0151] In some embodiments, the guide sequence is an RNA sequence of between 10 to 50 nt in length, but more particularly of about 20-30 nt advantageously about 20 nt, 23-25 nt or 24 nt. The guide sequence is selected so as to ensure that it hybridizes to the target sequence. This is described more in detail below. Selection can encompass further steps which increase efficacy and specificity.

[0152] In some embodiments, the guide sequence has a canonical length (e.g., about 15-30 nt) is used to hybridize with the target RNA or DNA. In some embodiments, a guide molecule is longer than the canonical length (e.g., >30 nt) is used to hybridize with the target RNA or DNA, such that a region of the guide sequence hybridizes with a region of the RNA or DNA strand outside of the Cas-guide target complex. This can be of interest where additional modifications, such deamination of nucleotides is of interest. In alternative embodiments, it is of interest to maintain the limitation of the canonical guide sequence length.

[0153] In some embodiments, the sequence of the guide molecule (direct repeat and/or spacer) is selected to reduce the degree secondary structure within the guide molecule. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0154] In some embodiments, a guide molecule is designed or selected to modulate intermolecular interactions among guide molecules, such as among stem-loop regions of different guide molecules. It will be appreciated that nucleotides within a guide that base-pair

to form a stem-loop are also capable of base-pairing to form an intermolecular duplex with a second guide and that such an intermolecular duplex would not have a secondary structure compatible with CRISPR complex formation. Accordingly, it is useful to select or design DR sequences in order to modulate stem-loop formation and CRISPR complex formation. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of nucleic acid-targeting guides are in intermolecular duplexes. It will be appreciated that stem-loop variation will often be within limits imposed by DR-CRISPR effector interactions. One way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to vary nucleotide pairs in the stem of the stem-loop of a DR. For example, in one embodiment, a G-C pair is replaced by an A-U or U-A pair. In another embodiment, an A-U pair is substituted for a G-C or a C-G pair. In another embodiment, a naturally occurring nucleotide is replaced by a nucleotide analog. Another way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to modify the loop of the stem-loop of a DR. Without be bound by theory, the loop can be viewed as an intervening sequence flanked by two sequences that are complementary to each other. When that intervening sequence is not self-complementary, its effect will be to destabilize intermolecular duplex formation. The same principle applies when guides are multiplexed: while the targeting sequences may differ, it may be advantageous to modify the stem-loop region in the DRs of the different guides. Moreover, when guides are multiplexed, the relative activities of the different guides can be modulated by balancing the activity of each individual guide. In certain embodiments, the equilibrium between intermolecular stem-loops vs. intermolecular duplexes is determined. The determination may be made by physical or biochemical means and can be in the presence or absence of a CRISPR effector.

[0155] In some embodiments, it is of interest to reduce the susceptibility of the guide molecule to RNA cleavage, such as cleavage by a CRISPR system that cleaves RNA. Accordingly, in particular embodiments, the guide molecule is adjusted to avoid cleavage by a CRISPR system or other RNA-cleaving enzymes.

[0156] In certain embodiments, the guide molecule comprises non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Preferably, these non-naturally occurring nucleic acids and non-naturally occurring nucleotides are located outside the guide sequence. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be

modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2′ and 4′ carbons of the ribose ring, or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2′-O-methyl (M), 2′-O-methyl 3′phosphorothioate (MS), *S*-constrained ethyl(cEt), or 2′-O-methyl 3′thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guides can comprise increased stability and increased activity as compared to unmodified guides, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015 Ragdarm et al., 0215, *PNAS*, E7110-E7111; Allerson et al., *J. Med. Chem*. 2005, 48:901-904; Bramsen et al., *Front. Genet*., 2012, 3:154; Deng et al., *PNAS*, 2015, 112:11870-11875; Sharma et al., *MedChemComm*., 2014, 5:1454-1471; Hendel et al., *Nat. Biotechnol*. (2015) 33(9): 985-989; Li et al., *Nature Biomedical Engineering*, 2017, 1, 0066 DOI:10.1038/s41551-017-0066). In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, *J. Biotech.* 233:74-83). In certain embodiments, a guide comprises ribonucleotides in a region that binds to a target RNA and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to a Cas effector. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, stem-loop regions, and the seed region. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl 3' phosphorothioate

(MS), *S*-constrained ethyl(cEt), or 2'-O-methyl 3' thioPACE (MSP). Such modification can enhance genome editing efficiency (see Hendel et al., *Nat. Biotechnol.* (2015) 33(9): 985-989). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or *S*-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, *PNAS*, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine, peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., *eLife*, 2017, 6:e25312, DOI:10.7554).

[0157] In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve *in vivo* editing and stability (see Finn et al., *Cell Reports* (2018), 22: 2227-2235). In some embodiments, more than 60 or 70 nucleotides of the guide are chemically modified. In some embodiments, this modification comprises replacement of nucleotides with 2'-O-methyl or 2'-fluoro nucleotide analogs or phosphorothioate (PS) modification of phosphodiester bonds. In some embodiments, the chemical modification comprises 2'-O-methyl or 2'-fluoro modification of guide nucleotides extending outside of the nuclease protein when the CRISPR complex is formed or PS modification of 20 to 30 or more nucleotides of the 3'-terminus of the guide. In a particular embodiment, the chemical modification further comprises 2'-O-methyl analogs at the 5' end of the guide or 2'-fluoro analogs in the seed and tail regions. Such chemical modifications improve stability to nuclease degradation and maintain or enhance genome-editing activity or efficiency, but modification of all nucleotides may abolish the function of the guide (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). Such chemical modifications may be guided by knowledge of the structure of the CRISPR

complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). In some embodiments, one or more guide RNA nucleotides may be replaced with DNA nucleotides. In some embodiments, up to 2, 4, 6, 8, 10, or 12 RNA nucleotides of the 5'-end tail/seed guide region are replaced with DNA nucleotides. In certain embodiments, the majority of guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 16 guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 8 guide RNA nucleotides of the 5'-end tail/seed region and 16 RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, guide RNA nucleotides that extend outside of the nuclease protein when the CRISPR complex is formed are replaced with DNA nucleotides. Such replacement of multiple RNA nucleotides with DNA nucleotides leads to decreased off-target activity but similar on-target activity compared to an unmodified guide; however, replacement of all RNA nucleotides at the 3' end may abolish the function of the guide (see Yin et al., *Nat. Chem. Biol.* (2018) 14, 311-316). Such modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., *Nat. Chem. Biol.* (2018) 14, 311-316).

[0158] In some embodiments, the guide molecule forms a stemloop with a separate non-covalently linked sequence, which can be DNA or RNA. In particular embodiments, the sequences forming the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In some embodiments, these sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sulfonyl, ally, propargyl, diene, alkyne, and azide. Once this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the direct repeat sequence. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, sulfones, sulfoxides, ureas, thioureas,

hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0159] In some embodiments, these stem-loop forming sequences can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0160] In certain embodiments, the guide molecule comprises (1) a guide sequence capable of hybridizing to a target locus and (2) a tracr mate or direct repeat sequence whereby the direct repeat sequence is located upstream (i.e., 5') or downstream (i.e. 3') from the guide sequence. In a particular embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the guide sequence is approximately within the first 10 nucleotides of the guide sequence.

[0161] In a particular embodiment, the guide molecule comprises a guide sequence linked to a direct repeat sequence, wherein the direct repeat sequence comprises one or more stem loops or optimized secondary structures. In particular embodiments, the direct repeat has a minimum length of 16 nts and a single stem loop. In further embodiments the direct repeat has a length longer than 16 nts, preferably more than 17 nts, and has more than one stem loops or optimized secondary structures. In particular embodiments the guide molecule comprises or consists of the guide sequence linked to all or part of the natural direct repeat sequence. A CRISPR-cas guide molecule comprises (in 3' to 5' direction or in 5' to 3' direction): a guide sequence a first complimentary stretch (the "repeat"), a loop (which is typically 4 or 5 nucleotides long), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), and a poly A (often poly U in RNA) tail (terminator). In certain embodiments, the direct repeat sequence retains its natural architecture and forms a single stem loop. In particular embodiments, certain aspects of the guide architecture can be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered guide molecule modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the guide molecule that are exposed when complexed with the CRISPR-Cas protein and/or target, for example the stemloop of the direct repeat sequence.

[0162] In particular embodiments, the stem comprises at least about 4bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or

fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-10 and Y2-10 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the loop will form a complete hairpin in the overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y basepairing sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire guide molecule is preserved. In one aspect, the loop that connects the stem made of X:Y basepairs can be any sequence of the same length (e.g., 4 or 5 nucleotides) or longer that does not interrupt the overall secondary structure of the guide molecule. In one aspect, the stemloop can further comprise, e.g. an MS2 aptamer. In one aspect, the stem comprises about 5-7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stemloop at that position.

[0163] In particular embodiments the natural hairpin or stemloop structure of the guide molecule is extended or replaced by an extended stemloop. It has been demonstrated that extension of the stem can enhance the assembly of the guide molecule with the CRISPR-Cas protein (Chen et al. Cell. (2013); 155(7): 1479–1491). In particular embodiments the stem of the stemloop is extended by at least 1, 2, 3, 4, 5 or more complementary basepairs (i.e. corresponding to the addition of 2,4, 6, 8, 10 or more nucleotides in the guide molecule). In particular embodiments these are located at the end of the stem, adjacent to the loop of the stemloop.

[0164] In particular embodiments, the susceptibility of the guide molecule to RNases or to decreased expression can be reduced by slight modifications of the sequence of the guide molecule which do not affect its function. For instance, in particular embodiments, premature termination of transcription, such as premature transcription of U6 Pol-III, can be removed by modifying a putative Pol-III terminator (4 consecutive U's) in the guide molecules sequence. Where such sequence modification is required in the stemloop of the guide molecule, it is preferably ensured by a basepair flip.

[0165] In a particular embodiment, the direct repeat may be modified to comprise one or more protein-binding RNA aptamers. In a particular embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein as detailed further herein.

[0166] In some embodiments, the guide molecule forms a duplex with a target RNA comprising at least one target cytosine residue to be edited. Upon hybridization of the guide RNA molecule to the target RNA, the cytidine deaminase binds to the single strand RNA in the duplex made accessible by the mismatch in the guide sequence and catalyzes deamination of one or more target cytosine residues comprised within the stretch of mismatching nucleotides.

[0167] A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be mRNA.

[0168] In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site); that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM.

[0169] Further, engineering of the PAM Interacting (PI) domain may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592.

[0170] In particular embodiments, the guide is an escorted guide. By "escorted" is meant that the CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the 3 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[0171] The escorted CRISPR-Cas systems or complexes have a guide molecule with a functional structure designed to improve guide molecule structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0172] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential

enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a que by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[0173] Accordingly, in particular embodiments, the guide molecule is modified, e.g., by one or more aptamer(s) designed to improve guide molecule delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide molecule deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends a guide molecule that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0174] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents.

Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0175] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0176] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0177] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0178] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0179] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel based system inducible by energy, heat or radio-wave (see, e.g., www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the CRISPR-Cas complex will be active and modulating target gene expression in cells.

[0180] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for *in vivo* applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0181] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under *in vivo* conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μs and 500 milliseconds, preferably between 1 μs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0182] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under *in vivo* conditions (see WO97/49450).

[0183] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc., as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0184] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The

ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0185] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0186] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0187] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under in vivo conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0188] Preferably, the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0189] Preferably, the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0190] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0191] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm2 to about 100 W/cm2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0192] As used herein, the term “ultrasound” refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0193] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm2 (FDA recommendation), although energy densities of up to 750 mW/cm2 have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[0194] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0195] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[0196] Preferably the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm-2. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm-2.

[0197] Preferably the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0198] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0199] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm-2 to about 10 Wcm-2 with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm-2, but for reduced periods of time, for example, 1000 Wcm-2 for periods in the millisecond range or less.

[0200] Preferably the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0201] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm-2 or 1.25 Wcm-2 as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0202] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[0203] In particular embodiments, the guide molecule is modified by a secondary structure to increase the specificity of the CRISPR-Cas system and the secondary structure can protect against exonuclease activity and allow for 5' additions to the guide sequence also referred to herein as a protected guide molecule.

[0204] In one aspect, the invention provides for hybridizing a "protector RNA" to a sequence of the guide molecule, wherein the "protector RNA" is an RNA strand complementary to the 3' end of the guide molecule to thereby generate a partially double-stranded guide RNA. In an embodiment of the invention, protecting mismatched bases (i.e. the bases of the guide molecule which do not form part of the guide sequence) with a perfectly complementary protector sequence decreases the likelihood of target RNA binding to the mismatched basepairs at the 3' end. In particular embodiments of the invention, additional sequences comprising an extended length may also be present within the guide molecule such that the guide comprises a protector sequence within the guide molecule. This "protector sequence" ensures that the guide molecule comprises a "protected sequence" in addition to an "exposed sequence" (comprising the part of the guide sequence hybridizing to the target sequence). In particular embodiments, the guide molecule is modified by the presence of the protector guide to comprise a secondary structure such as a hairpin. Advantageously there are three or four to thirty or more, e.g., about 10 or more, contiguous base pairs having complementarity to the protected sequence, the guide sequence or both. It is advantageous that the protected portion does not impede thermodynamics of the CRISPR-Cas system interacting with its target. By providing such an extension including a partially double stranded guide molecule, the guide molecule is considered protected and results in improved specific binding of the CRISPR-Cas complex, while maintaining specific activity.

[0205] In particular embodiments, use is made of a truncated guide (tru-guide), i.e. a guide molecule which comprises a guide sequence which is truncated in length with respect to the canonical guide sequence length. As described by Nowak et al. (Nucleic Acids Res (2016) 44 (20): 9555-9564), such guides may allow catalytically active CRISPR-Cas enzyme to bind its target without cleaving the target RNA. In particular embodiments, a truncated guide is used which allows the binding of the target but retains only nickase activity of the CRISPR-Cas enzyme.

[0206] The methods and tools provided herein are exemplified for certain Cas effectors. Further nucleases with similar properties can be identified using methods described in the art (Shmakov et al. 2015, 60:385–397; Abudayeh et al. 2016, Science, 5;353(6299)) . In particular embodiments, such methods for identifying novel CRISPR effector proteins may comprise the

steps of selecting sequences from the database encoding a seed which identifies the presence of a CRISPR Cas locus, identifying loci located within 10 kb of the seed comprising Open Reading Frames (ORFs) in the selected sequences, selecting therefrom loci comprising ORFs of which only a single ORF encodes a novel CRISPR effector having greater than 700 amino acids and no more than 90% homology to a known CRISPR effector. In particular embodiments, the seed is a protein that is common to the CRISPR-Cas system, such as Cas1. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[0207] Also, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[0208] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, AAV, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: US Patents Nos. 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, 8,945,839, 8,993,233 and 8,999,641; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App. Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO

2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809), WO 2015/089351 (PCT/US2014/069897), WO 2015/089354 (PCT/US2014/069902), WO 2015/089364 (PCT/US2014/069925), WO 2015/089427 (PCT/US2014/070068), WO 2015/089462 (PCT/US2014/070127), WO 2015/089419 (PCT/US2014/070057), WO 2015/089465 (PCT/US2014/070135), WO 2015/089486 (PCT/US2014/070175), PCT/US2015/051691, PCT/US2015/051830.

[0209] [0210] RReference is also made to US Provisional Application Nos. 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US Provisional Application No. 61/836,123, filed on June 17, 2013. Reference is additionally made to US Provisional Application Nos. 61/835,931, 61/835,936, 61/836,127, 61/836, 101, 61/836,080 and 61/835,973, each filed June 17, 2013. Further reference is made to US Provisional Application Nos. 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961,980 filed on October 28, 2013. Reference is yet further made to PCT Patent Application Nos: PCT/US2014/041803, PCT/US2014/041800, PCT/US2014/041809, PCT/US2014/041804 and PCT/US2014/041806, each filed June 10, 2014 6/10/14; PCT/US2014/041808 filed June 11, 2014; and PCT/US2014/62558 filed October 28, 2014, and US Provisional Applications Nos.: 61/915,150, 61/915,301, 61/915,267 and 61/915,260, each filed December 12, 2013; 61/757,972 and 61/768,959, filed on January 29, 2013 and February 25, 2013; 61/835,936, 61/836,127, 61/836,101, 61/836,080, 61/835,973, and 61/835,931, filed June 17, 2013; 62/010,888 and 62/010,879, both filed June 11, 2014; 62/010,329 and 62/010,441, each filed June 10, 2014; 61/939,228 and 61/939,242, each filed February 12, 2014; 61/980,012, filed April 15,2014; 62/038,358, filed August 17, 2014; 62/054,490, 62/055,484, 62/055,460 and 62/055,487, each filed September 25, 2014; and 62/069,243, filed October 27, 2014. Reference is also made to US Provisional Application Nos. 62/055,484, 62/055,460, and 62/055,487, filed September 25, 2014; US Provisional Application No. 61/980,012, filed April 15, 2014; and US provisional patent application

61/939,242 filed February 12, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US Provisional Application No. 61/930,214 filed on January 22, 2014. Reference is made to US Provisional Application Nos. 61/915,251; 61/915,260 and 61/915,267, each filed on December 12, 2013. Reference is made to US Provisional Application No. 61/980,012 filed April 15, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014.

[0210] Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-F EB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-

14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0211] Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 19-Aug-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, and

62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US Application 14/975,085, 18-Dec-2015, European application No. 16150428.7, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-Jul-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[0212] In addition, mention is made of PCT application PCT/US14/70057, Attorney Reference 47627.99.2060 and BI-2013/107 entitled "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS (claiming priority from one or more or all of US provisional patent applications: 62/054,490, filed September 24, 2014; 62/010,441, filed June 10, 2014; and 61/915,118, 61/915,215 and 61/915,148, each filed on December 12, 2013) ("the Particle Delivery PCT"), incorporated herein by reference, and of PCT application PCT/US14/70127, Attorney Reference 47627.99.2091 and BI-2013/101 entitled "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING " (claiming priority from one or more or all of US provisional patent applications: 61/915,176; 61/915,192; 61/915,215; 61/915,107, 61/915,145; 61/915,148; and 61/915,153 each filed December 12, 2013) ("the Eye PCT"), incorporated herein by reference, with respect to a method of preparing an sgRNA-and-Cas protein containing particle comprising admixing a mixture comprising an sgRNA and Cas effector protein (and optionally HDR template) with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol; and particles from such a process. For example, wherein the Cas protein and sgRNA were mixed together at a suitable, e.g., 3:1 to 1:3 or 2:1 to 1:2 or 1:1 molar ratio, at a suitable temperature, e.g., 15-30C, e.g., 20-25C, e.g., room temperature, for a suitable time, e.g., 15-45, such as 30 minutes, advantageously in sterile, nuclease free buffer, e.g., 1X PBS. Separately, particle components such as or comprising: a surfactant, e.g., cationic lipid, e.g.,

1,2-dioleoyl-3-trimethylammonium-propane (DOTAP); phospholipid, e.g., dimyristoylphosphatidylcholine (DMPC); biodegradable polymer, such as an ethylene-glycol polymer or PEG, and a lipoprotein, such as a low-density lipoprotein, e.g., cholesterol were dissolved in an alcohol, advantageously a C1-6 alkyl alcohol, such as methanol, ethanol, isopropanol, e.g., 100% ethanol. The two solutions were mixed together to form particles containing the Cas9-sgRNA complexes. Accordingly, sgRNA may be pre-complexed with the Cas protein, before formulating the entire complex in a particle. Formulations may be made with a different molar ratio of different components known to promote delivery of nucleic acids into cells (e.g. 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC), polyethylene glycol (PEG), and cholesterol) For example DOTAP : DMPC : PEG : Cholesterol Molar Ratios may be DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 5, Cholesterol 5. DOTAP 100, DMPC 0, PEG 0, Cholesterol 0. Other example nucleotide-binding systems and proteins.

Other exemplary site-specific nuclease or nucleic acid binding enzymes

[0213] In certain example embodiments, the retrotransposons may be used with other nucleotide-binding molecule that are not CRISPR-Cas system. Examples of the other nucleotide-binding molecules may be components of transcription activator-like effector nuclease (TALEN), Zn finger nucleases, meganucleases, a functional fragment thereof, a variant thereof, of any combination thereof.

TALE Systems

[0214] In some embodiments, the nucleotide-binding molecule in the systems may be a transcription activator-like effector nuclease, a functional fragment thereof, or a variant thereof. The present disclosure also includes nucleotide sequences that are or encode one or more components of a TALE system. As disclosed herein editing can be made by way of the transcription activator-like effector nucleases (TALENs) system. Transcription activator-like effectors (TALEs) can be engineered to bind practically any desired DNA sequence. Exemplary methods of genome editing using the TALEN system can be found for example in Cermak T. Doyle EL. Christian M. Wang L. Zhang Y. Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39:e82; Zhang F. Cong L. Lodato S. Kosuri S. Church GM. Arlotta P Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol. 2011;29:149–153 and US Patent Nos. 8,450,471, 8,440,431 and 8,440,432, all of which are specifically incorporated by reference.

[0215] In some embodiments, provided herein include isolated, non-naturally occurring, recombinant or engineered DNA binding proteins that comprise TALE monomers as a part of their organizational structure that enable the targeting of nucleic acid sequences with improved efficiency and expanded specificity.

[0216] Naturally occurring TALEs or "wild type TALEs" are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term "polypeptide monomers", or "TALE monomers" will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term "repeat variable di-residues" or "RVD" will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is X1-11-(X12X13)-X14-33 or 34 or 35, where the subscript indicates the amino acid position and X represents any amino acid. X12X13 indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such polypeptide monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as X*, where X represents X12 and (*) indicates that X13 is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as (X1-11-(X12X13)-X14-33 or 34 or 35)z, where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26.

[0217] The TALE monomers have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI preferentially bind to adenine (A), polypeptide monomers with an RVD of NG preferentially bind to thymine (T), polypeptide monomers with an RVD of HD preferentially bind to cytosine (C) and polypeptide monomers with an RVD of NN preferentially bind to both adenine (A) and guanine (G). In yet another embodiment of the invention, polypeptide monomers with an RVD of IG preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In still further embodiments of the invention, polypeptide monomers with an RVD of NS recognize all four base pairs and may bind to A, T, G or C. The

structure and function of TALEs is further described in, for example, Moscou et al., Science 326:1501 (2009); Boch et al., Science 326:1509-1512 (2009); and Zhang et al., Nature Biotechnology 29:149-153 (2011), each of which is incorporated by reference in its entirety.

[0218] The TALE polypeptides used in methods of the invention are isolated, non-naturally occurring, recombinant or engineered nucleic acid-binding proteins that have nucleic acid or DNA binding regions containing polypeptide monomer repeats that are designed to target specific nucleic acid sequences.

[0219] As described herein, polypeptide monomers having an RVD of HN or NH preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In a preferred embodiment of the invention, polypeptide monomers having RVDs RN, NN, NK, SN, NH, KN, HN, NQ, HH, RG, KH, RH and SS preferentially bind to guanine. In a much more advantageous embodiment of the invention, polypeptide monomers having RVDs RN, NK, NQ, HH, KH, RH, SS and SN preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In an even more advantageous embodiment of the invention, polypeptide monomers having RVDs HH, KH, NH, NK, NQ, RH, RN and SS preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In a further advantageous embodiment, the RVDs that have high binding specificity for guanine are RN, NH RH and KH. Furthermore, polypeptide monomers having an RVD of NV preferentially bind to adenine and guanine. In more preferred embodiments of the invention, polypeptide monomers having RVDs of H*, HA, KA, N*, NA, NC, NS, RA, and S* bind to adenine, guanine, cytosine and thymine with comparable affinity.

[0220] The predetermined N-terminal to C-terminal order of the one or more polypeptide monomers of the nucleic acid or DNA binding domain determines the corresponding predetermined target nucleic acid sequence to which the TALE polypeptides will bind. As used herein the polypeptide monomers and at least one or more half polypeptide monomers are "specifically ordered to target" the genomic locus or gene of interest. In plant genomes, the natural TALE-binding sites always begin with a thymine (T), which may be specified by a cryptic signal within the non-repetitive N-terminus of the TALE polypeptide; in some cases this region may be referred to as repeat 0. In animal genomes, TALE binding sites do not necessarily have to begin with a thymine (T) and TALE polypeptides may target DNA sequences that begin with T, A, G or C. The tandem repeat of TALE monomers always ends

with a half-length repeat or a stretch of sequence that may share identity with only the first 20 amino acids of a repetitive full length TALE monomer and this half repeat may be referred to as a half-monomer (FIG. 8), which is included in the term "TALE monomer". Therefore, it follows that the length of the nucleic acid or DNA being targeted is equal to the number of full polypeptide monomers plus two.

[0221] As described in Zhang et al., Nature Biotechnology 29:149-153 (2011), TALE polypeptide binding efficiency may be increased by including amino acid sequences from the "capping regions" that are directly N-terminal or C-terminal of the DNA binding region of naturally occurring TALEs into the engineered TALEs at positions N-terminal or C-terminal of the engineered TALE DNA binding region. Thus, in certain embodiments, the TALE polypeptides described herein further comprise an N-terminal capping region and/or a C-terminal capping region.

[0222] An exemplary amino acid sequence of a N-terminal capping region is:

M D P I R S R T P S P A R E L L S G P Q P D G V Q P T A D R G V S P
P A G G P L D G L P A R R T M S R T R L P S P P A P S P A F S A D S
F S D L L R Q F D P S L F N T S L F D S L P P F G A H H T E A A T G
E W D E V Q S G L R A A D A P P P T M R V A V T A A R P P R A K P A
P R R R A A Q P S D A S P A A Q V D L R T L G Y S Q Q Q Q E K I K P
K V R S T V A Q H H E A L V G H G F T H A H I V A L S Q H P A A L G
T V A V K Y Q D M I A A L P E A T H E A I V G V G K Q W S G A R A L
E A L L T V A G E L R G P P L Q L D T G Q L L K I A K R G G V T A V
E A V H A W R N A L T G A P L N (SEQ ID NO: 85)

[0223] An exemplary amino acid sequence of a C-terminal capping region is:

R P A L E S I V A Q L S R P D P A L A A L T N D H L V A L A C L G
G R P A L D A V K K G L P H A P A L I K R T N R R I P E R T S H R
V A D H A Q V V R V L G F F Q C H S H P A Q A F D D A M T Q F G M
S R H G L L Q L F R R V G V T E L E A R S G T L P P A S Q R W D R
I L Q A S G M K R A K P S P T S T Q T P D Q A S L H A F A D S L E
R D L D A P S P M H E G D Q T R A S (SEQ ID NO: 86)

[0224] As used herein the predetermined "N-terminus" to "C terminus" orientation of the N-terminal capping region, the DNA binding domain comprising the repeat TALE monomers and the C-terminal capping region provide structural basis for the organization of different domains in the d-TALEs or polypeptides of the invention.

[0225] The entire N-terminal and/or C-terminal capping regions are not necessary to enhance the binding activity of the DNA binding region. Therefore, in certain embodiments, fragments of the N-terminal and/or C-terminal capping regions are included in the TALE polypeptides described herein.

[0226] In certain embodiments, the TALE polypeptides described herein contain a N-terminal capping region fragment that included at least 10, 20, 30, 40, 50, 54, 60, 70, 80, 87, 90, 94, 100, 102, 110, 117, 120, 130, 140, 147, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260 or 270 amino acids of an N-terminal capping region. In certain embodiments, the N-terminal capping region fragment amino acids are of the C-terminus (the DNA-binding region proximal end) of an N-terminal capping region. As described in Zhang et al., Nature Biotechnology 29:149-153 (2011), N-terminal capping region fragments that include the C-terminal 240 amino acids enhance binding activity equal to the full length capping region, while fragments that include the C-terminal 147 amino acids retain greater than 80% of the efficacy of the full length capping region, and fragments that include the C-terminal 117 amino acids retain greater than 50% of the activity of the full-length capping region.

[0227] In some embodiments, the TALE polypeptides described herein contain a C-terminal capping region fragment that included at least 6, 10, 20, 30, 37, 40, 50, 60, 68, 70, 80, 90, 100, 110, 120, 127, 130, 140, 150, 155, 160, 170, 180 amino acids of a C-terminal capping region. In certain embodiments, the C-terminal capping region fragment amino acids are of the N-terminus (the DNA-binding region proximal end) of a C-terminal capping region. As described in Zhang et al., Nature Biotechnology 29:149-153 (2011), C-terminal capping region fragments that include the C-terminal 68 amino acids enhance binding activity equal to the full length capping region, while fragments that include the C-terminal 20 amino acids retain greater than 50% of the efficacy of the full length capping region.

[0228] In certain embodiments, the capping regions of the TALE polypeptides described herein do not need to have identical sequences to the capping region sequences provided herein. Thus, in some embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical or share identity to the capping region amino acid sequences provided herein. Sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences. In some preferred

embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 95% identical or share identity to the capping region amino acid sequences provided herein.

[0229] Sequence homologies may be generated by any of a number of computer programs known in the art, which include but are not limited to BLAST or FASTA. Suitable computer program for carrying out alignments like the GCG Wisconsin Bestfit package may also be used. Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

[0230] In some embodiments described herein, the TALE polypeptides of the invention include a nucleic acid binding domain linked to the one or more effector domains. The terms "effector domain" or "regulatory and functional domain" refer to a polypeptide sequence that has an activity other than binding to the nucleic acid sequence recognized by the nucleic acid binding domain. By combining a nucleic acid binding domain with one or more effector domains, the polypeptides of the invention may be used to target the one or more functions or activities mediated by the effector domain to a particular target DNA sequence to which the nucleic acid binding domain specifically binds.

[0231] In some embodiments of the TALE polypeptides described herein, the activity mediated by the effector domain is a biological activity. For example, in some embodiments the effector domain is a transcriptional inhibitor (i.e., a repressor domain), such as an mSin interaction domain (SID). SID4X domain or a Krüppel-associated box (KRAB) or fragments of the KRAB domain. In some embodiments the effector domain is an enhancer of transcription (i.e. an activation domain), such as the VP16, VP64 or p65 activation domain. In some embodiments, the nucleic acid binding is linked, for example, with an effector domain that includes but is not limited to a transposase, integrase, recombinase, resolvase, invertase, protease, DNA methyltransferase, DNA demethylase, histone acetylase, histone deacetylase, nuclease, transcriptional repressor, transcriptional activator, transcription factor recruiting, protein nuclear-localization signal or cellular uptake signal.

[0232] In some embodiments, the effector domain is a protein domain which exhibits activities which include but are not limited to transposase activity, integrase activity, recombinase activity, resolvase activity, invertase activity, protease activity, DNA methyltransferase activity, DNA demethylase activity, histone acetylase activity, histone deacetylase activity, nuclease activity, nuclear-localization signaling activity, transcriptional repressor activity, transcriptional activator activity, transcription factor recruiting activity, or

cellular uptake signaling activity. Other preferred embodiments of the invention may include any combination the activities described herein.

Zn-Finger Nucleases

[0233] In some embodiment, the nucleotide-binding molecule of the systems may be a Zn-finger nuclease, a functional fragment thereof, or a variant thereof. The composition may comprise one or more Zn-finger nucleases or nucleic acids encoding thereof. In some cases, the nucleotide sequences may comprise coding sequences for Zn-Finger nucleases. Other preferred tools for genome editing for use in the context of this invention include zinc finger systems and TALE systems. One type of programmable DNA-binding domain is provided by artificial zinc-finger (ZF) technology, which involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP).

[0234] ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, Proc. Natl. Acad. Sci. U.S.A. 91, 883−887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93, 1156−1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods 8, 74−79). ZFPs can also be designed as transcription activators and repressors and have been used to target many genes in a wide variety of organisms. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Patent Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference.

Meganucleases

[0235] In some embodiment, the nucleotide-binding domain may be a meganuclease, a functional fragment thereof, or a variant thereof. The composition may comprise one or more meganucleases or nucleic acids encoding thereof. As disclosed herein editing can be made by way of meganucleases, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). In some cases, the nucleotide sequences may comprise coding sequences for meganucleases. Exemplary method for using

meganucleases can be found in US Patent Nos: 8,163,514; 8,133,697; 8,021,867; 8,119,361; 8,119,381; 8,124,369; and 8,129,134, which are specifically incorporated by reference.

[0236] In certain embodiments, any of the nucleases, including the modified nucleases as described herein, may be used in the methods, compositions, and kits according to the invention. In particular embodiments, nuclease activity of an unmodified nuclease may be compared with nuclease activity of any of the modified nucleases as described herein, e.g. to compare for instance off-target or on-target effects. Alternatively, nuclease activity (or a modified activity as described herein) of different modified nucleases may be compared, e.g. to compare for instance off-target or on-target effects.

RNase Domains

[0237] The compositions and systems herein may further comprise one or more RNase domains. The RNase domain may be connected to the Cas polypeptide and/or the non-LTR retrotransposon polypeptide. Ribonucleases (RNases) are a type of nuclease that catalyzes the degradation of RNA into smaller components. RNases can be divided into endoribonucleases and exoribonucleases and play key roles in the maturation of all RNA molecules, both messenger RNAs that carry genetic material for making proteins, and non-coding RNAs that function in varied cellular processes. In addition, active RNA degradation systems are a first defense against RNA viruses, and provide the underlying machinery for more advanced cellular immune strategies such as RNAi. Examples of RNase domain include RNase A, RNaseH, RNaseIII, RNase L, and RNase P. In a particular example, the RNase domain is RNaseH.

[0238] RNase A is an RNase that is one of the hardiest enzymes in common laboratory usage; one method of isolating it is to boil a crude cellular extract until all enzymes other than RNase A are denatured. It is specific for single-stranded RNAs, where it cleaves the 3'-end of unpaired C and U residues, ultimately forming a 3'-phosphorylated product via a 2',3'-cyclic monophosphate intermediate. It does not require any cofactors for its activity.

[0239] RNaseH is a non-sequence-specific endonuclease that cleaves the RNA in a DNA/RNA duplex to via a hydrolytic mechanism to produce ssDNA. Members of the RNase H family can be found in nearly all organisms, from bacteria to archaea to eukaryotes. Ribonuclease H enzymes cleave the phosphodiester bonds of RNA in a double-stranded RNA:DNA hybrid, leaving a 3’ hydroxyl and a 5’ phosphate group on either end of the cut site. RNase H1 and H2 have distinct substrate preferences and distinct but overlapping functions in the cell. In prokaryotes and lower eukaryotes, neither enzyme is essential, whereas both are believed to be essential in higher eukaryotes. The combined activity of both H1 and

H2 enzymes is associated with maintenance of genome stability due to the enzymes' degradation of the RNA component of R loops.

[0240] RNase III is a type of ribonuclease that cleaves rRNA (16s rRNA and 23s rRNA) from transcribed polycistronic RNA operon in prokaryotes. It also digests double stranded RNA (dsRNA)-Dicer family of RNAse, cutting pre-miRNA (60–70bp long) at a specific site and transforming it in miRNA (22–30bp), that is actively involved in the regulation of transcription and mRNA life-time.

[0241] RNase L is an interferon-induced nuclease that, upon activation, destroys all RNA within the cell.

[0242] RNase P is a type of ribonuclease that is unique in that it is a ribozyme – a ribonucleic acid that acts as a catalyst in the same way as an enzyme. One of its functions is to cleave off a leader sequence from the 5' end of one stranded pre-tRNA. RNase P is one of two known multiple turnover ribozymes in nature (the other being the ribosome). In bacteria RNase P is also responsible for the catalytic activity of holoenzymes, which consist of an apoenzyme that forms an active enzyme system by combination with a coenzyme and determines the specificity of this system for a substrate.

[0243] In some embodiments, the engineered systems described herein, further comprise an RNase domain. In specific embodiments, the RNase domain may comprise, but is not necessarily limited to, an RNase H domain.

Nuclear localization sequences

[0244] In some embodiments, the polypeptides herein (e.g., site-specific nuclease polypeptides, the non-LTR retrotransposon polypeptide, or fusion protein thereof) may further comprise (e.g., fused to) one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the polypeptides and proteins comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g. zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). The NLS(s) may be at an internal location of the protein, i.e., not at the C-terminus or N-terminus. When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In a preferred embodiment of the invention, the polypeptides comprise at most 6 NLSs. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about

1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus.

[0245] In the cases of fusion protein comprising a site-specific nuclease polypeptide and a retrotransposon polypeptide, the one or more NLSs may be on any part of the fusion protein. In some examples, the NLS(s) is at the N-terminus of the fusion protein. In some examples, the NLS(s) is at the C-terminus of the fusion protein . In some example, the NLS(s) is at an internal location of the fusion protein, e.g., between the site-specific nuclease polypeptide and the retrotransposon polypeptide.Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO: 87); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 88); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 89) or RQRRNELKRSP (SEQ ID NO: 90); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 91); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 92) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 93) and PPKKARED (SEQ ID NO: 94) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 95) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 96) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 97) and PKQKKRK (SEQ ID NO: 98) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: 99) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: 100) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 101) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: 102) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the polypeptides in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the polypeptides, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the polypeptides, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of complex formation

(e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by complex formation and/or enzyme activity), as compared to a control no exposed to the polypeptides or complex, or exposed to a polypeptides lacking the one or more NLSs. In certain embodiments of the herein described polypeptides or complexes and systems the codon optimized polypeptides comprise an NLS attached to the C-terminal of the protein. In certain embodiments, other localization tags may be fused to the polypeptides, such as without limitation for localizing the polypeptides to particular sites in a cell, such as organelles, such as mitochondria, plastids, chloroplast, vesicles, golgi, (nuclear or cellular) membranes, ribosomes, nucleolus, ER, cytoskeleton, vacuoles, centrosome, nucleosome, granules, centrioles, etc.

[0246] In certain embodiments of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the polypeptides. In preferred embodiments at least one or more C-terminal or N-terminal NLSs are attached (and hence nucleic acid molecule(s) coding for the Cas protein can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. The invention also encompasses methods for delivering multiple nucleic acid components, wherein each nucleic acid component is specific for a different target locus of interest thereby modifying multiple target loci of interest. The nucleic acid component of the complex may comprise one or more protein-binding RNA aptamers. The one or more aptamers may be capable of binding a bacteriophage coat protein.

EXEMPLARY SYSTEMS AND COMPOSITIONS

[0247] In some embodiments, the non-naturally occurring or engineered systems or compositions comprise a Cas nickase fused with one or more retrotransposon polypeptides, a guide RNA for Cas targeting insertion site on genome of a cell, and one or more vectors comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas9 nickase (e.g. with D10A and/or H840A mutations) fused with retrotransposon polypeptide. R2 from B. mori, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cpf1 nickase fused with retrotransposon R2 from B. mori, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the

R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas12b nickase fused with retrotransposon R2 from B. mori, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0248] In some embodiments, the non-naturally occurring or engineered systems or compositions comprise a Cas nickase fused with one or more retrotransposon polypeptides, where the one or more retrotransposon polypeptides comprises a nuclease that is inactivated, a guide RNA for Cas targeting insertion site on genome in a cell, and one or more vectors comprising expression cassette comprising nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas9 nickase (with D10A and/or H840A mutations) fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cpf1 nickase fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein is inactivated, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas12b nickase fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0249] In some examples, the non-naturally occurring or engineered systems or compositions comprise a wildtype Cas fused with one or more retrotransposon polypeptides, a guide RNA for Cas targeting insertion site on genome, and one or more vectors expression cassette comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a wildtype Cas9 fused with retrotransposon R2 from B. mori, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise

wildtype Cpf1 fused with retrotransposon R2 from B. mori, guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas12b fused with retrotransposon R2 from B. mori, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0250] In some examples, the non-naturally occurring or engineered systems or compositions comprise a wildtype Cas fused with one or more retrotransposon polypeptides, where the one or more retrotransposon polypeptides comprises a nuclease domain that is inactivated, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas9 fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cpf1 fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas12b fused with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0251] In these examples, R2 be in the form of a dimer (e.g., see FIG. 4). In this case, a tandem fusion of R2 may be used. The construct may be dCas9 or Cas9 nickase with fusion to tandem R2. One of the R2 in the tandem dimer may be inactivated. So the construct may be dCas9 or Cas9 nickase fused to tandem R2 with one R2's nuclease domain inactivated. In some examples the 5' and 3' RNA for the R2 retrotransposon may include sequences shown in FIG. 5.

[0252] In some embodiments, the retrotransposon may comprise sequences encoding multiple polypeptides, e.g., comprise multiple open reading frames (ORFs). An exemplary mechanism of insertion is shown in FIG. 6. In some cases, the retrotransposon is L1.

[0253] In some embodiments, the systems or compositions comprise a Cas nickase fused with a first retrotransposon polypeptide (e.g., from a first ORF of a retrotransposon), a second retrotransposon polypeptide (e.g., from a second ORF of the retrotransposon), a guide RNA for Cas targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the systems or compositions comprise a Cas9 nickase (D10A or H840A) fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 nickase fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b nickase fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0254] In some embodiments, the systems or compositions comprise a dead Cas (dCas) fused with a first retrotransposon polypeptide (e.g., from a first ORF of a retrotransposon), a second retrotransposon polypeptide (e.g., from a second ORF of the retrotransposon), a guide RNA for Cas targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the systems or compositions comprise a dCas9 fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b fused with ORF2 of

LINE1, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0255] In some embodiments, the systems or compositions comprise a Cas nickase fused with a first retrotransposon polypeptide (e.g., from a first ORF of a retrotransposon) where the polypeptide contains a nuclease domain that is inactivated, a second retrotransposon polypeptide (e.g., from a second ORF of the retrotransposon), a guide RNA for Cas targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the systems or compositions comprise a Cas9 nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0256] In some embodiments, the systems or compositions comprise a wildtype Cas fused with a first retrotransposon polypeptide (e.g., from a first ORF of a retrotransposon) where the polypeptide contains a nuclease domain that is inactivated, a second retrotransposon polypeptide (e.g., from a second ORF of the retrotransposon), a guide RNA for Cas targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette comprising a nucleic acid polymerase promoter driving the expression of the retrotransposon RNA. In some examples, the systems or compositions comprise a wildtype Cas9 fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a wildtype Cpf1 fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a

polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a wildtype Cas12b fused with ORF2 of LINE1 where the nuclease domain has been inactivated, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0257] In some embodiments, the complexes of Cas and retrotransposon polypeptide(s) may be fused with one or more functional domains. In some embodiments, the complexes of Cas and retrotransposon polypeptide(s) may be fused with RNaseH domain. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas9 nickase fused with retrotransposon R2 from B. mori, where the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas9 nickase fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas9 fuse with retrotransposon R2 from B. mori, where the Cas9-R2 complex is also attached with RNaseH, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas9 fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0258] In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cpf1 nickase fused with retrotransposon R2 from B. mori, where the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally

occurring or engineered systems or compositions comprise a Cpf1 nickase fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas9 fuse with retrotransposon R2 from B. mori, where the Cpf1-R2 complex is also attached with RNaseH, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cpf1 fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0259] In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas12b nickase fused with retrotransposon R2 from B. mori, where the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise a Cas12b nickase fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas12b fuse with retrotransposon R2 from B. mori, where the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA. In some examples, the non-naturally occurring or engineered systems or compositions comprise wildtype Cas12b fuse with retrotransposon R2 from B. mori, where the nuclease domain in the R2 protein has been inactivated and the Cas-R2 complex is also attached with RNaseH, a guide RNA for Cas12b targeting insertion site on genome, and one or more

vectors expression cassette comprising Pol2 promoter driving the expression of the R2 transposon RNA.

[0260] In some examples, the systems or compositions comprise a Cas9 nickase (D10A or H840A) fused with ORF2 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 nickase fused with ORF2 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b nickase fused with ORF2 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0261] In some examples, the systems or compositions comprise a dCas9 fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a guide RNA for Cas9 targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a guide RNA for Cpf1 targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b fused with ORF2 of LINE1, a polypeptide expressed by ORF1 of LINE1 where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a guide RNA for Cas12b targeting insertion site on the genome of a cell, and one or more vectors comprising expression cassette consisting of Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0262] In some examples, the systems or compositions comprise a Cas9 nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising

expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cpf1 nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a Cas12b nickase fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0263] In some examples, the systems or compositions comprise a wildtype Cas9 fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas9 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a wildtype Cpf1 fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cpf1 targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA. In some examples, the systems or compositions comprise a wildtype Cas12b fused with ORF2 of LINE1 where the nuclease domain has been inactivated where the Cas-LINE1/ORF2 complex is also attached with RNaseH, a polypeptide expressed by ORF1 of LINE1, a guide RNA for Cas12b targeting insertion site on genome, and one or more vectors comprising expression cassette comprising Pol2 promoter driving the expression of the LINE1 transposon RNA.

[0264] In some embodiments, the systems and compositions may comprise two Cas proteins, each is associated with (e.g., fused to) a retrotransposon polypeptide. Directed by their guide RNA, the Cas proteins bind to different target sites on a target polynucleotide. Each Cas protein may make a break (double-stranded or single-stranded) on its target site. The systems further comprise a retrotransposon RNA bound with one or both of the retrotransposon polypeptide. An overhand from one strand of the target polynucleotide may hybridize a portion

of the retrotransposon RNA, which functions as the primer to synthesize a single-stranded cDNA using the retrotransposon RNA as the template. A second overhang (e.g., from the other strand of the target polynucleotide) may hybridize with a portion of the single-stranded cDNA and function as the primer to synthesize a second strand of the cDNA. The generated double-stranded cDNA may comprise a donor polynucleotide sequence to be inserted to a position in the target polynucleotide. The position may be between the two target sites of the Cas proteins. In some examples, the Cas proteins may be Type II Cas, e.g., Cas9. In certain examples, the Cas proteins may be Type V Cas, e.g., Cas12a, Cas12b, or Cas12c. In certain examples, the Cas protein may be a nickase, e.g., a Cas9 with an HNH domain inactivated. In certain examples, the retrotransposon polypeptides may be R2. In certain examples, the retrotransposon polypeptides may be L1, e.g., a polypeptide encoded by ORF of L1. The retrotransposon polypeptides may have an inactivated nuclease domain. Examples of the systems and compositions are shown in FIG. 7.

POLYNUCLEOTIDES AND VECTORS

[0265] The systems herein may comprise one or more polynucleotides. The polynucleotide(s) may comprise coding sequences of Cas protein(s), guide sequences, or any combination thereof. The present disclosure further provides vectors or vector systems comprising one or more polynucleotides herein. The vectors or vector systems include those described in the delivery sections herein.

[0266] The terms “polynucleotide”, “nucleotide”, “nucleotide sequence”, “nucleic acid” and “oligonucleotide” are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The

sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. A "wild type" can be a base line. As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature. The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions. As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Where reference is made to a polynucleotide sequence, then complementary or partially complementary sequences are also envisaged. These are preferably capable of hybridizing to the reference sequence under highly stringent conditions. Generally, in order to maximize the hybridization rate, relatively

low-stringency hybridization conditions are selected: about 20 to 25° C lower than the thermal melting point (Tm). The Tm is the temperature at which 50% of specific target sequence hybridizes to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridized sequences, highly stringent washing conditions are selected to be about 5 to 15° C lower than the Tm. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

[0267] As used herein, the term "genomic locus" or "locus" (plural loci) is the specific location of a gene or DNA sequence on a chromosome. A "gene" refers to stretches of DNA or RNA that encode a polypeptide or an RNA chain that has functional role to play in an organism and hence is the molecular unit of heredity in living organisms. For the purpose of this invention, it may be considered that genes include regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. As used herein, "expression of a genomic locus" or "gene expression" is the process by which information from a gene is used in the synthesis of a functional gene product. The products of gene expression are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is functional RNA. The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses to generate functional products to survive. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context. As used herein, "expression" also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond

formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. As used herein, the term "domain" or "protein domain" refers to a part of a protein sequence that may exist and function independently of the rest of the protein chain. As described in aspects of the invention, sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences.

[0268] In certain embodiments, the polynucleotide sequence is recombinant DNA. In further embodiments, the polynucleotide sequence further comprises additional sequences as described elsewhere herein. In certain embodiments, the nucleic acid sequence is synthesized *in vitro*.

[0269] Aspects of the invention relate to polynucleotide molecules that encode one or more components of the CRISPR-Cas system or Cas protein as referred to in any embodiment herein. In certain embodiments, the polynucleotide molecules may comprise further regulatory sequences. By means of guidance and not limitation, the polynucleotide sequence can be part of an expression plasmid, a minicircle, a lentiviral vector, a retroviral vector, an adenoviral or adeno-associated viral vector, a piggyback vector, or a tol2 vector. In certain embodiments, the polynucleotide sequence may be a bicistronic expression construct. In further embodiments, the isolated polynucleotide sequence may be incorporated in a cellular genome. In yet further embodiments, the isolated polynucleotide sequence may be part of a cellular genome. In further embodiments, the isolated polynucleotide sequence may be comprised in an artificial chromosome. In certain embodiments, the 5' and/or 3' end of the isolated polynucleotide sequence may be modified to improve the stability of the sequence of actively avoid degradation. In certain embodiments, the isolated polynucleotide sequence may be comprised in a bacteriophage. In other embodiments, the isolated polynucleotide sequence may be contained in agrobacterium species. In certain embodiments, the isolated polynucleotide sequence is lyophilized.

mRNA

[0270] In some embodiments, the composition comprises mRNA molecules comprising coding sequences of (i) the site-specific nuclease polypeptide(s) and/or (ii) the non-LTR

retrotransposon polypeptide(s). In certain examples, a single mRNA molecule comprises coding sequences of (i) the site-specific nuclease polypeptide(s) and (ii) the non-LTR retrotransposon polypeptide(s), e.g., a fusion protein comprising (i) and (ii).

[0271] In some embodiments, the mRNA molecules comprise a poly-A tail (e.g., at its 3' end). A poly-A tail refers to a sequence a sequence of adenyl (A) residues located on the end (e.g., 3' end) of the RNA molecule. In some examples, an mRNA molecule comprising one or more coding sequences of the site-specific nuclease polypeptide(s) comprises a poly-A tail. In some examples, an mRNA molecule comprising one or more coding sequences of the non-LTR retrotransposon polypeptide(s) comprises a poly-A tail. In some examples, an mRNA molecule comprising one or more coding sequences of both (i) the site-specific nuclease polypeptide(s) and (ii) the non-LTR retrotransposon polypeptide(s) (e.g., a fusion protein comprising (i) and (ii)) comprises a poly-A tail.

[0272] For example, the poly-A tail may comprise from 1 to 500, from 50 to 400, from 50 to 350, from 50 to 300, from 100 to 250, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350 adenyl (A) residues.

Codon optimization

[0273] Aspects of the invention relate to polynucleotide molecules that encode one or more components of one or more CRISPR-Cas systems as described in any of the embodiments

herein, wherein at least one or more regions of the polynucleotide molecule may be codon optimized for expression in a eukaryotic cell. In certain embodiments, the polynucleotide molecules that encode one or more components of one or more CRISPR-Cas systems as described in any of the embodiments herein are optimized for expression in a mammalian cell or a plant cell.

[0274] An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in International Patent Publication No. WO 2014/093622 (PCT/US2013/074667) as an example of a codon optimized sequence (from knowledge in the art and this disclosure, codon optimizing coding nucleic acid molecule(s), especially as to effector protein is within the ambit of the skilled artisan). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a DNA/RNA-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence.

[0275] Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon

usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid.

Vector systems

[0276] The present disclosure provides vector systems one or more vectors, the one or more vectors comprising one or more polynucleotides encoding components in retrotransposon herein, or combination thereof. The one or more polynucleotides in the vector systems may comprise one or more regulatory elements operably configures to express the polypeptide(s) and/or the nucleic acid component(s), optionally wherein the one or more regulatory elements comprise inducible promoters. The polynucleotide molecule encoding the Cas polypeptide is codon optimized for expression in a eukaryotic cell.

[0277] As described previously and as used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. The term "vector" includes cloning and expression vectors, as well as viral vectors and integrating vectors. An "expression vector" is a vector that includes one or more expression control sequences, and an "expression control sequence" is a DNA sequence that controls and regulates the transcription and/or translation of another DNA sequence. Suitable expression vectors include, without limitation, plasmids and viral vectors derived from, for example, bacteriophage, baculoviruses, tobacco mosaic virus, herpes viruses, cytomegalovirus, retroviruses, vaccinia viruses, adenoviruses, and adeno-associated viruses. Numerous vectors and expression systems are commercially available from such corporations as Novagen (Madison, WI), Clontech (Palo Alto, CA), Stratagene (La Jolla, CA), and Invitrogen/Life Technologies (Carlsbad, CA). By way of example, some vectors used in recombinant DNA techniques allow entities, such as a segment of DNA (such as a heterologous DNA segment, such as a heterologous cDNA segment), to be transferred into a target cell. The present invention comprehends recombinant vectors that may include viral vectors, bacterial

vectors, protozoan vectors, DNA vectors, or recombinants thereof. With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, the contents of which are herein incorporated by reference in their entirety.

[0278] A vector may have one or more restriction endonuclease recognition sites (whether type I, II or IIs) at which the sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a nucleic acid fragment may be spliced or inserted in order to bring about its replication and cloning. Vectors may also comprise one or more recombination sites that permit exchange of nucleic acid sequences between two nucleic acid molecules. Vectors may further provide primer sites, e.g., for PCR, transcriptional and/or translational initiation and/or regulation sites, recombinational signals, replicons, selectable markers, etc. A vector may further contain one or more selectable markers suitable for use in the identification of cells transformed with the vector.

[0279] As mentioned previously, vectors capable of directing the expression of genes and/or nucleic acid sequence to which they are operatively linked, in an appropriate host cell (e.g., a prokaryotic cell, eukaryotic cell, or mammalian cell), are referred to herein as "expression vectors." If translation of the desired nucleic acid sequence is required, such as for example, the mRNA encoding a TALE polypeptide, the vector also typically may comprise sequences required for proper translation of the nucleotide sequence. The term "expression" as used herein with regards to expression vectors, refers to the biosynthesis of a nucleic acid sequence product, i.e., to the transcription and/or translation of a nucleotide sequence, for example, a nucleic acid sequence encoding a TALE polypeptide in a cell. Expression also refers to biosynthesis of a microRNA or RNAi molecule, which refers to expression and transcription of an RNAi agent such as siRNA, shRNA, and antisense DNA, that do not require translation to polypeptide sequences.

[0280] In general, expression vectors of utility in the methods of generating and compositions which may comprise polypeptides of the invention described herein are often in the form of "plasmids," which refer to circular double-stranded DNA loops which, in their vector form, are not bound to a chromosome. In some embodiments of the aspects described herein, all components of a given polypeptide may be encoded in a single vector. For example, in some embodiments, a vector may be constructed that contains or may comprise all components necessary for a functional polypeptide as described herein. In some embodiments, individual components (e.g., one or more monomer units and one or more effector domains) may be separately encoded in different vectors and introduced into one or more cells separately. Moreover, any vector described herein may itself comprise predetermined Cas and/or

retrotransposon polypeptides encoding component sequences, such as an effector domain and/or other polypeptides, at any location or combination of locations, such as 5′ to, 3′ to, or both 5′ and 3′ to the exogenous nucleic acid molecule which may comprise one or more component Cas and/or retrotransposon polypeptides encoding sequences to be cloned in. Such expression vectors are termed herein as which may comprise "backbone sequences."

[0281] Several embodiments of the invention relate to vectors that include but are not limited to plasmids, episomes, bacteriophages, or viral vectors, and such vectors may integrate into a host cell's genome or replicate autonomously in the particular cellular system used. In some embodiments of the compositions and methods described herein, the vector used is an episomal vector, i.e., a nucleic acid capable of extra-chromosomal replication and may include sequences from bacteria, viruses or phages. Other embodiments of the invention relate to vectors derived from bacterial plasmids, bacteriophages, yeast episomes, yeast chromosomal elements, and viruses, vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, cosmids and phagemids. In some embodiments, a vector may be a plasmid, bacteriophage, bacterial artificial chromosome (BAC) or yeast artificial chromosome (YAC). A vector may be a single- or double-stranded DNA, RNA, or phage vector.

[0282] Viral vectors include, but are not limited to, retroviral vectors, such as lentiviral vectors or gammaretroviral vectors, adenoviral vectors, and baculoviral vectors. For example, a lentiviral vector may be used in the form of lentiviral particles. Other forms of expression vectors known by those skilled in the art which serve equivalent functions may also be used. Expression vectors may be used for stable or transient expression of the polypeptide encoded by the nucleic acid sequence being expressed. A vector may be a self-replicating extrachromosomal vector or a vector which integrates into a host genome. One type of vector is a genomic integrated vector, or "integrated vector", which may become integrated into the chromosomal DNA or RNA of a host cell, cellular system, or non-cellular system. In some embodiments, the nucleic acid sequence encoding the Cas and/or retrotransposon polypeptides described herein, integrates into the chromosomal DNA or RNA of a host cell, cellular system, or non-cellular system along with components of the vector sequence.

[0283] The recombinant expression vectors used herein comprise a Cas and/or retrotransposon nucleic acid in a form suitable for expression of the nucleic acid in a host cell, which indicates that the recombinant expression vector(s) include one or more regulatory sequences, selected on the basis of the host cell(s) to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed.

[0284] As used herein, the term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., 5′ and 3′ untranslated regions (UTRs) and polyadenylation signals). With regards to regulatory sequences, mention is made of U.S. patent application 10/491,026, the contents of which are incorporated by reference herein in their entirety.

[0285] The terms "promoter", "promoter element" or "promoter sequence" are equivalents and as used herein, refer to a DNA sequence which when operatively linked to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. Promoters may be constitutive, inducible or regulatable. The term "tissue-specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue. Tissue specificity of a promoter may be evaluated by methods known in the art. The term "cell-type specific" as applied to a promoter refers to a promoter, which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell-type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell-type specificity of a promoter may be assessed using methods well known in the art, e.g., GUS activity staining or immunohistochemical staining. The term "minimal promoter" as used herein refers to the minimal nucleic acid sequence which may comprise a promoter element while also maintaining a functional promoter. A minimal promoter may comprise an inducible, constitutive or tissue-specific promoter. With regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application 12/511,940, the contents of which are incorporated by reference herein in their entirety.

[0286] In advantageous embodiments of the invention, the expression vectors described herein may be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., Cas and/or retrotransposon polypeptides, variant forms thereof).

[0287] In some embodiments, the recombinant expression vectors which may comprise a nucleic acid encoding a Cas and/or retrotransposon polypeptide described herein further comprise a 5′UTR sequence and/or a 3′ UTR sequence, thereby providing the nucleic acid sequence transcribed from the expression vector additional stability and translational efficiency.

[0288] Certain embodiments of the invention may relate to the use of prokaryotic vectors and variants and derivatives thereof. Other embodiments of the invention may relate to the use of eukaryotic expression vectors. With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety.

[0289] In some embodiments of the aspects described herein, a Cas and/or retrotransposon polypeptide is expressed using a yeast expression vector. Examples of vectors for expression in yeast S. cerevisiae include, but are not limited to, pYepSec1 (Baldari, et al., (1987) EMBO J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

[0290] In other embodiments of the invention, Cas and/or retrotransposon polypeptides are expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include, but are not limited to, the pAc series (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39).

[0291] In some embodiments of the aspects described herein, Cas and/or retrotransposon polypeptides are expressed in mammalian cells using a mammalian expression vector. Non-limiting examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. With regards to viral regulatory elements, mention is made of U.S. patent application 13/248,967, the contents of which are incorporated by reference herein in their entirety.

[0292] In some such embodiments, the mammalian expression vector is capable of directing expression of the nucleic acid encoding the Cas and/or retrotransposon polypeptides in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety.

[0293] The vectors which may comprise nucleic acid sequences encoding the Cas and/or retrotransposon polypeptides described herein may be "introduced" into cells as

polynucleotides, preferably DNA, by techniques well known in the art for introducing DNA and RNA into cells. The term "transduction" refers to any method whereby a nucleic acid sequence is introduced into a cell, e.g., by transfection, lipofection, electroporation (methods whereby an instrument is used to create micro-sized holes transiently in the plasma membrane of cells under an electric discharge, see, e.g., Banerjee et al., Med. Chem. 42:4292-99 (1999); Godbey et al., Gene Ther. 6:1380-88 (1999); Kichler et al., Gene Ther. 5:855-60 (1998); Birchaa et al., J. Pharm. 183:195-207 (1999)), biolistics, passive uptake, lipid:nucleic acid complexes, viral vector transduction, injection, contacting with naked DNA, gene gun (whereby the nucleic acid is coupled to a nanoparticle of an inert solid (commonly gold) which is then "shot" directly into the target cell's nucleus), calcium phosphate, DEAE dextran, lipofectin, lipofectamine, DIMRIE C, Superfect, and Effectin (Qiagen), unifectin, maxifectin, DOTMA, DOGS (Transfectam; dioctadecylamidoglycylspermine), DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine), DOTAP (1,2-dioleoyl-3-trimethylammonium propane), DDAB (dimethyl dioctadecylammonium bromide), DHDEAB (N,N-di-n-hexadecyl-N,N-dihydroxyethyl ammonium bromide), HDEAB (N-n-hexadecyl-N,N-dihydroxyethylammonium bromide), polybrene, poly(ethylenimine) (PEI), sono-poration (transfection via the application of sonic forces to cells), optical transfection (methods whereby a tiny (~1 μm diameter) hole is transiently generated in the plasma membrane of a cell using a highly focused laser), magnetofection (refers to a transfection method, that uses magnetic force to deliver exogenous nucleic acids coupled to magnetic nanoparticles into target cells), impalefection (carried out by impaling cells by elongated nanostructures, such as carbon nanofibers or silicon nanowires which were coupled to exogenous nucleic acids), and the like. In this regard, mention is made of U.S. Patent Application 13/088,009, the contents of which are incorporated by reference herein in their entirety.

[0294] The nucleic acid sequences encoding the Cas and/or retrotransposon polypeptides or the vectors which may comprise the nucleic acid sequences encoding the Cas and/or retrotransposon polypeptides described herein may be introduced into a cell using any method known to one of skill in the art. The term "transformation" as used herein refers to the introduction of genetic material (e.g., a vector which may comprise a nucleic acid sequence encoding a Cas and/or retrotransposon polypeptides) into a cell, tissue or organism. Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA), which detects

the presence of a polypeptide encoded by one or more of the transgenes. For example, a nucleic acid sequence encoding Cas and/or retrotransposon polypeptides may further comprise a constitutive promoter operably linked to a second output product, such as a reporter protein. Expression of that reporter protein indicates that a cell has been transformed or transfected with the nucleic acid sequence encoding Cas and/or retrotransposon polypeptides. Alternatively, or in combination, transient transformation may be detected by detecting the activity of the Cas and/or retrotransposon polypeptides. The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes.

[0295] In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell or cellular system, preferably resulting in chromosomal integration and stable heritability through meiosis. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences, which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell, which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene. Transformation also includes introduction of genetic material into plant cells in the form of plant viral vectors involving epichromosomal replication and gene expression, which may exhibit variable properties with respect to meiotic stability. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

[0296] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable biomarker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable biomarker may be introduced into a host cell on the same vector as that encoding Cas and/or retrotransposon polypeptides or may be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid may be identified by drug selection (e.g., cells that have incorporated the selectable biomarker gene survive, while the other cells die). With

regards to transformation, mention is made to U.S. Patent 6,620,986, the contents of which are incorporated by reference herein in their entirety.

METHOD OF INSERTING POLYNUCLEOTIDES

[0297] The present disclosure further provides methods of inserting a polynucleotide into a target nucleic acid. Examples of the methods comprise introducing the engineered or non-naturally occurring systems or compositions herein to a cell or population of cells, wherein the CRISPR-Cas complex directs the non-LTR retrotransposon to the target sequence, and wherein the non-LTR retrotransposon inserts the donor polynucleotide encoded by the retrotransposon RNA at or adjacent to the target sequence.

Immune orthogonal orthologs

[0298] In some embodiments, when components of the systems and compositions need to be expressed or administered in a subject, immunogenicity of components of the systems and compositions may be reduced by sequentially expressing or administering immune orthogonal orthologs of the components of the systems and compositions to the subject. As used herein, the term "immune orthogonal orthologs" refer to orthologous proteins that have similar or substantially the same function or activity, but have no or low cross-reactivity with the immune response generated by one another. In some embodiments, sequential expression or administration of such orthologs elicits low or no secondary immune response. The immune orthogonal orthologs can avoid being neutralized by antibodies (e.g., existing antibodies in the host before the orthologs are expressed or administered). Cells expressing the orthologs can avoid being cleared by the host's immune system (e.g., by activated CTLs). In some examples, CRISPR enzyme orthologs from different species may be immune orthogonal orthologs.

[0299] Immune orthogonal orthologs may be identified by analyzing the sequences, structures, and/or immunogenicity of a set of candidates orthologs. In an example method, a set of immune orthogonal orthologs may be identified by a) comparing the sequences of a set of candidate orthologs (e.g., orthologs from different species) to identify a subset of candidates that have low or no sequence similarity; b) assessing immune overlap among the members of the subset of candidates to identify candidates that have no or low immune overlap. In some cases, immune overlap among candidates may be assessed by determining the binding (e.g., affinity) between a candidate ortholog and MHC (e.g., MHC type I and/or MHC II) of the host. Alternatively or additionally, immune overlap among candidates may be assessed by determining B-cell epitopes for the candidate orthologs. In one example, immune orthogonal

orthologs may be identified using the method described in Moreno AM et al., BioRxiv, published online January 10, 2018, doi: doi.org/10.1101/245985.

DELIVERY

[0300] The present disclosure also provides delivery systems for introducing components of the systems and compositions herein to cells, tissues, organs, or organisms. A delivery system may comprise one or more delivery vehicles and/or cargos. Exemplary delivery systems and methods include those described in paragraphs [00117] to [00278] of Feng Zhang et al., (WO2016106236A1), and pages 1241-1251 and Table 1 of Lino CA et al., Delivering CRISPR: a review of the challenges and approaches, DRUG DELIVERY, 2018, VOL. 25, NO. 1, 1234–1257, which are incorporated by reference herein in their entireties.

[0301] In some embodiments, the delivery systems may be used to introduce the components of the systems and compositions to plant cells. For example, the components may be delivered to plant using electroporation, microinjection, aerosol beam injection of plant cell protoplasts, biolistic methods, DNA particle bombardment, and/or Agrobacterium-mediated transformation. Examples of methods and delivery systems for plants include those described in Fu et al., Transgenic Res. 2000 Feb;9(1):11-9; Klein RM, et al., Biotechnology. 1992;24:384-6; Casas AM et al., Proc Natl Acad Sci U S A. 1993 Dec 1; 90(23): 11212–11216; and U.S. Pat. No. 5,563,055, Davey MR et al., Plant Mol Biol. 1989 Sep;13(3):273-85, which are incorporated by reference herein in their entireties.

Cargos

[0302] The delivery systems may comprise one or more cargos. The cargos may comprise one or more components of the systems and compositions herein. A cargo may comprise one or more of the following: i) one or more plasmids encoding the engineered proteins; (ii) mRNA molecules encoding the engineered proteins; (iii) the engineered proteins. In some examples, a cargo may comprise a plasmid encoding one or more engineered proteins herein.

Physical delivery

[0303] In some embodiments, the cargos may be introduced to cells by physical delivery methods. Examples of physical methods include microinjection, electroporation, and hydrodynamic delivery. Both nucleic acid and proteins may be delivered using such methods. For example, the engineered protein or mRNA thereof may be prepared *in vitro*, isolated, (refolded, purified if needed), and introduced to cells.

Microinjection

[0304] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In some embodiments, microinjection may be performed using a

microscope and a needle (e.g., with 0.5–5.0 μm in diameter) to pierce a cell membrane and deliver the cargo directly to a target site within the cell. Microinjection may be used for *in vitro* and *ex vivo* delivery.

[0305] Plasmids comprising coding sequences for the engineered proteins may be microinjected. In some cases, microinjection may be used i) to deliver DNA directly to a cell nucleus, and/or ii) to deliver mRNA (e.g., *in vitro* transcribed) to a cell nucleus or cytoplasm.

[0306] Microinjection may be used to generate genetically modified animals. For example, gene editing cargos may be injected into zygotes to allow for efficient germline modification. Such approach can yield normal embryos and full-term mouse pups harboring the desired modification(s).

Electroporation

[0307] In some embodiments, the cargos and/or delivery vehicles may be delivered by electroporation. Electroporation may use pulsed high-voltage electrical currents to transiently open nanometer-sized pores within the cellular membrane of cells suspended in buffer, allowing for components with hydrodynamic diameters of tens of nanometers to flow into the cell. In some cases, electroporation may be used on various cell types and efficiently transfer cargo into cells. Electroporation may be used for in vitro and ex vivo delivery.

[0308] Electroporation may also be used to deliver the cargo to into the nuclei of mammalian cells by applying specific voltage and reagents, e.g., by nucleofection. Such approaches include those described in Wu Y, et al. (2015). Cell Res 25:67–79; Ye L, et al. (2014). Proc Natl Acad Sci USA 111:9591–6; Choi PS, Meyerson M. (2014). Nat Commun 5:3728; Wang J, Quake SR. (2014). Proc Natl Acad Sci 111:13157–62. Electroporation may also be used to deliver the cargo *in vivo*, e.g., with methods described in Zuckermann M, et al. (2015). Nat Commun 6:7391.

Hydrodynamic delivery

[0309] Hydrodynamic delivery may also be used for delivering the cargos, e.g., for *in vivo* delivery. In some examples, hydrodynamic delivery may be performed by rapidly pushing a large volume (8–10% body weight) solution containing the gene editing cargo into the bloodstream of a subject (e.g., an animal or human), e.g., for mice, via the tail vein. As blood is incompressible, the large bolus of liquid may result in an increase in hydrodynamic pressure that temporarily enhances permeability into endothelial and parenchymal cells, allowing for cargo not normally capable of crossing a cellular membrane to pass into cells. This approach may be used for delivering naked DNA plasmids and proteins. The delivered cargos may be enriched in liver, kidney, lung, muscle, and/or heart.

Transfection

[0310] The cargos, e.g., nucleic acids, may be introduced to cells by transfection methods for introducing nucleic acids into cells. Examples of transfection methods include calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acid.

Delivery vehicles

[0311] The delivery systems may comprise one or more delivery vehicles. The delivery vehicles may deliver the cargo into cells, tissues, organs, or organisms (e.g., animals or plants). The cargos may be packaged, carried, or otherwise associated with the delivery vehicles. The delivery vehicles may be selected based on the types of cargo to be delivered, and/or the delivery is *in vitro* and/or *in vivo*. Examples of delivery vehicles include vectors, viruses, non-viral vehicles, and other delivery reagents described herein.

[0312] The delivery vehicles in accordance with the present invention may a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, the delivery vehicles have a greatest dimension of less than 10 μm. In some embodiments, the delivery vehicles may have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension (e.g., diameter) of less than 900 nm, less than 800 nm, less than 700 nm, less than 600 nm, less than 500 nm, less than 400 nm, less than 300 nm, less than 200 nm, less than 150nm, or less than 100nm, less than 50nm. In some embodiments, the delivery vehicles may have a greatest dimension ranging between 25 nm and 200 nm.

[0313] In some embodiments, the delivery vehicles may be or comprise particles. For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension (e.g., diameter) no greater than 1000nm. The particles may be provided in different forms, e.g., as solid particles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of particles, or combinations thereof. Metal, dielectric, and semiconductor particles may be prepared, as well as hybrid structures (e.g., core–shell particles).

[0314] Nanoparticles may also be used to deliver the compositions and systems to plant cells, e.g., as described in WO 2008042156, US 20130185823, and WO2015089419.

Vectors

[0315] The systems, compositions, and/or delivery systems may comprise one or more vectors. The present disclosure also includes vector systems. A vector system may comprise one or more vectors. In some embodiments, a vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. A vector may be a plasmid, e.g., a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Certain vectors may be capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Some vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. In certain examples, vectors may be expression vectors, e.g., capable of directing the expression of genes to which they are operatively-linked. In some cases, the expression vectors may be for expression in eukaryotic cells. Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0316] Examples of vectors include pGEX, pMAL, pRIT5, E. coli expression vectors (e.g., pTrc, pET 11d, yeast expression vectors (e.g., pYepSec1, pMFa, pJRY88, pYES2, and picZ, Baculovirus vectors (e.g., for expression in insect cells such as SF9 cells) (e.g., pAc series and the pVL series), mammalian expression vectors (e.g., pCDM8 and pMT2PC.

[0317] In a single vector there can be a promoter for each RNA coding sequence. Alternatively or additionally, in a single vector, there may be a promoter controlling (e.g., driving transcription and/or expression) multiple RNA encoding sequences.

Regulatory elements

[0318] A vector may comprise one or more regulatory elements. The regulatory element(s) may be operably linked to coding sequences of the engineered proteins. The term "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0319] Examples of regulatory elements include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination

signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific.

[0320] Examples of promoters include one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter.

Viral vectors

[0321] The cargos may be delivered by viruses. In some embodiments, viral vectors are used. A viral vector may comprise virally-derived DNA or RNA sequences for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Viruses and viral vectors may be used for *in vitro*, *ex vivo*, and/or *in vivo* deliveries.

Adeno associated virus (AAV)

[0322] The systems and compositions herein may be delivered by adeno associated virus (AAV). AAV vectors may be used for such delivery. AAV, of the Dependovirus genus and Parvoviridae family, is a single stranded DNA virus. In some embodiments, AAV may provide a persistent source of the provided DNA, as AAV delivered genomic material can exist indefinitely in cells, e.g., either as exogenous DNA or, with some modification, be directly integrated into the host DNA. In some embodiments, AAV do not cause or relate with any

diseases in humans. The virus itself is able to efficiently infect cells while provoking little to no innate or adaptive immune response or associated toxicity.

[0323] Examples of AAV that can be used herein include AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, and AAV-9. The type of AAV may be selected with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium. Examples of cell types targeted by AAV are described in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)). In some examples, AAV particles may be created in HEK 293 T cells. Once particles with specific tropism have been created, they are used to infect the target cell line much in the same way that native viral particles do. This may allow for persistent presence of engineered proteins in the infected cell type, and what makes this version of delivery particularly suited to cases where long-term expression is desirable. Examples of doses and formulations for AAV that can be used include those describe in US Patent Nos. 8,454,972 and 8,404,658.

[0324] Various strategies may be used for delivery the systems and compositions herein with AAVs. In some examples, coding sequences of engineered proteins may be packaged directly onto one DNA plasmid vector and delivered via one AAV particle. In some examples, AAVs may be used to deliver gRNAs into cells that have been previously engineered to express the engineered protein. In some examples, coding sequences of two or more engineered proteins may be made into two separate AAV particles, which are used for co-transfection of target cells.

Lentiviruses

[0325] The systems and compositions herein may be delivered by lentiviruses. Lentiviral vectors may be used for such delivery. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells.

[0326] Examples of lentiviruses include human immunodeficiency virus (HIV), which may use its envelope glycoproteins of other viruses to target a broad range of cell types; minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV), which may be used for ocular therapies. In certain embodiments, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al.

(2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the nucleic acid-targeting system herein.

[0327] Lentiviruses may be pseudo-typed with other viral proteins, such as the G protein of vesicular stomatitis virus. In doing so, the cellular tropism of the lentiviruses can be altered to be as broad or narrow as desired. In some cases, to improve safety, second- and third-generation lentiviral systems may split essential genes across three plasmids, which may reduce the likelihood of accidental reconstitution of viable viral particles within cells.

[0328] In some examples, leveraging the integration ability, lentiviruses may be used to create libraries of cells comprising various genetic modifications, e.g., for screening and/or studying genes and signaling pathways.

Adenoviruses

[0329] The systems and compositions herein may be delivered by adenoviruses. Adenoviral vectors may be used for such delivery. Adenoviruses include nonenveloped viruses with an icosahedral nucleocapsid containing a double stranded DNA genome. Adenoviruses may infect dividing and non-dividing cells.

Viral vehicles for delivery to plants

[0330] The systems and compositions may be delivered to plant cells using viral vehicles. In particular embodiments, the compositions and systems may be introduced in the plant cells using a plant viral vector (e.g., as described in Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). Such viral vector may be a vector from a DNA virus, e.g., geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). The viral vector may be a vector from an RNA virus, e.g., tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses may be non-integrative vectors.

Non-viral vehicles

[0331] The delivery vehicles may comprise non-viral vehicles. In general, methods and vehicles capable of delivering nucleic acids and/or proteins may be used for delivering the systems compositions herein. Examples of non-viral vehicles include lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, gold nanoparticles, streptolysin O, multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Lipid particles

[0332] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes.

Lipid nanoparticles (LNPs)

[0333] LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes), and may be delivered to cells with relative ease. In some examples, lipid nanoparticles do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro, ex vivo*, and *in vivo* deliveries. Lipid particles may be used for various scales of cell populations.

[0334] In some examples. LNPs may be used for delivering DNA molecules and/or RNA molecules. In certain cases, LNPs may be use for delivering RNP complexes.

[0335] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), l,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), l,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-l,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011).

Liposomes

[0336] In some embodiments, a lipid particle may be liposome. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. In some embodiments, liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB).

[0337] Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as l,2-distearoryl-sn-glycero-3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

[0338] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin,

and/or l,2-dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

Stable nucleic-acid-lipid particles (SNALPs)

[0339] In some embodiments, the lipid particles may be stable nucleic acid lipid particles (SNALPs). SNALPs may comprise an ionizable lipid (DLinDMA) (e.g., cationic at low pH), a neutral helper lipid, cholesterol, a diffusible polyethylene glycol (PEG)-lipid, or any combination thereof. In some examples, SNALPs may comprise synthetic cholesterol, dipalmitoylphosphatidylcholine, 3-N-[(w-methoxy polyethylene glycol)2000)carbamoyl]-l,2-dimyrestyloxypropylamine, and cationic l,2-dilinoleyloxy-3-N,Ndimethylaminopropane. In some examples, SNALPs may comprise synthetic cholesterol, l,2-distearoyl-sn-glycero-3-phosphocholine, PEG- cDMA, and l,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA)

Other lipids

[0340] The lipid particles may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,3]- dioxolane (DLin-KC2-DMA), DLin-KC2-DMA4, C12- 200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG.

Lipoplexes/polyplexes

[0341] In some embodiments, the delivery vehicles comprise lipoplexes and/or polyplexes. Lipoplexes may bind to negatively charged cell membrane and induce endocytosis into the cells. Examples of lipoplexes may be complexes comprising lipid(s) and non-lipid components. Examples of lipoplexes and polyplexes include FuGENE-6 reagent, a non-liposomal solution containing lipids and other components, zwitterionic amino lipids (ZALs), Ca2þ (e.g., forming DNA/Ca^{2+} microcomplexes), polyethenimine (PEI) (e.g., branched PEI), and poly(L-lysine) (PLL).

Cell penetrating peptides

[0342] In some embodiments, the delivery vehicles comprise cell penetrating peptides (CPPs). CPPs are short peptides that facilitate cellular uptake of various molecular cargo (e.g., from nanosized particles to small chemical molecules and large fragments of DNA).

[0343] CPPs may be of different sizes, amino acid sequences, and charges. In some examples, CPPs can translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPPs may be introduced into cells via

different mechanisms, e.g., direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

[0344] CPPs may have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. Another type of CPPs is the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1). Examples of CPPs include to Penetratin, Tat (48-60), Transportan, and (R-AhX-R4) (Ahx refers to aminohexanoyl), Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence, polyarginine peptide Args sequence, Guanine rich-molecular transporters, and sweet arrow peptide. Examples of CPPs and related applications also include those described in US Patent 8,372,951.

[0345] CPPs can be used for *in vitro* and *ex vivo* work quite readily, and extensive optimization for each cargo and cell type is usually required. In some examples, CPPs may be covalently attached to the engineered protein directly, which is then complexed with the gRNA and delivered to cells. CPP may also be used to delivery RNPs.

[0346] CPPs may be used to deliver the compositions and systems to plants. In some examples, CPPs may be used to deliver the components to plant protoplasts, which are then regenerated to plant cells and further to plants.

DNA nanoclews

[0347] In some embodiments, the delivery vehicles comprise DNA nanoclews. A DNA nanoclew refers to a sphere-like structure of DNA (e.g., with a shape of a ball of yarn). The nanoclew may be synthesized by rolling circle amplification with palindromic sequences that aide in the self-assembly of the structure. The sphere may then be loaded with a payload. An example of DNA nanoclew is described in Sun W et al, J Am Chem Soc. 2014 Oct 22;136(42):14722-5; and Sun W et al, Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. A DNA nanoclew may be coated, e.g., coated with PEI to induce endosomal escape.

Gold nanoparticles

[0348] In some embodiments, the delivery vehicles comprise gold nanoparticles (also referred to AuNPs or colloidal gold). Gold nanoparticles may form complex with cargos. Gold nanoparticles may be coated, e.g., coated in a silicate and an endosomal disruptive polymer, PAsp(DET). Examples of gold nanoparticles include AuraSense Therapeutics' Spherical

Nucleic Acid (SNA™) constructs, and those described in Mout R, et al. (2017). ACS Nano 11:2452–8; Lee K, et al. (2017). Nat Biomed Eng 1:889–901.

iTOP

[0349] In some embodiments, the delivery vehicles comprise iTOP. iTOP refers to a combination of small molecules drives the highly efficient intracellular delivery of native proteins, independent of any transduction peptide. iTOP may be used for induced transduction by osmocytosis and propanebetaine, using NaCl-mediated hyperosmolality together with a transduction compound (propanebetaine) to trigger macropinocytotic uptake into cells of extracellular macromolecules. Examples of iTOP methods and reagents include those described in D'Astolfo DS, Pagliero RJ, Pras A, et al. (2015). Cell 161:674–690.

Polymer-based particles

[0350] In some embodiments, the delivery vehicles may comprise polymer-based particles (e.g., nanoparticles). In some embodiments, the polymer-based particles may mimic a viral mechanism of membrane fusion. The polymer-based particles may be a synthetic copy of Influenza virus machinery and form transfection complexes with various types of nucleic acids ((siRNA, miRNA, plasmid DNA or shRNA, mRNA) that cells take up via the endocytosis pathway, a process that involves the formation of an acidic compartment. The low pH in late endosomes acts as a chemical switch that renders the particle surface hydrophobic and facilitates membrane crossing. Once in the cytosol, the particle releases its payload for cellular action. This Active Endosome Escape technology is safe and maximizes transfection efficiency as it is using a natural uptake pathway.

Streptolysin O (SLO)

[0351] The delivery vehicles may be streptolysin O (SLO). SLO is a toxin produced by Group A streptococci that works by creating pores in mammalian cell membranes. SLO may act in a reversible manner, which allows for the delivery of proteins (e.g., up to 100 kDa) to the cytosol of cells without compromising overall viability. Examples of SLO include those described in Sierig G, et al. (2003). Infect Immun 71:446–55; Walev I, et al. (2001). Proc Natl Acad Sci U S A 98:3185–90; Teng KW, et al. (2017). Elife 6:e25460.

Multifunctional envelope-type nanodevice (MEND)

[0352] The delivery vehicles may comprise multifunctional envelope-type nanodevice (MENDs). MENDs may comprise condensed plasmid DNA, a PLL core, and a lipid film shell. A MEND may further comprise cell-penetrating peptide (e.g., stearyl octaarginine). The cell penetrating peptide may be in the lipid shell. The lipid envelope may be modified with one or more functional components, e.g., one or more of: polyethylene glycol (e.g., to increase

vascular circulation time), ligands for targeting of specific tissues/cells, additional cell-penetrating peptides (e.g., for greater cellular delivery), lipids to enhance endosomal escape, and nuclear delivery tags. In some examples, the MEND may be a tetra-lamellar MEND (T-MEND), which may target the cellular nucleus and mitochondria. In certain examples, a MEND may be a PEG-peptide-DOPE-conjugated MEND (PPD-MEND), which may target bladder cancer cells. Examples of MENDs include those described in Kogure K, et al. (2004). J Control Release 98:317–23; Nakamura T, et al. (2012). Acc Chem Res 45:1113–21.

Lipid-coated mesoporous silica particles

[0353] The delivery vehicles may comprise lipid-coated mesoporous silica particles. Lipid-coated mesoporous silica particles may comprise a mesoporous silica nanoparticle core and a lipid membrane shell. The silica core may have a large internal surface area, leading to high cargo loading capacities. In some embodiments, pore sizes, pore chemistry, and overall particle sizes may be modified for loading different types of cargos. The lipid coating of the particle may also be modified to maximize cargo loading, increase circulation times, and provide precise targeting and cargo release. Examples of lipid-coated mesoporous silica particles include those described in Du X, et al. (2014). Biomaterials 35:5580–90; Durfee PN, et al. (2016). ACS Nano 10:8325–45.

Inorganic nanoparticles

[0354] The delivery vehicles may comprise inorganic nanoparticles. Examples of inorganic nanoparticles include carbon nanotubes (CNTs) (e.g., as described in Bates K and Kostarelos K. (2013). Adv Drug Deliv Rev 65:2023–33.), bare mesoporous silica nanoparticles (MSNPs) (e.g., as described in Luo GF, et al. (2014). Sci Rep 4:6064), and dense silica nanoparticles (SiNPs) (as described in Luo D and Saltzman WM. (2000). Nat Biotechnol 18:893–5).

Exosomes

[0355] The delivery vehicles may comprise exosomes. Exosomes include membrane bound extracellular vesicles, which can be used to contain and delivery various types of biomolecules, such as proteins, carbohydrates, lipids, and nucleic acids, and complexes thereof (e.g., RNPs). Examples of exosomes include those described in Schroeder A, et al., J Intern Med. 2010 Jan;267(1):9-21; El-Andaloussi S, et al., Nat Protoc. 2012 Dec;7(12):2112-26; Uno Y, et al., Hum Gene Ther. 2011 Jun;22(6):711-9; Zou W, et al., Hum Gene Ther. 2011 Apr;22(4):465-75.

[0356] In some examples, the exosome may form a complex (e.g., by binding directly or indirectly) to one or more components of the cargo. In certain examples, a molecule of an

exosome may be fused with first adapter protein and a component of the cargo may be fused with a second adapter protein. The first and the second adapter protein may specifically bind each other, thus associating the cargo with the exosome. Examples of such exosomes include those described in Ye Y, et al., Biomater Sci. 2020 Apr 28. doi: 10.1039/d0bm00427h.

APPLICATIONS IN PLANTS AND FUNGI

[0357] The compositions, systems, and methods described herein can be used to perform gene or genome interrogation or editing or manipulation in plants and fungi. For example, the applications include investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant or fugus genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The compositions, systems, and methods can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques.

[0358] The compositions, systems, and methods herein may be used to confer desired traits (e.g., enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds) on essentially any plants and fungi, and their cells and tissues. The compositions, systems, and methods may be used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of any foreign gene.

[0359] In some embodiments, compositions, systems, and methods may be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; see, e.g., Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system," Plant Methods 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system," Plant Physiology September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system," Nature Biotechnology 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR/Cas system," Cell Research (2013) 23:1229–1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system," Mol Plant. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice," Rice 2014, 7:5 (2014), Zhou et al., "Exploiting SNPs for biallelic CRISPR

mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy," New Phytologist (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliando et al, "Targeted DNA degradation using a CRISPR device stably carried in the host genome, NATURE COMMUNICATIONS 6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, Morrell et al "Crop genomics: advances and applications," Nat Rev Genet. 2011 Dec 29;13(2):85-96, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. Aspects of utilizing the compositions, systems, and methods may be analogous to the use of the CRISPR-Cas (e.g. CRISPR-Cas9) system in plants, and mention is made of the University of Arizona website "CRISPR-PLANT" (www.genome.arizona.edu/crispr/) (supported by Penn State and AGI).

[0360] The compositions, systems, and methods may also be used on protoplasts. A "protoplast" refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

[0361] The compositions, systems, and methods may be used for screening genes (e.g., endogenous, mutations) of interest. In some examples, genes of interest include those encoding enzymes involved in the production of a component of added nutritional value or generally genes affecting agronomic traits of interest, across species, phyla, and plant kingdom. By selectively targeting e.g. genes encoding enzymes of metabolic pathways, the genes responsible for certain nutritional aspects of a plant can be identified. Similarly, by selectively targeting genes which may affect a desirable agronomic trait, the relevant genes can be identified. Accordingly, the present invention encompasses screening methods for genes encoding enzymes involved in the production of compounds with a particular nutritional value and/or agronomic traits.

[0362] It is also understood that reference herein to animal cells may also apply, *mutatis mutandis*, to plant or fungal cells unless otherwise apparent; and, the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein.

[0363] In some cases, nucleic acids introduced to plants and fungi may be codon optimized for expression in the plants and fungi. Methods of codon optimization include those described

in Kwon KC, et al., Codon Optimization to Enhance Expression Yields Insights into Chloroplast Translation, Plant Physiol. 2016 Sep;172(1):62-77.

[0364] The components in the compositions and systems may further comprise one or more functional domains described herein. In some examples, the functional domains may be an exonuclease. Such exonuclease may increase the efficiency of the component's function, e.g., mutagenesis efficiency. An example of the functional domain is Trex2, as described in Weiss T et al., www.biorxiv.org/content/10.1101/2020.04.11.037572v1, doi: doi.org/10.1101/2020.04.11.037572.

Examples of plants

[0365] The compositions, systems, and methods herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics. In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants.

[0366] The compositions, systems, and methods may be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders *Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales*; monocotyledonous plants such as those belonging to the orders *Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales*, and *Orchid ales*, or with plants belonging to *Gymnospermae*, e.g. those belonging to the orders *Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales* and *Gnetales*.

[0367] The compositions, systems, and methods herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera

hereunder: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vilis, and Vigna*; and the *genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus*, and *Pseudotsuga*.

[0368] In some embodiments, target plants and plant cells for engineering include those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini.

[0369] The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants. The compositions, systems, and methods can be used over a broad range of "algae" or "algae cells." Examples of algae include eukaryotic phyla, including the *Rhodophyta* (red algae), *Chlorophyta* (green algae), *Phaeophyta* (brown algae), *Bacillariophyta* (diatoms), *Eustigmatophyta* and dinoflagellates as well as the prokaryotic phylum *Cyanobacteria* (blue-green algae). Examples of algae species include those of *Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira, and Trichodesmium*.

Plant promoters

[0370] In order to ensure appropriate expression in a plant cell, the components of the components and systems herein may be placed under control of a plant promoter. A plant promoter is a promoter operable in plant cells. A plant promoter is capable of initiating transcription in plant cells, whether or not its origin is a plant cell. The use of different types of promoters is envisaged.

[0371] In some examples, the plant promoter is a constitutive plant promoter, which is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. In some examples, the plant promoter is a regulated promoter, which directs gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In some examples, the plant promoter is a tissue-preferred promoters, which can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed.

[0372] Exemplary plant promoters include those obtained from plants, plant viruses, and bacteria such as Agrobacterium or Rhizobium which comprise genes expressed in plant cells. Additional examples of promoters include those described in Kawamata et al., (1997) Plant

Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[0373] In some examples, a plant promoter may be an inducible promoter, which is inducible and allows for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome), such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. In a particular example, of the components of a light inducible system include a component of the system, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana), and a transcriptional activation/repression domain.

[0374] In some examples, the promoter may be a chemical-regulated promotor (where the application of an exogenous chemical induces gene expression) or a chemical-repressible promoter (where application of the chemical represses gene expression). Examples of chemical-inducible promoters include maize ln2-2 promoter (activated by benzene sulfonamide herbicide safeners), the maize GST promoter (activated by hydrophobic electrophilic compounds used as pre-emergent herbicides), the tobacco PR-1 a promoter (activated by salicylic acid), promoters regulated by antibiotics (such as tetracycline-inducible and tetracycline-repressible promoters).

Stable integration in the genome of plants

[0375] In some embodiments, polynucleotides encoding the components of the compositions and systems may be introduced for stable integration into the genome of a plant cell. In some cases, vectors or expression systems may be used for such integration. The design of the vector or the expression system can be adjusted depending on for when, where and under what conditions the guide RNA and/or the component(s) in the system are expressed. In some cases, the polynucleotides may be integrated into an organelle of a plant, such as a plastid, mitochondrion or a chloroplast. The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

[0376] In some embodiments, the method of integration generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host

tissue, and regenerating plant cells or plants therefrom. In some examples, the expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the RNA and/or component(s) of the system in a plant cell; a 5' untranslated region to enhance expression ; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the guide RNA and/or the gene sequences of component(s) of the system and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

Transient expression in plants

[0377] In some embodiments, the components of the compositions and systems may be transiently expressed in the plant cell. In some examples, the compositions and systems may modify a target nucleic acid only when both the guide RNA and the component(s) of the system are present in a cell, such that genomic modification can further be controlled. As the expression of the component(s) of the system is transient, plants regenerated from such plant cells typically contain no foreign DNA. In certain examples, the component(s) of the system is stably expressed and the guide sequence is transiently expressed.

[0378] DNA and/or RNA (e.g., mRNA) may be introduced to plant cells for transient expression. In such cases, the introduced nucleic acid may be provided in sufficient quantity to modify the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions.

[0379] The transient expression may be achieved using suitable vectors. Exemplary vectors that may be used for transient expression include a pEAQ vector (may be tailored for Agrobacterium-mediated transient expression) and Cabbage Leaf Curl virus (CaLCuV), and vectors described in Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93; and Yin K et al., Scientific Reports volume 5, Article number: 14926 (2015).

[0380] Combinations of the different methods described above are also envisaged.

Translocation to and/or expression in specific plant organelles

[0381] The compositions and systems herein may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[0382] In some embodiments, it is envisaged that the compositions and systems are used to specifically modify chloroplast genes or to ensure expression in the chloroplast. The compositions and systems (e.g., component(s) of the system such as reverse transcriptases, Cas proteins, guide molecules, or their encoding polynucleotides) may be transformed,

compartmentalized, and/or targeted to the chloroplast. In an example, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[0383] Examples of methods of chloroplast transformation include Particle bombardment, PEG treatment, and microinjection, and the translocation of transformation cassettes from the nuclear genome to the plastid. In some examples, targeting of chloroplasts may be achieved by incorporating in chloroplast localization sequence, and/or the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the components of the compositions and systems. Additional examples of transforming, targeting and localization of chloroplasts include those described in WO2010061186, Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180, and US 20040142476, which are incorporated by reference herein in their entireties.

Exemplary applications in plants

[0384] The compositions, systems, and methods may be used to generate genetic variation(s) in a plant (e.g., crop) of interest. One or more, e.g., a library of, guide molecules targeting one or more locations in a genome may be provided and introduced into plant cells together with the component(s) of the system. For example, a collection of genome-scale point mutations and gene knock-outs can be generated. In some examples, the compositions, systems, and methods may be used to generate a plant part or plant from the cells so obtained and screening the cells for a trait of interest. The target genes may include both coding and non-coding regions. In some cases, the trait is stress tolerance and the method is a method for the generation of stress-tolerant crop varieties.

[0385] In some embodiments, the compositions, systems, and methods are used to modify endogenous genes or to modify their expression. The expression of the components may induce targeted modification of the genome, either by direct activity of the component(s) of the system and optionally introduction of template DNA, or by modification of genes targeted. The different strategies described herein above allow Cas-mediated targeted genome editing without requiring the introduction of the components into the plant genome.

[0386] In some cases, the modification may be performed without the permanent introduction into the genome of the plant of any foreign gene, including those encoding components, so as to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous.

Components which are transiently introduced into the plant cell are typically removed upon crossing.

[0387] For example, the modification may be performed by transient expression of the components of the compositions and systems. The transient expression may be performed by delivering the components of the compositions and systems with viral vectors, delivery into protoplasts, with the aid of particulate molecules such as nanoparticles or CPPs.

Generation of plants with desired traits

[0388] The compositions, systems, and methods herein may be used to introduce desired traits to plants. The approaches include introduction of one or more foreign genes to confer a trait of interest, editing or modulating endogenous genes to confer a trait of interest.

Agronomic traits

[0389] In some embodiments, crop plants can be improved by influencing specific plant traits. Examples of the traits include improved agronomic traits such as herbicide resistance, disease resistance, abiotic stress tolerance, high yield, and superior quality, pesticide-resistance, disease resistance, insect and nematode resistance, resistance against parasitic weeds, drought tolerance, nutritional value, stress tolerance, self-pollination voidance, forage digestibility biomass, and grain yield.

[0390] In some embodiments, genes that confer resistance to pests or diseases may be introduced to plants. In cases there are endogenous genes that confer such resistance in plants, their expression and function may be enhanced (e.g., by introducing extra copies, modifications that enhance expression and/or activity).

[0391] Examples of genes that confer resistance include plant disease resistance genes (e.g., Cf- 9, Pto, RSP2, SlDMR6-1), genes conferring resistance to a pest (e.g., those described in WO96/30517), Bacillus thuringiensis proteins, lectins, Vitamin-binding proteins (e.g., avidin), enzyme inhibitors (e.g., protease or proteinase inhibitors or amylase inhibitors), insect-specific hormones or pheromones (e.g., ecdysteroid or a juvenile hormone, variant thereof, a mimetic based thereon, or an antagonist or agonist thereof) or genes involved in the production and regulation of such hormone and pheromones, insect-specific peptides or neuropeptide, Insect-specific venom (e.g., produced by a snake, a wasp, etc., or analog thereof), Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity, Enzymes involved in the modification of biologically active molecule (e.g., a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a

chitinase and a glucanase, whether natural or synthetic), molecules that stimulates signal transduction, Viral-invasive proteins or a complex toxin derived therefrom, Developmental-arrestive proteins produced in nature by a pathogen or a parasite, a developmental-arrestive protein produced in nature by a plant, or any combination thereof.

[0392] The compositions, systems, and methods may be used to identify, screen, introduce or remove mutations or sequences lead to genetic variability that give rise to susceptibility to certain pathogens, e.g., host specific pathogens. Such approach may generate plants that are non-host resistance, e.g., the host and pathogen are incompatible or there can be partial resistance against all races of a pathogen, typically controlled by many genes and/or also complete resistance to some races of a pathogen but not to other races.

[0393] In some embodiments, compositions, systems, and methods may be used to modify genes involved in plant diseases. Such genes may be removed, inactivated, or otherwise regulated or modified. Examples of plant diseases include those described in [0045]-[0080] of US20140213619A1, which is incorporated by reference herein in its entirety.

[0394] In some embodiments, genes that confer resistance to herbicides may be introduced to plants. Examples of genes that confer resistance to herbicides include genes conferring resistance to herbicides that inhibit the growing point or meristem, such as an imidazolinone or a sulfonylurea, genes conferring glyphosate tolerance (e.g., resistance conferred by, e.g., mutant 5-enolpyruvylshikimate-3- phosphate synthase genes, aroA genes and glyphosate acetyl transferase (GAT) genes, respectively), or resistance to other phosphono compounds such as by glufosinate (phosphinothricin acetyl transferase (PAT) genes from Streptomyces species, including Streptomyces hygroscopicus and Streptomyces viridichromogenes), and to pyridinoxy or phenoxy proprionic acids and cyclohexones by ACCase inhibitor-encoding genes), genes conferring resistance to herbicides that inhibit photosynthesis (such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene), and glutathione S-transferase), genes encoding enzymes detoxifying the herbicide or a mutant glutamine synthase enzyme that is resistant to inhibition, genes encoding a detoxifying enzyme is an enzyme encoding a phosphinothricin acetyltransferase (such as the bar or pat protein from Streptomyces species), genes encoding hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, e.g., naturally occurring HPPD resistant enzymes, and genes encoding a mutated or chimeric HPPD enzyme.

[0395] In some embodiments, genes involved in Abiotic stress tolerance may be introduced to plants. Examples of genes include those capable of reducing the expression and/or the activity of poly(ADP-ribose) polymerase (PARP) gene, transgenes capable of reducing the expression and/or the activity of the PARG encoding genes, genes coding for a plant-functional

enzyme of the nicotineamide adenine dinucleotide salvage synthesis pathway including nicotinamidase, nicotinate phosphoribosyltransferase, nicotinic acid mononucleotide adenyl transferase, nicotinamide adenine dinucleotide synthetase or nicotine amide phosphorybosyltransferase, enzymes involved in carbohydrate biosynthesis, enzymes involved in the production of polyfructose (e.g., the inulin and levan-type), the production of alpha-1,6 branched alpha-1,4-glucans, the production of alternan, the production of hyaluronan.

[0396] In some embodiments, genes that improve drought resistance may be introduced to plants. Examples of genes Ubiquitin Protein Ligase protein (UPL) protein (UPL3), DR02, DR03, ABC transporter, and DREB1A.

Nutritionally improved plants

[0397] In some embodiments, the compositions, systems, and methods may be used to produce nutritionally improved plants. In some examples, such plants may provide functional foods, e.g., a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains. In certain examples, such plants may provide nutraceuticals foods, e.g., substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. The nutraceutical foods may be useful in the prevention and/or treatment of diseases in animals and humans, e.g., cancers, diabetes, cardiovascular disease, and hypertension.

[0398] An improved plant may naturally produce one or more desired compounds and the modification may enhance the level or activity or quality of the compounds. In some cases, the improved plant may not naturally produce the compound(s), while the modification enables the plant to produce such compound(s). In some cases, the compositions, systems, and methods used to modify the endogenous synthesis of these compounds indirectly, e.g. by modifying one or more transcription factors that controls the metabolism of this compound.

[0399] Examples of nutritionally improved plants include plants comprising modified protein quality, content and/or amino acid composition, essential amino acid contents, oils and fatty acids, carbohydrates, vitamins and carotenoids, functional secondary metabolites, and minerals. In some examples, the improved plants may comprise or produce compounds with health benefits. Examples of nutritionally improved plants include those described in Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939–953.

[0400] Examples of compounds that can be produced include carotenoids (e.g., α-Carotene or β-Carotene), lutein, lycopene, Zeaxanthin, Dietary fiber (e.g., insoluble fibers, β-Glucan, soluble fibers, fatty acids (e.g., ω-3 fatty acids, Conjugated linoleic acid, GLA), Flavonoids (e.g., Hydroxycinnamates, flavonols, catechins and tannins), Glucosinolates, indoles,

isothiocyanates (e.g., Sulforaphane), Phenolics (e.g., stilbenes, caffeic acid and ferulic acid, epicatechin), Plant stanols/sterols, Fructans, inulins, fructo-oligosaccharides, Saponins, Soybean proteins, Phytoestrogens (e.g., isoflavones, lignans), Sulfides and thiols such as diallyl sulphide, Allyl methyl trisulfide, dithiolthiones, Tannins, such as proanthocyanidins, or any combination thereof.

[0401] The compositions, systems, and methods may also be used to modify protein/starch functionality, shelf life, taste/aesthetics, fiber quality, and allergen, antinutrient, and toxin reduction traits.

[0402] Examples of genes and nucleic acids that can be modified to introduce the traits include stearyl-ACP desaturase, DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid, Tf RAP2.2 and its interacting partner SINAT2, Tf Dof1, and DOF Tf AtDof1.1 (OBP2).

Modification of polyploid plants

[0403] The compositions, systems, and methods may be used to modify polyploid plants. Polyploid plants carry duplicate copies of their genomes (e.g. as many as six, such as in wheat). In some cases, the compositions, systems, and methods may be can be multiplexed to affect all copies of a gene, or to target dozens of genes at once. For instance, the compositions, systems, and methods may be used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defenses against a disease. The modification may be simultaneous suppression the expression of the TaMLO-Al, TaMLO-Bl and TaMLO-Dl nucleic acid sequence in a wheat plant cell and regenerating a wheat plant therefrom, in order to ensure that the wheat plant is resistant to powdery mildew (e.g., as described in WO2015109752).

Regulation of fruit-ripening

[0404] The compositions, systems, and methods may be used to regulate ripening of fruits. Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it may render a fruit or vegetable inedible, which can bring significant losses to both farmers and consumers.

[0405] In some embodiments, the compositions, systems, and methods are used to reduce ethylene production. In some examples, the compositions, systems, and methods may be used to suppress the expression and/or activity of ACC synthase, insert a ACC deaminase gene or a functional fragment thereof, insert a SAM hydrolase gene or functional fragment thereof, suppress ACC oxidase gene expression

[0406] Alternatively or additionally, the compositions, systems, and methods may be used to modify ethylene receptors (e.g., suppressing ETR1) and/or Polygalacturonase (PG). Suppression of a gene may be achieved by introducing a mutation, an antisense sequence, and/or a truncated copy of the gene to the genome.

Increasing storage life of plants

[0407] In some embodiments, the compositions, systems, and methods are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. The modification may be in a gene that prevents the accumulation of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In particular embodiments, the methods provided herein are used to reduce or inhibit expression of the vacuolar invertase gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose.

Reducing allergens in plants

[0408] In some embodiments, the compositions, systems, and methods are used to generate plants with a reduced level of allergens, making them safer for consumers. To this end, the compositions, systems, and methods may be used to identify and modify (e.g., suppress) one or more genes responsible for the production of plant allergens. Examples of such genes include Lol p5, as well as those in peanuts, soybeans, lentils, peas, lupin, green beans, mung beans, such as those described in Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222), which is incorporated by reference herein in its entirety.

Generation of male sterile plants

[0409] The compositions, systems, and methods may be used to generate male sterile plants. Hybrid plants typically have advantageous agronomic traits compared to inbred plants. However, for self-pollinating plants, the generation of hybrids can be challenging. In different plant types (e.g., maize and rice), genes have been identified which are important for plant fertility, more particularly male fertility. Plants that are as such genetically altered can be used in hybrid breeding programs.

[0410] The compositions, systems, and methods may be used to modify genes involved male fertility, e.g., inactivating (such as by introducing mutations to) genes required for male fertility. Examples of the genes involved in male fertility include cytochrome P450-like gene (MS26) or the meganuclease gene (MS45), and those described in Wan X et al., Mol Plant. 2019 Mar 4;12(3):321-342; and Kim YJ, et al., Trends Plant Sci. 2018 Jan;23(1):53-65.

Increasing the fertility stage in plants

[0411] In some embodiments, the compositions, systems, and methods may be used to prolong the fertility stage of a plant such as of a rice. For instance, a rice fertility stage gene such as Ehd3 can be targeted in order to generate a mutation in the gene and plantlets can be selected for a prolonged regeneration plant fertility stage.

Production of early yield of products

[0412] In some embodiments, the compositions, systems, and methods may be used to produce early yield of the product. For example, flowering process may be modulated, e.g., by mutating flowering repressor gene such as SP5G. Examples of such approaches include those described in Soyk S, et al., Nat Genet. 2017 Jan;49(1):162-168.

Oil and biofuel production

[0413] The compositions, systems, and methods may be used to generate plants for oil and biofuel production. Biofuels include fuels made from plant and plant-derived resources. Biofuels may be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. Biofuels include bioethanol and biodiesel. Bioethanol can be produced by the sugar fermentation process of cellulose (starch), which may be derived from maize and sugar cane. Biodiesel can be produced from oil crops such as rapeseed, palm, and soybean. Biofuels can be used for transportation.

Generation of plants for production of vegetable oils and biofuels

[0414] The compositions, systems, and methods may be used to generate algae (e.g., diatom) and other plants (e.g., grapes) that express or overexpress high levels of oil or biofuels.

[0415] In some cases, the compositions, systems, and methods may be used to modify genes involved in the modification of the quantity of lipids and/or the quality of the lipids. Examples of such genes include those involved in the pathways of fatty acid synthesis, e.g., acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl_acyl- carrier protein synthase III, glycerol-3-phospate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phoshatidate phosphatase, fatty acid thioesterase such as palmitoyl protein thioesterase, or malic enzyme activities.

[0416] In further embodiments it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolization. Examples of genes include those involved in the activation of triacylglycerol and free fatty acids, β-oxidation of fatty acids, such as genes of acyl-CoA synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucomutase.

[0417] In some examples, algae may be modified for production of oil and biofuels, including fatty acids (e.g., fatty esters such as acid methyl esters (FAME) and fatty acid ethyl esters (FAEE)). Examples of methods of modifying microalgae include those described in Stovicek et al. Metab. Eng. Comm., 2015; 2:1; US 8945839; and WO 2015086795.

[0418] In some examples, one or more genes may be introduced (e.g., overexpressed) to the plants (e.g., algae) to produce oils and biofuels (e.g., fatty acids) from a carbon source (e.g., alcohol). Examples of the genes include genes encoding acyl-CoA synthases, ester synthases, thioesterases (e.g., tesA, 'tesA, tesB, fatB, fatB2, fatB3, fatAl, or fatA), acyl-CoA synthases (e.g., fadD, JadK, BH3103, pfl-4354, EAV15023, fadDl, fadD2, RPC_4074,fadDD35, fadDD22, faa39), ester synthases (e.g., synthase/acyl-CoA:diacylglycerl acyltransferase from *Simmondsia chinensis*, *Acinetobacter sp. ADP*, *Alcanivorax borkumensis*, *Pseudomonas aeruginosa*, *Fundibacter jadensis*, *Arabidopsis thaliana*, or *Alkaligenes eutrophus*, or variants thereof).

[0419] Additionally or alternatively, one or more genes in the plants (e.g., algae) may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. Examples of such genes include genes encoding acyl-CoA dehydrogenases (e.g., fade), outer membrane protein receptors, and transcriptional regulator (e.g., repressor) of fatty acid biosynthesis (e.g., fabR), pyruvate formate lyases (e.g., pflB), lactate dehydrogenases (e.g., IdhA).

Organic acid production

[0420] In some embodiments, plants may be modified to produce organic acids such as lactic acid. The plants may produce organic acids using sugars, pentose or hexose sugars. To this end, one or more genes may be introduced (e.g., and overexpressed) in the plants. An example of such genes include the LDH gene.

[0421] In some examples, one or more genes may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. The genes may include those encoding proteins involved an endogenous metabolic pathway which

produces a metabolite other than the organic acid of interest and/or wherein the endogenous metabolic pathway consumes the organic acid.

[0422] Examples of genes that can be modified or introduced include those encoding pyruvate decarboxylases (pdc), fumarate reductases, alcohol dehydrogenases (adh), acetaldehyde dehydrogenases, phosphoenolpyruvate carboxylases (ppc), D-lactate dehydrogenases (d-ldh), L-lactate dehydrogenases (l-ldh), lactate 2-monooxygenases, lactate dehydrogenase, cytochrome-dependent lactate dehydrogenases (e.g., cytochrome B2-dependent L-lactate dehydrogenases).

Enhancing plant properties for biofuel production

[0423] In some embodiments, the compositions, systems, and methods are used to alter the properties of the cell wall of plants to facilitate access by key hydrolyzing agents for a more efficient release of sugars for fermentation. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, lignin biosynthesis may be downregulated in the plant so as to increase fermentable carbohydrates.

[0424] In some examples, one or more lignin biosynthesis genes may be down regulated. Examples of such genes include 4-coumarate 3-hydroxylases (C3H), phenylalanine ammonia-lyases (PAL), cinnamate 4-hydroxylases (C4H), hydroxycinnamoyl transferases (HCT), caffeic acid O-methyltransferases (COMT), caffeoyl CoA 3-O-methyltransferases (CCoAOMT), ferulate 5- hydroxylases (F5H), cinnamyl alcohol dehydrogenases (CAD), cinnamoyl CoA-reductases (CCR), 4- coumarate-CoA ligases (4CL), monolignol-lignin-specific glycosyltransferases, and aldehyde dehydrogenases (ALDH), and those described in WO 2008064289.

[0425] In some examples, plant mass that produces lower level of acetic acid during fermentation may be reduced. To this end, genes involved in polysaccharide acetylation (e.g., Cas1L and those described in WO 2010096488) may be inactivated.

Other microorganisms for oils and biofuel production

[0426] In some embodiments, microorganisms other than plants may be used for production of oils and biofuels using the compositions, systems, and methods herein. Examples of the microorganisms include those of the genus of *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Synechococcus*, *Synechoystis*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophtora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophamonas*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*.

Plant cultures and regeneration

[0427] In some embodiments, the modified plants or plant cells may be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Examples of regeneration techniques include those relying on manipulation of certain phytohormones in a tissue culture growth medium, relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences, obtaining from cultured protoplasts, plant callus, explants, organs, pollens, embryos or parts thereof.

Detecting modifications in the plant genome- selectable markers

[0428] When the compositions, systems, and methods are used to modify a plant, suitable methods may be used to confirm and detect the modification made in the plant. In some examples, when a variety of modifications are made, one or more desired modifications or traits resulting from the modifications may be selected and detected. The detection and confirmation may be performed by biochemical and molecular biology techniques such as Southern analysis, PCR, Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR, enzymatic assays, ribozyme activity, gel electrophoresis, Western blot, immunoprecipitation, enzyme-linked immunoassays, *in situ* hybridization, enzyme staining, and immunostaining.

[0429] In some cases, one or more markers, such as selectable and detectable markers, may be introduced to the plants. Such markers may be used for selecting, monitoring, isolating cells and plants with desired modifications and traits. A selectable marker can confer positive or negative selection and is conditional or non-conditional on the presence of external substrates. Examples of such markers include genes and proteins that confer resistance to antibiotics, such as hygromycin (hpt) and kanamycin (nptII), and genes that confer resistance to herbicides, such as phosphinothricin (bar) and chlorosulfuron (als), enzyme capable of producing or processing a colored substances (e.g., the β-glucuronidase, luciferase, B or C1 genes).

Applications in fungi

[0430] The compositions, systems, and methods described herein can be used to perform efficient and cost effective gene or genome interrogation or editing or manipulation in fungi or fungal cells, such as yeast. The approaches and applications in plants may be applied to fungi as well.

[0431] A fungal cell may be any type of eukaryotic cell within the kingdom of fungi, such as phyla of *Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota,*

Glomeromycota, Microsporidia, and *Neocallimastigomycota*. Examples of fungi or fungal cells in include yeasts, molds, and filamentous fungi.

[0432] In some embodiments, the fungal cell is a yeast cell. A yeast cell refers to any fungal cell within the phyla *Ascomycota and Basidiomycota*. Examples of yeasts include budding yeast, fission yeast, and mold, *S. cerevisiae, Kluyveromyces marxianus, Issatchenkia orientalis, Candida spp*. (e.g., *Candida albicans*), *Yarrowia spp*. (e.g., *Yarrowia lipolytica*), *Pichia spp*. (e.g., *Pichia pastoris*), *Kluyveromyces spp*. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora spp*. (e.g., *Neurospora crassa*), *Fusarium spp*. (e.g., *Fusarium oxysporum*), and *Issatchenkia spp*. (e.g., *Issatchenkia orientalis*, *Pichia kudriavzevii* and *Candida acidothermophilum*).

[0433] In some embodiments, the fungal cell is a filamentous fungal cell, which grow in filaments, e.g., hyphae or mycelia. Examples of filamentous fungal cells include *Aspergillus spp*. (e.g., *Aspergillus niger*), *Trichoderma spp.* (e.g., *Trichoderma reesei*), *Rhizopus spp*. (e.g., *Rhizopus oryzae*), and *Mortierella spp*. (e.g., *Mortierella isabellina*).

[0434] In some embodiments, the fungal cell is of an industrial strain. Industrial strains include any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains include, without limitation, JAY270 and ATCC4124.

[0435] In some embodiments, the fungal cell is a polyploid cell whose genome is present in more than one copy. Polyploid cells include cells naturally found in a polyploid state, and cells that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may be a cell whose entire genome is polyploid, or a cell that is polyploid in a particular genomic locus of interest. In some examples, the abundance of guide RNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the composition and system described herein may take advantage of using certain fungal cell types.

[0436] In some embodiments, the fungal cell is a diploid cell, whose genome is present in two copies. Diploid cells include cells naturally found in a diploid state, and cells that have been induced to exist in a diploid state (e.g., through specific regulation, alteration,

inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest.

[0437] In some embodiments, the fungal cell is a haploid cell, whose genome is present in one copy. Haploid cells include cells naturally found in a haploid state, or cells that have been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0438] The compositions and systems, and nucleic acid encoding thereof may be introduced to fungi cells using the delivery systems and methods herein. Examples of delivery systems include lithium acetate treatment, bombardment, electroporation, and those described in Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395–403.

[0439] In some examples, a yeast expression vector (e.g., those with one or more regulatory elements) may be used. Examples of such vectors include a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

Biofuel and materials production by fungi

[0440] In some embodiments, the compositions, systems, and methods may be used for generating modified fungi for biofuel and material productions. For instance, the modified fungi for production of biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. Foreign genes required for biofuel production and synthesis may be introduced in to fungi In some examples, the genes may encode enzymes involved in the conversion of pyruvate to ethanol or another product of interest, degrade cellulose (e.g., cellulase), endogenous metabolic pathways which compete with the biofuel production pathway.

[0441] In some examples, the compositions, systems, and methods may be used for generating and/or selecting yeast strains with improved xylose or cellobiose utilization,

isoprenoid biosynthesis, and/or lactic acid production. One or more genes involved in the metabolism and synthesis of these compounds may be modified and/or introduced to yeast cells. Examples of the methods and genes include lactate dehydrogenase, PDC1 and PDC5, and those described in Ha, S.J., et al. (2011) Proc. Natl. Acad. Sci. USA 108(2):504-9 and Galazka, J.M., et al. (2010) Science 330(6000):84-6; Jakočiūnas T et al., Metab Eng. 2015 Mar;28:213-222; Stovicek V, et al., FEMS Yeast Res. 2017 Aug 1;17(5).

Improved plants and yeast cells

[0442] The present disclosure further provides improved plants and fungi. The improved and fungi may comprise one or more genes introduced, and/or one or more genes modified by the compositions, systems, and methods herein. The improved plants and fungi may have increased food or feed production (e.g., higher protein, carbohydrate, nutrient or vitamin levels), oil and biofuel production (e.g., methanol, ethanol), tolerance to pests, herbicides, drought, low or high temperatures, excessive water, etc.

[0443] The plants or fungi may have one or more parts that are improved, e.g., leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. The parts may be viable, nonviable, regeneratable, and/or non- regeneratable.

[0444] The improved plants and fungi may include gametes, seeds, embryos, either zygotic or somatic, progeny and/or hybrids of improved plants and fungi. The progeny may be a clone of the produced plant or fungi, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly plants.

Further applications in plants

[0445] Further applications of the compositions, systems, and methods on plants and fungi include visualization of genetic element dynamics (e.g., as described in Chen B, et al., Cell. 2013 Dec 19;155(7):1479-91), targeted gene disruption positive-selection *in vitro* and *in vivo* (as described in Malina A et al., Genes Dev. 2013 Dec 1;27(23):2602-14), epigenetic modification such as using fusion of component(s) of the system and histone-modifying enzymes (e.g., as described in Rusk N, Nat Methods. 2014 Jan;11(1):28), identifying transcription regulators (e.g., as described in Waldrip ZJ, Epigenetics. 2014 Sep;9(9):1207-11), anti-virus treatment for both RNA and DNA viruses (e.g., as described in Price AA, et al., Proc Natl Acad Sci U S A. 2015 May 12;112(19):6164-9; Ramanan V et al., Sci Rep. 2015 Jun 2;5:10833), alteration of genome complexity such as chromosome numbers (e.g., as described in Karimi-Ashtiyani R et al., Proc Natl Acad Sci U S A. 2015 Sep 8;112(36):11211-6; Anton T, et al., Nucleus. 2014 Mar-Apr;5(2):163-72), self-cleavage of the composition and system

for controlled inactivation/activation (e.g., as described Sugano SS et al., Plant Cell Physiol. 2014 Mar;55(3):475-81), multiplexed gene editing (as described in Kabadi AM et al., Nucleic Acids Res. 2014 Oct 29;42(19):e147), development of kits for multiplex genome editing (as described in Xing HL et al., BMC Plant Biol. 2014 Nov 29;14:327), starch production (as described in Hebelstrup KH et al., Front Plant Sci. 2015 Apr 23;6:247), targeting multiple genes in a family or pathway (e.g., as described in Ma X et al., Mol Plant. 2015 Aug;8(8):1274-84), regulation of non-coding genes and sequences (e.g., as described in Lowder LG, et al., Plant Physiol. 2015 Oct;169(2):971-85), editing genes in trees (e.g., as described in Belhaj K et al., Plant Methods. 2013 Oct 11;9(1):39; Harrison MM, et al., Genes Dev. 2014 Sep 1;28(17):1859-72; Zhou X et al., New Phytol. 2015 Oct;208(2):298-301), introduction of mutations for resistance to host-specific pathogens and pests.

[0446] Additional examples of modifications of plants and fungi that may be performed using the compositions, systems, and methods include those described in WO2016/099887, WO2016/025131, WO2016/073433, WO2017/066175, WO2017/100158, WO 2017/105991, WO2017/106414, WO2016/100272, WO2016/100571, WO 2016/100568, WO 2016/100562, and WO 2017/019867.

APPLICATIONS IN NON-HUMAN ANIMALS

[0447] The compositions, systems, and methods may be used to study and modify non-human animals, e.g., introducing desirable traits and disease resilience, treating diseases, facilitating breeding, etc. In some embodiments, the compositions, systems, and methods may be used to improve breeding and introducing desired traits, e.g., increasing the frequency of trait-associated alleles, introgression of alleles from other breeds/species without linkage drag, and creation of de novo favorable alleles. Genes and other genetic elements that can be targeted may be screened and identified. Examples of application and approaches include those described in Tait-Burkard C, et al., Livestock 2.0 - genome editing for fitter, healthier, and more productive farmed animals. Genome Biol. 2018 Nov 26;19(1):204; Lillico S, Agricultural applications of genome editing in farmed animals. Transgenic Res. 2019 Aug;28(Suppl 2):57-60; Houston RD, et al., Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020 Apr 16. doi: 10.1038/s41576-020-0227-y, which are incorporated herein by reference in their entireties. Applications described in other sections such as therapeutic, diagnostic, etc. can also be used on the animals herein.

[0448] The compositions, systems, and methods may be used on animals such as fish, amphibians, reptiles, mammals, and birds. The animals may be farm and agriculture animals, or pets. Examples of farm and agriculture animals include horses, goats, sheep, swine, cattle,

llamas, alpacas, and birds, e.g., chickens, turkeys, ducks, and geese. The animals may be a non-human primate, e.g., baboons, capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. Examples of pets include dogs, cats, horses, wolfs, rabbits, ferrets, gerbils, hamsters, chinchillas, fancy rats, guinea pigs, canaries, parakeets, and parrots.

[0449] In some embodiments, one or more genes may be introduced (e.g., overexpressed) in the animals to obtain or enhance one or more desired traits. Growth hormones, insulin-like growth factors (IGF-1) may be introduced to increase the growth of the animals, e.g., pigs or salmon (such as described in Pursel VG et al., J Reprod Fertil Suppl. 1990;40:235-45; Waltz E, Nature. 2017;548:148). Fat-1 gene (e.g., from C elegans) may be introduced for production of larger ratio of n-3 to n-6 fatty acids may be induced, e.g. in pigs (such as described in Li M, et al., Genetics. 2018;8:1747–54). Phytase (e.g., from E coli) xylanase (e.g., from Aspergillus niger), beta-glucanase (e.g., from bacillus lichenformis) may be introduced to reduce the environmental impact through phosphorous and nitrogen release reduction, e.g. in pigs (such as described in Golovan SP, et al., Nat Biotechnol. 2001;19:741–5; Zhang X et al., elife. 2018). shRNA decoy may be introduced to induce avian influenza resilience e.g. in chicken (such as described in Lyall et al., Science. 2011;331:223–6). Lysozyme or lysostaphin may be introduced to induce mastitis resilience e.g., in goat and cow (such as described in Maga EA et al., Foodborne Pathog Dis. 2006;3:384–92; Wall RJ, et al., Nat Biotechnol. 2005;23:445–51). Histone deacetylase such as HDAC6 may be introduced to induce PRRSV resilience, e.g., in pig (such as described in Lu T., et al., PLoS One. 2017;12:e0169317). CD163 may be modified (e.g., inactivated or removed) to introduce PRRSV resilience in pigs (such as described in Prather RS et al.., Sci Rep. 2017 Oct 17;7(1):13371). Similar approaches may be used to inhibit or remove viruses and bacteria (e.g., Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema) that may be transmitted from animals to humans.

[0450] In some embodiments, one or more genes may be modified or edited for disease resistance and production traits. Myostatin (e.g., GDF8) may be modified to increase muscle growth, e.g., in cow, sheep, goat, catfish, and pig (such as described in Crispo M et al., PLoS One. 2015;10:e0136690; Wang X, et al., Anim Genet. 2018;49:43–51; Khalil K, et al., Sci Rep. 2017;7:7301; Kang J-D, et al., RSC Adv. 2017;7:12541–9). Pc POLLED may be modified to induce horlessness, e.g., in cow (such as described in Carlson DF et al., Nat Biotechnol. 2016;34:479–81). KISS1R may be modified to induce boretaint (hormone release during

sexual maturity leading to undesired meat taste), e.g., in pigs. Dead end protein (dnd) may be modified to induce sterility, e.g., in salmon (such as described in Wargelius A, et al., Sci Rep. 2016;6:21284). Nano2 and DDX may be modified to induce sterility (e.g., in surrogate hosts), e.g., in pigs and chicken (such as described Park K-E, et al., Sci Rep. 2017;7:40176; Taylor L et al., Development. 2017;144:928–34). CD163 may be modified to induce PRRSV resistance, e.g., in pigs (such as described in Whitworth KM, et al., Nat Biotechnol. 2015;34:20–2). RELA may be modified to induce ASFV resilience, e.g., in pigs (such as described in Lillico SG, et al., Sci Rep. 2016;6:21645). CD18 may be modified to induce Mannheimia (Pasteurella) haemolytica resilience, e.g., in cows (such as described in Shanthalingam S, et al., roc Natl Acad Sci U S A. 2016;113:13186–90). NRAMP1 may be modified to induce tuberculosis resilience, e.g., in cows (such as described in Gao Y et al., Genome Biol. 2017;18:13). Endogenous retrovirus genes may be modified or removed for xenotransplantation such as described in Yang L, et al. Science. 2015;350:1101–4; Niu D et al., Science. 2017;357:1303–7). Negative regulators of muscle mass (e.g., Myostatin) may be modified (e.g., inactivated) to increase muscle mass, e.g., in dogs (as described in Zou Q et al., J Mol Cell Biol. 2015 Dec;7(6):580-3).

[0451] Animals such as pigs with severe combined immunodeficiency (SCID) may generated (e.g., by modifying RAG2) to provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development. Examples of methods and approaches include those described Lee K, et al., Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5; and Schomberg et al. FASEB Journal, April 2016; 30(1):Suppl 571.1.

[0452] SNPs in the animals may be modified. Examples of methods and approaches include those described Tan W. et al., Proc Natl Acad Sci U S A. 2013 Oct 8;110(41):16526-31; Mali P, et al., Science. 2013 Feb 15;339(6121):823-6.

[0453] Stem cells (e.g., induced pluripotent stem cells) may be modified and differentiated into desired progeny cells, e.g., as described in Heo YT et al., Stem Cells Dev. 2015 Feb 1;24(3):393-402.

[0454] Profile analysis (such as Igenity) may be performed on animals to screen and identify genetic variations related to economic traits. The genetic variations may be modified to introduce or improve the traits, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain.

MODELS OF GENETIC AND EPIGENETIC CONDITIONS

[0455] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a

model of mutations of interest or a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences associated with a disease, or a plant, animal or cell in which the expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[0456] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0457] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0458] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of components of the system; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0459] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the systems and methods herein on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0460] Several disease models have been specifically investigated. These include *de novo* autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are, of course, preferred, but serve to show the broad applicability of the invention across genes and corresponding models. An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0461] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0462] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E. coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[0463] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0464] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Patent No. 5,210,015.

[0465] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[0466] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for example, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized

on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Patent No. 5,445,934.

[0467] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, ß-galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0468] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphoimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[0469] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent:protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[0470] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during

stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agent:polypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and, hence, generating a detectable signal.

[0471] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[0472] The amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[0473] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0474] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example, anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Anti-phosphotyrosine antibodies are particularly useful in detecting proteins that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2α). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody

technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

[0475] In practicing the subject method, it may be desirable to discern the expression pattern of an protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[0476] An altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a signaling biochemical pathway will dependent on the biological activity and/or the signal transduction pathway that is under investigation. For example, where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays such as AlphaScreen™ (available from Perkin Elmer) and eTag™ assay (Chan-Hui, et al. (2003) Clinical Immunology 111: 162-174).

[0477] Where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPRTM (Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a millisecond.

[0478] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection,

optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection.

[0479] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA).

[0480] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

[0481] The target polynucleotide of the system herein can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. Further,

engineering of the PAM Interacting (PI) domain may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas, e.g. Cas9, genome engineering platform. Cas proteins, such as Cas9 proteins may be engineered to alter their PAM specificity, for example as described in Kleinstiver BP et al. *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592.

[0482] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

THERAPEUTIC APPLICATIONS

[0483] Also provided herein are methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject. Generally, the methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject can include modifying a polynucleotide in a subject or cell thereof using a composition, system, or component thereof described herein and/or include detecting a diseased or healthy polynucleotide in a subject or cell thereof using a composition, system, or component thereof described herein. In some embodiments, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism (e.g. bacterial or virus) within a subject or cell thereof. In some embodiments, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism or symbiotic organism within a subject. The composition, system, and components thereof can be used to develop models of diseases, states, or conditions. The composition, system, and components thereof can be used to detect a disease state or correction thereof, such as by a method of treatment or prevention

described herein. The composition, system, and components thereof can be used to screen and select cells that can be used, for example, as treatments or preventions described herein. The composition, system, and components thereof can be used to develop biologically active agents that can be used to modify one or more biologic functions or activities in a subject or a cell thereof.

[0484] In general, the method can include delivering a composition, system, and/or component thereof to a subject or cell thereof, or to an infectious or symbiotic organism by a suitable delivery technique and/or composition. Once administered the components can operate as described elsewhere herein to elicit a nucleic acid modification event. In some aspects, the nucleic acid modification event can occur at the genomic, epigenomic, and/or transcriptomic level. DNA and/or RNA cleavage, gene activation, and/or gene deactivation can occur. Additional features, uses, and advantages are described in greater detail below. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. In addition to treating and/or preventing a disease in a subject, the compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling *in vivo* (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions to establish cell lines and transgenic animals for optimization and screening purposes).

[0485] The composition, system, and components thereof described elsewhere herein can be used to treat and/or prevent a disease, such as a genetic and/or epigenetic disease, in a subject. The composition, system, and components thereof described elsewhere herein can be used to treat and/or prevent genetic infectious diseases in a subject, such as bacterial infections, viral infections, fungal infections, parasite infections, and combinations thereof. The composition, system, and components thereof described elsewhere herein can be used to modify the composition or profile of a microbiome in a subject, which can in turn modify the health status of the subject. The composition, system, described herein can be used to modify cells *ex vivo*, which can then be administered to the subject whereby the modified cells can treat or prevent a disease or symptom thereof. This is also referred to in some contexts as adoptive therapy. The composition, system, described herein can be used to treat mitochondrial diseases, where the mitochondrial disease etiology involves a mutation in the mitochondrial DNA.

[0486] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding one or more components of the composition, system, or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. The repair template may be a recombination template herein. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises one or more components of composition, system, complex or component thereof comprising multiple Cas effectors. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0487] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the Cas effector(s), advantageously encoding and expressing *in vivo* the remaining portions of the composition, system, (e.g., RNA, guides). A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression by transforming the subject with the systems or compositions herein. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0488] One or more components of the composition and system described herein can be included in a composition, such as a pharmaceutical composition, and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[0489] Thus, also described herein are methods of inducing one or more polynucleotide modifications in a eukaryotic or prokaryotic cell or component thereof (e.g. a mitochondria) of a subject, infectious organism, and/or organism of the microbiome of the subject. The modification can include the introduction, deletion, or substitution of one or more nucleotides

at a target sequence of a polynucleotide of one or more cell(s). The modification can occur *in vitro*, *ex vivo*, *in situ*, or *in vivo*.

[0490] In some embodiments, the method of treating or inhibiting a condition or a disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non-human organism can include manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus in a target sequence in a subject or a non-human subject in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition or disease is susceptible to treatment or inhibition by manipulation of the target sequence including providing treatment comprising delivering a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment.

[0491] Also provided herein is the use of the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment in *ex vivo* or *in vivo* gene or genome editing; or for use in *in vitro, ex vivo* or *in vivo* gene therapy. Also provided herein are particle delivery systems, non-viral delivery systems, and/or the virus particle of any one of the above embodiments or the cell of any one of the above embodiments used in the manufacture of a medicament for *in vitro, ex vivo* or *in vivo* gene or genome editing or for use in *in vitro, ex vivo* or *in vivo* gene therapy or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus associated with a disease or in a method of treating or inhibiting a condition or disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non- human organism.

[0492] In some embodiments, polynucleotide modification can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said polynucleotide of said cell(s). The modification can include the introduction, deletion, or substitution of at least 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence. The modification can include the introduction, deletion, or substitution of at least 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35,

40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, or 9900 to 10000 nucleotides at each target sequence of said cell(s).

[0493] In some embodiments, the modifications can include the introduction, deletion, or substitution of nucleotides at each target sequence of said cell(s) via nucleic acid components (e.g. guide(s) RNA(s) or sgRNA(s)), such as those mediated by a composition, system, or a component thereof described elsewhere herein. In some embodiments, the modifications can include the introduction, deletion, or substitution of nucleotides at a target or random sequence of said cell(s) via a composition, system, or technique.

[0494] In some embodiments, the composition, system, or component thereof can promote Non-Homologous End-Joining (NHEJ). In some embodiments, modification of a polynucleotide by a composition, system, or a component thereof, such as a diseased polynucleotide, can include NHEJ. In some embodiments, promotion of this repair pathway by the composition, system, or a component thereof can be used to target gene or polynucleotide specific knock-outs and/or knock-ins. In some embodiments, promotion of this repair pathway by the composition, system, or a component thereof can be used to generate NHEJ-mediated indels. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. The indel can range in size from 1-50 or more base pairs. In some embodiments thee indel can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,

46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, or 500 base pairs or more. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences.

[0495] In some embodiments, composition, system, mediated NHEJ can be used in the method to delete small sequence motifs. In some embodiments, composition, system, mediated NHEJ can be used in the method to generate NHEJ-mediate indels that can be targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of

a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp). In an embodiment, in which a guide RNA and Cas effector generate a double strand break for the purpose of inducing NHEJ-mediated indels, a guide RNA may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position). In an embodiment, in which two guide RNAs complexing with one or more Cas nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

[0496] For minimization of toxicity and off-target effect, it may be important to control the concentration of each components delivered. For example, optimal concentrations of Cas mRNA and guide RNA, and/or other functional domains or components can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. In some examples, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in International Patent Publication No. WO 2014/093622 (PCT/US2013/074667); or, via mutation.

[0497] In some embodiments, formation of system or complex results in cleavage, nicking, and/or another modification of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence.

[0498] In some embodiments, a method of modifying a target polynucleotide in a cell to treat or prevent a disease can include allowing a composition, system, or component thereof to bind to the target polynucleotide, e.g., to effect cleavage, nicking, or other modification as the composition, system, is capable of said target polynucleotide, thereby modifying the target polynucleotide, wherein the composition, system, or component thereof, complex with a guide sequence, and hybridize said guide sequence to a target sequence within the target polynucleotide, wherein said guide sequence is optionally linked to a tracr mate sequence, which in turn can hybridize to a tracr sequence. In some embodiments, modification can

include cleaving or nicking one or two strands at the location of the target sequence by one or more components of the composition, system, or component thereof.

[0499] The cleavage, nicking, or other modification capable of being performed by the composition, system, can modify transcription of a target polynucleotide. In some embodiments, modification of transcription can include decreasing transcription of a target polynucleotide. In some embodiments, modification can include increasing transcription of a target polynucleotide. In some embodiments, the method includes repairing said cleaved target polynucleotide by homologous recombination with a recombination template polynucleotide, wherein said repair results in a modification such as, but not limited to, an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said modification results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the modification imparted by the composition, system, or component thereof provides a transcript and/or protein that can correct a disease or a symptom thereof, including but not limited to, any of those described in greater detail elsewhere herein.

[0500] In some embodiments, the method of treating or preventing a disease can include delivering one or more vectors or vector systems to a cell, such as a eukaryotic or prokaryotic cell, wherein one or more vectors or vector systems include the composition, system, or component thereof. In some embodiments, the vector(s) or vector system(s) can be a viral vector or vector system, such as an AAV or lentiviral vector system, which are described in greater detail elsewhere herein. In some embodiments, the method of treating or preventing a disease can include delivering one or more viral particles, such as an AAV or lentiviral particle, containing the composition, system, or component thereof. In some embodiments, the viral particle has a tissue specific tropism. In some embodiments, the viral particle has a liver, muscle, eye, heart, pancreas, kidney, neuron, epithelial cell, endothelial cell, astrocyte, glial cell, immune cell, or red blood cell specific tropism.

[0501] It will be understood that the composition and system, such as the composition and system, for use in the methods as described herein, may be suitably used for any type of application known for composition, system, preferably in eukaryotes. In certain aspects, the application is therapeutic, preferably therapeutic in a eukaryote organism, such as including but not limited to animals (including human), plants, algae, fungi (including yeasts), etc. Alternatively, or in addition, in certain aspects, the application may involve accomplishing or inducing one or more particular traits or characteristics, such as genotypic and/or phenotypic traits or characteristics, as also described elsewhere herein.

Treating Diseases of the Circulatory System

[0502] In some embodiments, the composition, system, and/or component thereof described herein can be used to treat and/or prevent a circulatory system disease. In some embodiments the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) can be used to deliver the composition, system, and/or component thereof described herein to the blood. In some embodiments, the circulatory system disease can be treated by using a lentivirus to deliver the composition, system, described herein to modify hematopoietic stem cells (HSCs) in vivo or *ex vivo* (see e.g. Drakopoulou, "Review Article, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia," Stem Cells International, Volume 2011, Article ID 987980, 10 pages, doi:10.4061/2011/987980, which can be adapted for use with the composition, system, herein in view of the description herein). In some embodiments, the circulatory system disorder can be treated by correcting HSCs as to the disease using a composition, system, herein or a component thereof, wherein the composition, system, optionally includes a suitable HDR repair template (see e.g. Cavazzana, "Outcomes of Gene Therapy for β-Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced *Ex Vivo* with a Lentiviral βA-T87Q-Globin Vector."; Cavazzana-Calvo, "Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia", Nature 467, 318–322 (16 September 2010) doi:10.1038/nature09328; Nienhuis, "Development of Gene Therapy for Thalassemia, Cold Spring Harbor Perspectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, a lentiviral vector containing an engineered β-globin gene (βA-T87Q); and Xie et al., "Seamless gene correction of β-thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback" Genome Research gr.173427.114 (2014) www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press; Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011), which can be adapted for use with the composition, system, herein in view of the description herein). In some embodiments, iPSCs can be modified using a composition, system, described herein to correct a disease polynucleotide associated with a circulatory disease. In this regard, the teachings of Xu et al. (Sci Rep. 2015 Jul 9;5:12065. doi: 10.1038/srep12065) and Song et al. (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5) with respect to modifying iPSCs can be adapted for use in view of the description herein with the composition, system, described herein.

[0503] The term "Hematopoietic Stem Cell" or "HSC" refers broadly those cells considered to be an HSC, e.g., blood cells that give rise to all the other blood cells and are derived from mesoderm; located in the red bone marrow, which is contained in the core of most bones. HSCs herein may include cells having a phenotype of hematopoietic stem cells, identified by small size, lack of lineage (lin) markers, and markers that belong to the cluster of differentiation series, like: CD34, CD38, CD90, CD133, CD105, CD45, and also c-kit, - the receptor for stem cell factor. Hematopoietic stem cells are negative for the markers that are used for detection of lineage commitment, and are, thus, called Lin-; and, during their purification by FACS, a number of up to 14 different mature blood-lineage markers, e.g., CD13 & CD33 for myeloid, CD71 for erythroid, CD19 for B cells, CD61 for megakaryocytic, etc. for humans; and, B220 (murine CD45) for B cells, Mac-1 (CD11b/CD18) for monocytes, Gr-1 for Granulocytes, Ter119 for erythroid cells, Il7Ra, CD3, CD4, CD5, CD8 for T cells, etc. Mouse HSC markers: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, and Human HSC markers: CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, and lin-. HSCs are identified by markers. Hence in embodiments discussed herein, the HSCs can be CD34+ cells. HSCs can also be hematopoietic stem cells that are CD34-/CD38-. Stem cells that may lack c-kit on the cell surface that are considered in the art as HSCs, as well as CD133+ cells likewise considered HSCs in the art.

[0504] In some embodiments, the treatment or prevention for treating a circulatory system or blood disease can include modifying a human cord blood cell with any modification described herein. In some embodiments, the treatment or prevention for treating a circulatory system or blood disease can include modifying a granulocyte colony-stimulating factor-mobilized peripheral blood cell (mPB) with any modification described herein. In some embodiments, the human cord blood cell or mPB can be CD34+. In some embodiments, the cord blood cell(s) or mPB cell(s) modified can be autologous. In some embodiments, the cord blood cell(s) or mPB cell(s) can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, described herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g. Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein. The modified cord blood cell(s) or mPB cell(s) can be optionally expanded *in vitro*. The modified cord blood cell(s) or mPB cell(s) can be derived to a subject in need thereof using any suitable delivery technique.

[0505] The composition and system may be engineered to target genetic locus or loci in HSCs. In some embodiments, the components of the systems can be codon-optimized for a eukaryotic cell and especially a mammalian cell, e.g., a human cell, for instance, HSC, or iPSC and sgRNA targeting a locus or loci in HSC, such as circulatory disease, can be prepared. These may be delivered via particles. The particles may be formed by the components of the systems herein being admixed. The components mixture can be, for example, admixed with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol, whereby particles containing the components of the systems may be formed. The disclosure comprehends so making particles and particles from such a method as well as uses thereof. Particles suitable delivery of the systems in the context of blood or circulatory system or HSC delivery to the blood or circulatory system are described in greater detail elsewhere herein.

[0506] In some embodiments, after *ex vivo* modification the HSCs or iPCS can be expanded prior to administration to the subject. Expansion of HSCs can be via any suitable method such as that described by, Lee, "Improved *ex vivo* expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4." Blood. 2013 May 16;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 2013 Mar 21.

[0507] In some embodiments, the HSCs or iPSCs modified can be autologous. In some embodiments, the HSCs or iPSCs can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, described herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g. Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein.

Treating Neurological Diseases

[0508] In some embodiments, the compositions, systems, described herein can be used to treat diseases of the brain and CNS. Delivery options for the brain include encapsulation of the systems in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing the systems. For instance, Xia CF and Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human

insulin receptor using avidin-biotin technology." Mol Pharm. 2009 May-Jun;6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and *in vivo*, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed *in vivo* following an intravenous administration of the targeted siRNA, the teachings of which can be adapted for use with the compositions, systems, herein. In other embodiments, an artificial virus can be generated for CNS and/or brain delivery. See e.g. Zhang et al. (Mol Ther. 2003 Jan;7(1):11-8.)), the teachings of which can be adapted for use with the compositions, systems, herein.

Treating Hearing Diseases

[0509] In some embodiments, the composition and system described herein can be used to treat a hearing disease or hearing loss in one or both ears. Deafness is often caused by lost or damaged hair cells that cannot relay signals to auditory neurons. In such cases, cochlear implants may be used to respond to sound and transmit electrical signals to the nerve cells. But these neurons often degenerate and retract from the cochlea as fewer growth factors are released by impaired hair cells.

[0510] In some embodiments, the composition, system, or modified cells can be delivered to one or both ears for treating or preventing hearing disease or loss by any suitable method or technique. Suitable methods and techniques include, but are not limited to those set forth in US Patent Publication No. 20120328580 describes injection of a pharmaceutical composition into the ear (e.g., auricular administration), such as into the luminae of the cochlea (e.g., the Scala media, Sc vestibulae, and Sc tympani), e.g., using a syringe, e.g., a single-dose syringe. For example, one or more of the compounds described herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner ear; administration in situ, via a catheter or pump (see e.g. McKenna et al., (U.S. Patent Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639); administration in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear (see e.g. U.S. Patent Publication No. 2007/0093878, which provides an exemplary cochlear implant suitable for delivery of the compositions, systems, described herein to the ear). Such methods are routinely used in the art, for example, for the administration of steroids and antibiotics into human ears. Injection can be, for example, through the round window of the ear or through the cochlear capsule. Other inner ear administration methods are known in the art (see, e.g., Salt and Plontke, Drug Discovery

Today, 10:1299-1306, 2005). In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient during a surgical procedure. In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient without the need for a surgical procedure.

[0511] In general, the cell therapy methods described in US Patent Publication No. 20120328580 can be used to promote complete or partial differentiation of a cell to or towards a mature cell type of the inner ear (e.g., a hair cell) *in vitro*. Cells resulting from such methods can then be transplanted or implanted into a patient in need of such treatment. The cell culture methods required to practice these methods, including methods for identifying and selecting suitable cell types, methods for promoting complete or partial differentiation of selected cells, methods for identifying complete or partially differentiated cell types, and methods for implanting complete or partially differentiated cells are described below.

[0512] Cells suitable for use in the present disclosure include, but are not limited to, cells that are capable of differentiating completely or partially into a mature cell of the inner ear, e.g., a hair cell (e.g., an inner and/or outer hair cell), when contacted, e.g., *in vitro*, with one or more of the compounds described herein. Exemplary cells that are capable of differentiating into a hair cell include, but are not limited to stem cells (e.g., inner ear stem cells, adult stem cells, bone marrow derived stem cells, embryonic stem cells, mesenchymal stem cells, skin stem cells, iPS cells, and fat derived stem cells), progenitor cells (e.g., inner ear progenitor cells), support cells (e.g., Deiters' cells, pillar cells, inner phalangeal cells, tectal cells and Hensen's cells), and/or germ cells. The use of stem cells for the replacement of inner ear sensory cells is described in Li et al., (U.S. Patent Publication No. 2005/0287127) and Li et al., (U.S. Patent Application No. 11/953,797). The use of bone marrow derived stem cells for the replacement of inner ear sensory cells is described in Edge et al., PCT/US2007/084654. iPS cells are described, e.g., at Takahashi et al., Cell, Volume 131, Issue 5, Pages 861-872 (2007); Takahashi and Yamanaka, Cell 126, 663-76 (2006); Okita et al., Nature 448, 260-262 (2007); Yu, J. et al., Science 318(5858):1917-1920 (2007); Nakagawa et al., Nat. Biotechnol. 26:101-106 (2008); and Zaehres and Scholer, Cell 131(5):834-835 (2007). Such suitable cells can be identified by analyzing (e.g., qualitatively or quantitatively) the presence of one or more tissue specific genes. For example, gene expression can be detected by detecting the protein product of one or more tissue-specific genes. Protein detection techniques involve staining proteins (e.g., using cell extracts or whole cells) using antibodies against the appropriate antigen. In this case, the appropriate antigen is the protein product of the tissue-specific gene expression. Although, in principle, a first antibody (i.e., the antibody that binds the antigen) can be labeled,

it is more common (and improves the visualization) to use a second antibody directed against the first (e.g., an anti-IgG). This second antibody is conjugated either with fluorochromes, or appropriate enzymes for colorimetric reactions, or gold beads (for electron microscopy), or with the biotin-avidin system, so that the location of the primary antibody, and thus the antigen, can be recognized.

[0513] The composition and system may be delivered to the ear by direct application of pharmaceutical composition to the outer ear, with compositions modified from US Patent Publication No. 20110142917. In some embodiments the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[0514] In some embodiments, the compositions, systems, or components thereof and/or vectors or vector systems can be delivered to ear via a transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the nucleic acid-targeting system (see, e.g., Qi et al., Gene Therapy (2013), 1-9). About 40 μl of 10mM RNA may be contemplated as the dosage for administration to the ear.

[0515] According to Rejali et al. (Hear Res. 2007 Jun;228(1-2):180-7), cochlear implant function can be improved by good preservation of the spiral ganglion neurons, which are the target of electrical stimulation by the implant and brain derived neurotrophic factor (BDNF) has previously been shown to enhance spiral ganglion survival in experimentally deafened ears. Rejali et al. tested a modified design of the cochlear implant electrode that includes a coating of fibroblast cells transduced by a viral vector with a BDNF gene insert. To accomplish this type of *ex vivo* gene transfer, Rejali et al. transduced guinea pig fibroblasts with an adenovirus with a BDNF gene cassette insert, and determined that these cells secreted BDNF and then attached BDNF-secreting cells to the cochlear implant electrode via an agarose gel, and implanted the electrode in the scala tympani. Rejali et al. determined that the BDNF expressing electrodes were able to preserve significantly more spiral ganglion neurons in the basal turns of the cochlea after 48 days of implantation when compared to control electrodes and demonstrated the feasibility of combining cochlear implant therapy with *ex vivo* gene transfer for enhancing spiral ganglion neuron survival. Such a system may be applied to the nucleic acid-targeting system for delivery to the ear.

[0516] In some embodiments, the system set forth in Mukherjea et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) can be adapted for transtympanic administration of the composition, system, or component thereof to the ear. In some embodiments, a dosage of about 2 mg to about 4 mg of CRISPR Cas for administration to a human.

[0517] In some embodiments, the system set forth in [Jung et al. (Molecular Therapy, vol. 21 no. 4, 834–841 apr. 2013) can be adapted for vestibular epithelial delivery of the composition, system, or component thereof to the ear. In some embodiments, a dosage of about 1 to about 30 mg of CRISPR Cas for administration to a human.

Treating Diseases in Non-Dividing Cells

[0518] In some embodiments, the gene or transcript to be corrected is in a non-dividing cell. Exemplary non-dividing cells are muscle cells or neurons. Non-dividing (especially non-dividing, fully differentiated) cell types present issues for gene targeting or genome engineering, for example because homologous recombination (HR) is generally suppressed in the G1 cell-cycle phase. However, while studying the mechanisms by which cells control normal DNA repair systems, Durocher discovered a previously unknown switch that keeps HR "off" in non-dividing cells and devised a strategy to toggle this switch back on. Orthwein et al. (Daniel Durocher's lab at the Mount Sinai Hospital in Ottawa, Canada) recently reported (Nature 16142, published online 9 Dec 2015) have shown that the suppression of HR can be lifted and gene targeting successfully concluded in both kidney (293T) and osteosarcoma (U2OS) cells. Tumor suppressors, BRCA1, PALB2 and BRAC2 are known to promote DNA DSB repair by HR. They found that formation of a complex of BRCA1 with PALB2 - BRAC2 is governed by a ubiquitin site on PALB2, such that action on the site by an E3 ubiquitin ligase. This E3 ubiquitin ligase is composed of KEAP1 (a PALB2 -interacting protein) in complex with cullin-3 (CUL3)–RBX1. PALB2 ubiquitylation suppresses its interaction with BRCA1 and is counteracted by the deubiquitylase USP11, which is itself under cell cycle control. Restoration of the BRCA1–PALB2 interaction combined with the activation of DNA-end resection is sufficient to induce homologous recombination in G1, as measured by a number of methods including a CRISPR–Cas-based gene-targeting assay directed at USP11 or KEAP1 (expressed from a pX459 vector). However, when the BRCA1–PALB2 interaction was restored in resection-competent G1 cells using either KEAP1 depletion or expression of the PALB2-KR mutant, a robust increase in gene-targeting events was detected. These teachings can be adapted for and/or applied to the systems described herein.

[0519] Thus, reactivation of HR in cells, especially non-dividing, fully differentiated cell types is preferred, in some embodiments. In some embodiments, promotion of the BRCA1–PALB2 interaction is preferred in some embodiments. In some embodiments, the target ell is a non-dividing cell. In some embodiments, the target cell is a neuron or muscle cell. In some embodiments, the target cell is targeted *in vivo*. In some embodiments, the cell is in G1 and HR is suppressed. In some embodiments, use of KEAP1 depletion, for example inhibition of

expression of KEAP1 activity, is preferred. KEAP1 depletion may be achieved through siRNA, for example as shown in Orthwein et al. Alternatively, expression of the PALB2-KR mutant (lacking all eight Lys residues in the BRCA1-interaction domain is preferred, either in combination with KEAP1 depletion or alone. PALB2-KR interacts with BRCA1 irrespective of cell cycle position. Thus, promotion or restoration of the BRCA1-PALB2 interaction, especially in G1 cells, is preferred in some embodiments, especially where the target cells are non-dividing, or where removal and return (*ex vivo* gene targeting) is problematic, for example neuron or muscle cells. KEAP1 siRNA is available from ThermoFischer. In some embodiments, a BRCA1–PALB2 complex may be delivered to the G1 cell. In some embodiments, PALB2 deubiquitylation may be promoted for example by increased expression of the deubiquitylase USP11, so it is envisaged that a construct may be provided to promote or up-regulate expression or activity of the deubiquitylase USP11.

Treating Diseases of the Eye

[0520] In some embodiments, the disease to be treated is a disease that affects the eyes. Thus, in some embodiments, the composition, system, or component thereof described herein is delivered to one or both eyes.

[0521] The composition, system can be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[0522] In some embodiments, the condition to be treated or targeted is an eye disorder. In some embodiments, the eye disorder may include glaucoma. In some embodiments, the eye disorder includes a retinal degenerative disease. In some embodiments, the retinal degenerative disease is selected from Stargardt disease, Bardet-Biedl Syndrome, Best disease, Blue Cone Monochromacy, Choroidermia, Cone-rod dystrophy, Congenital Stationary Night Blindness, Enhanced S-Cone Syndrome, Juvenile X-Linked Retinoschisis, Leber Congenital Amaurosis, Malattia Leventinesse, Norrie Disease or X-linked Familial Exudative Vitreoretinopathy, Pattern Dystrophy, Sorsby Dystrophy, Usher Syndrome, Retinitis Pigmentosa, Achromatopsia or Macular dystrophies or degeneration, Retinitis Pigmentosa, Achromatopsia, and age related macular degeneration. In some embodiments, the retinal degenerative disease is Leber Congenital Amaurosis (LCA) or Retinitis Pigmentosa. Other exemplary eye diseases are described in greater detail elsewhere herein.

[0523] In some embodiments, the composition, system is delivered to the eye, optionally via intravitreal injection or subretinal injection. Intraocular injections may be performed with the aid of an operating microscope. For subretinal and intravitreal injections, eyes may be

prolapsed by gentle digital pressure and fundi visualized using a contact lens system consisting of a drop of a coupling medium solution on the cornea covered with a glass microscope slide coverslip. For subretinal injections, the tip of a 10-mm 34-gauge needle, mounted on a 5-µl Hamilton syringe may be advanced under direct visualization through the superior equatorial sclera tangentially towards the posterior pole until the aperture of the needle was visible in the subretinal space. Then, 2 µl of vector suspension may be injected to produce a superior bullous retinal detachment, thus confirming subretinal vector administration. This approach creates a self-sealing sclerotomy allowing the vector suspension to be retained in the subretinal space until it is absorbed by the RPE, usually within 48 h of the procedure. This procedure may be repeated in the inferior hemisphere to produce an inferior retinal detachment. This technique results in the exposure of approximately 70% of neurosensory retina and RPE to the vector suspension. For intravitreal injections, the needle tip may be advanced through the sclera 1 mm posterior to the corneoscleral limbus and 2 µl of vector suspension injected into the vitreous cavity. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. These vectors may be injected at titers of either $1.0–1.4 \times 10^{10}$ or $1.0–1.4 \times 10^{9}$ transducing units (TU)/ml.

[0524] In some embodiments, for administration to the eye, lentiviral vectors can be used. In some embodiments, the lentiviral vector is an equine infectious anemia virus (EIAV) vector. Exemplary EIAV vectors for eye delivery are described in Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845; Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012), which can be adapted for use with the composition, system, described herein. In some embodiments, the dosage can be 1.1×10^{5} transducing units per eye (TU/eye) in a total volume of 100 µl.

[0525] Other viral vectors can also be used for delivery to the eye, such as AAV vectors, such as those described in Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006), Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642–649 apr. 2011; Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)), which can be adapted for use with the composition, system, described herein. In some embodiments, the dose can range from about 10^{6} to $10^{9.5}$ particle units. In the context of the Millington-Ward AAV vectors, a dose of about 2×10^{11} to

about 6 x 10^{13} virus particles can be administered. In the context of Dalkara vectors, a dose of about 1 x 10^{15} to about 1 x 10^{16} vg/ml administered to a human.

[0526] In some embodiments, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering composition, system, to the eye. In this system, a single intravitreal administration of 3 μg of sd-rxRNA results in sequence-specific reduction of PPIB mRNA levels for 14 days. The sd-rxRNA® system may be applied to the nucleic acid-targeting system, contemplating a dose of about 3 to 20 mg of CRISPR administered to a human.

[0527] In other embodiments, the methods of US Patent Publication No. 20130183282, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the nucleic acid-targeting system.

[0528] In other embodiments, the methods of US Patent Publication No. 20130202678 for treating retinopathies and sight-threatening ophthalmologic disorders relating to delivering of the Puf-A gene (which is expressed in retinal ganglion and pigmented cells of eye tissues and displays a unique anti-apoptotic activity) to the sub-retinal or intravitreal space in the eye may be used or adapted. In particular, desirable targets are zgc:193933, prdm1a, spata2, tex10, rbb4, ddx3, zp2.2, Blimp-1 and HtrA2, all of which may be targeted by the composition, system.

[0529] Wu (Cell Stem Cell,13:659–62, 2013) designed a guide RNA that led Cas9to a single base pair mutation that causes cataracts in mice, where it induced DNA cleavage. Then using either the other wild-type allele or oligos given to the zygotes repair mechanisms corrected the sequence of the broken allele and corrected the cataract-causing genetic defect in mutant mouse. This approach can be adapted to and/or applied to the compositions, systems, described herein.

[0530] US Patent Publication No. 20120159653 describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeneration (MD), the teachings of which can be applied to and/or adapted for the compositions, systems, described herein.

[0531] One aspect of US Patent Publication No. 20120159653 relates to editing of any chromosomal sequences that encode proteins associated with MD which may be applied to the nucleic acid-targeting system.

[0532] Methods and target genes using the systems herein in treating eye disease also include gene therapy that need long coding sequence, e.g., USH2A and ABCA4, such as those described in Fry LE, et al., Int J Mol Sci. 2020 Jan 25;21(3):777.

Treating Muscle Diseases and Cardiovascular Diseases

[0533] In some embodiments, the composition, system can be used to treat and/or prevent a muscle disease and associated circulatory or cardiovascular disease or disorder. The present disclosure also contemplates delivering the composition, system, described herein to the heart. For the heart, a myocardium tropic adeno-associated virus (AAVM) is preferred, in particular AAVM41 which showed preferential gene transfer in the heart (see, e.g., Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). Administration may be systemic or local. A dosage of about 1-10 x 10^{14} vector genomes are contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492: 376 and Somasuntharam et al. (2013) Biomaterials 34: 7790, the teachings of which can be adapted for and/or applied to the compositions, systems, described herein.

[0534] For example, US Patent Publication No. 20110023139, the teachings of which can be adapted for and/or applied to the compositions, systems, described herein describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with cardiovascular disease. Cardiovascular diseases generally include high blood pressure, heart attacks, heart failure, and stroke and TIA. Any chromosomal sequence involved in cardiovascular disease or the protein encoded by any chromosomal sequence involved in cardiovascular disease may be utilized in the methods described in this disclosure. The cardiovascular-related proteins are typically selected based on an experimental association of the cardiovascular-related protein to the development of cardiovascular disease. For example, the production rate or circulating concentration of a cardiovascular-related protein may be elevated or depressed in a population having a cardiovascular disorder relative to a population lacking the cardiovascular disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the cardiovascular-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR). Exemplary chromosomal sequences can be found in Table 2.

[0535] The compositions, systems, herein can be used for treating diseases of the muscular system. The present disclosure also contemplates delivering the composition, system, described herein, effector protein systems, to muscle(s).

[0536] In some embodiments, the muscle disease to be treated is a muscle dystrophy such as DMD. In some embodiments, the composition, system, such as a system capable of RNA modification, described herein can be used to achieve exon skipping to achieve correction of the diseased gene. As used herein, the term "exon skipping" refers to the modification of pre-mRNA splicing by the targeting of splice donor and/or acceptor sites within a pre-mRNA with one or more complementary antisense oligonucleotide(s) (AONs). By blocking access of a spliceosome to one or more splice donor or acceptor site, an AON may prevent a splicing reaction thereby causing the deletion of one or more exons from a fully-processed mRNA. Exon skipping may be achieved in the nucleus during the maturation process of pre-mRNAs. In some examples, exon skipping may include the masking of key sequences involved in the splicing of targeted exons by using a composition, system, described herein capable of RNA modification. In some embodiments, exon skipping can be achieved in dystrophin mRNA. In some embodiments, the composition, system, can induce exon skipping at exon 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 45, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or any combination thereof of the dystrophin mRNA. In some embodiments, the composition, system, can induce exon skipping at exon 43, 44, 50, 51, 52, 55, or any combination thereof of the dystrophin mRNA. Mutations in these exons, can also be corrected using non-exon skipping polynucleotide modification methods.

[0537] In some embodiments, for treatment of a muscle disease, the method of Bortolanza et al. Molecular Therapy vol. 19 no. 11, 2055–2064 Nov. 2011) may be applied to an AAV expressing CRISPR Cas and injected into humans at a dosage of about 2×10^{15} or 2×10^{16} vg of vector. The teachings of Bortolanza et al., can be adapted for and/or applied to the compositions, systems, described herein.

[0538] In some embodiments, the method of Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881–887 May 2010) may be applied to an AAV expressing CRISPR Cas and injected into humans, for example, at a dosage of about 10^{14} to about 10^{15} vg of vector. The teachings of Dumonceaux described herein can be adapted for and/or applied to the compositions, systems, described herein.

[0539] In some embodiments, the method of Kinouchi et al. (Gene Therapy (2008) 15, 1126–1130) may be applied to CRISPR Cas systems described herein and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 μM solution into the muscle.

[0540] In some embodiments, the method of Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) can be adapted for and/or applied to the compositions, systems, herein and injected at a dose of about 15 to about 50 mg into the great saphenous vein of a human.

[0541] In some embodiments, the method comprises treating a sickle cell related disease, e.g., sickle cell trait, sickle cell disease such as sickle cell anemia, β-thalassaemia. For example, the method and system may be used to modify the genome of the sickle cell, e.g., by correcting one or more mutations of the β-globin gene. In the case of β-thalassaemia, sickle cell anemia can be corrected by modifying HSCs with the systems. The system allows the specific editing of the cell's genome by cutting its DNA and then letting it repair itself. The Cas protein is inserted and directed by a RNA guide to the mutated point and then it cuts the DNA at that point. Simultaneously, a healthy version of the sequence is inserted. This sequence is used by the cell's own repair system to fix the induced cut. In this way, the systems allow the correction of the mutation in the previously obtained stem cells. The methods and systems may be used to correct HSCs as to sickle cell anemia using a systems that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β-globin, advantageously non-sickling β-globin); specifically, the guide RNA can target mutation that give rise to sickle cell anemia, and the HDR can provide coding for proper expression of β-globin. A guide RNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of β-globin; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. The HDR template can provide for the HSC to express an engineered β-globin gene (e.g., βA-T87Q), or β-globin.

Treating Diseases of the Liver and Kidney

[0542] In some embodiments, the composition, system, or component thereof described herein can be used to treat a disease of the kidney or liver. Thus, in some embodiments, delivery of the system or component thereof described herein is to the liver or kidney.

[0543] Delivery strategies to induce cellular uptake of the therapeutic nucleic acid include physical force or vector systems such as viral-, lipid- or complex- based delivery, or nanocarriers. From the initial applications with less possible clinical relevance, when nucleic acids were addressed to renal cells with hydrodynamic high-pressure injection systemically, a wide range of gene therapeutic viral and non-viral carriers have been applied already to target posttranscriptional events in different animal kidney disease models *in vivo* (Csaba Révész and Péter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy

Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). Delivery methods to the kidney may include those in Yuan et al. (Am J Physiol Renal Physiol 295: F605–F617, 2008). The method of Yuang et al. may be applied to the system contemplating a 1-2 g subcutaneous injection of the systems conjugated with cholesterol to a human for delivery to the kidneys. In some embodiments, the method of Molitoris et al. (J Am Soc Nephrol 20: 1754–1764, 2009) can be adapted to the system of and a cumulative dose of 12- 20 mg/kg to a human can be used for delivery to the proximal tubule cells of the kidneys. In some embodiments, the methods of Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) can be adapted to the system and a dose of up to 25 mg/kg can be delivered via i.v. administration. In some embodiments, the method of Shimizu et al. (J Am Soc Nephrol 21: 622–633, 2010) can be adapted to the system and a dose of about of 10-20 μmol CRISPR Cas complexed with nanocarriers in about 1-2 liters of a physiologic fluid for i.p. administration can be used.

[0544] Other various delivery vehicles can be used to deliver the composition, system to the kidney such as viral, hydrodynamic, lipid, polymer nanoparticles, aptamers and various combinations thereof (see e.g. Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269); Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888); Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980); Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289); Q. Zhang et al., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13); Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331); Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103); Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693); Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157); Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8 pp. (1754-1764); Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226); Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101); Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251); Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108); Shimizu et al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633); Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737); Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212); Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637); Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178).

[0545] In some embodiments, delivery is to liver cells. In some embodiments, the liver cell is a hepatocyte. Delivery of the composition and system herein may be via viral vectors, especially AAV (and in particular AAV2/6) vectors. These can be administered by intravenous injection. A preferred target for the liver, whether *in vitro* or *in vivo*, is the albumin gene. This is a so-called 'safe harbor" as albumin is expressed at very high levels and so some reduction in the production of albumin following successful gene editing is tolerated. It is also preferred as the high levels of expression seen from the albumin promoter/enhancer allows for useful levels of correct or transgene production (from the inserted recombination template) to be achieved even if only a small fraction of hepatocytes are edited. See sites identified by Wechsler et al. (reported at the 57th Annual Meeting and Exposition of the American Society of Hematology - abstract available online at ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015) which can be adapted for use with the compositions, systems, herein.

[0546] Exemplary liver and kidney diseases that can be treated and/or prevented are described elsewhere herein.

Treating Epithelial and Lung Diseases

[0547] In some embodiments, the disease treated or prevented by the composition and system described herein can be a lung or epithelial disease. The compositions and systems described herein can be used for treating epithelial and/or lung diseases. The present disclosure also contemplates delivering the composition, system, described herein, to one or both lungs.

[0548] In some embodiments, a viral vector can be used to deliver the composition, system, or component thereof to the lungs. In some embodiments, the AAV is an AAV-1, AAV-2, AAV-5, AAV-6, and/or AAV-9 for delivery to the lungs. (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). In some embodiments, the MOI can vary from 1×10^3 to 4×10^5 vector genomes/cell. In some embodiments, the delivery vector can be an RSV vector as in Zamora et al. (Am J Respir Crit Care Med Vol 183. pp 531–538, 2011. The method of Zamora et al. may be applied to the nucleic acid-targeting system and an aerosolized the systems, for example with a dosage of 0.6 mg/kg, may be contemplated.

[0549] Subjects treated for a lung disease may for example receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a

promoter for Cas, U6 or H1 promoter for guide RNA),: A preferred arrangement is to use a CFTRdelta508 targeting guide, a repair template for deltaF508 mutation and a codon optimized Cas enzyme, with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs.

Treating Diseases of the Skin

[0550] The compositions and systems described herein can be used for the treatment of skin diseases. The present disclosure also contemplates delivering the composition and system, described herein, to the skin.

[0551] In some embodiments, delivery to the skin (intradermal delivery) of the composition, system, or component thereof can be via one or more microneedles or microneedle containing device. For example, in some embodiments the device and methods of Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) can be used and/or adapted to deliver the composition, system, described herein, for example, at a dosage of up to 300 μl of 0.1 mg/ml CRISPR-Cas system to the skin.

[0552] In some embodiments, the methods and techniques of Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442–446 Feb. 2010) can be used and/or adapted for delivery of a system described herein to the skin.

[0553] In some embodiments, the methods and techniques of Zheng et al. (PNAS, July 24, 2012, vol. 109, no. 30, 11975–11980) can be used and/or adapted for nanoparticle delivery of a system described herein to the skin. In some embodiments, as dosage of about 25 nM applied in a single application can achieve gene knockdown in the skin.

Treating Cancer

[0554] The compositions, systems, described herein can be used for the treatment of cancer. The present disclosure also contemplates delivering the composition, system, described herein, to a cancer cell. Also, as is described elsewhere herein the compositions, systems, can be used to modify an immune cell, such as a CAR or CAR T cell, which can then in turn be used to treat and/or prevent cancer. This is also described in International Patent Publication No. WO 2015/161276, the disclosure of which is hereby incorporated by reference and described herein below.

[0555] Target genes suitable for the treatment or prophylaxis of cancer can include those set forth in Tables 2 and 3. In some embodiments, target genes for cancer treatment and prevention can also include those described in International Patent Publication No. WO 2015/048577 the disclosure of which is hereby incorporated by reference and can be adapted for and/or applied to the composition, system, described herein.

Adoptive Cell Therapy

[0556] The compositions, systems, and components thereof described herein can be used to modify cells for an adoptive cell therapy. In an aspect, methods and compositions which involve editing a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with cancer immunotherapy are comprehended by adapting the composition, system. In some examples, the compositions, systems, and methods may be used to modify a stem cell (e.g., induced pluripotent cell) to derive modified natural killer cells, gamma delta T cells, and alpha beta T cells, which can be used for the adoptive cell therapy. In certain examples, the compositions, systems, and methods may be used to modify modified natural killer cells, gamma delta T cells, and alpha beta T cells.

[0557] As used herein, "ACT", "adoptive cell therapy" and "adoptive cell transfer" may be used interchangeably. In certain embodiments, Adoptive cell therapy (ACT) can refer to the transfer of cells to a patient with the goal of transferring the functionality and characteristics into the new host by engraftment of the cells (see, e.g., Mettananda et al., Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia, Nat Commun. 2017 Sep 4;8(1):424). As used herein, the term "engraft" or "engraftment" refers to the process of cell incorporation into a tissue of interest *in vivo* through contact with existing cells of the tissue. Adoptive cell therapy (ACT) can refer to the transfer of cells, most commonly immune-derived cells, back into the same patient or into a new recipient host with the goal of transferring the immunologic functionality and characteristics into the new host. If possible, use of autologous cells helps the recipient by minimizing GVHD issues. The adoptive transfer of autologous tumor infiltrating lymphocytes (TIL) (Zacharakis et al., (2018) Nat Med. 2018 Jun;24(6):724-730; Besser et al., (2010) Clin. Cancer Res 16 (9) 2646–55; Dudley et al., (2002) Science 298 (5594): 850–4; and Dudley et al., (2005) Journal of Clinical Oncology 23 (10): 2346–57.) or genetically re-directed peripheral blood mononuclear cells (Johnson et al., (2009) Blood 114 (3): 535–46; and Morgan et al., (2006) Science 314(5796) 126-9) has been used to successfully treat patients with advanced solid tumors, including melanoma, metastatic breast cancer and colorectal carcinoma, as well as patients with CD19-expressing hematologic malignancies (Kalos et al., (2011) Science Translational Medicine 3 (95): 95ra73). In certain embodiments, allogenic cells immune cells are transferred (see, e.g., Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266). As described further herein, allogenic cells can be edited to reduce alloreactivity and prevent graft-versus-host disease. Thus, use of allogenic cells allows

for cells to be obtained from healthy donors and prepared for use in patients as opposed to preparing autologous cells from a patient after diagnosis.

[0558] Aspects involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens or tumor specific neoantigens (see, e.g., Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62-68; Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127–144; and Rajasagi et al., 2014, Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood. 2014 Jul 17;124(3):453-62).

[0559] In certain embodiments, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: MR1 (see, e.g., Crowther, et al., 2020, Genome-wide CRISPR–Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1, Nature Immunology volume 21, pages178–185), B cell maturation antigen (BCMA) (see, e.g., Friedman et al., Effective Targeting of Multiple BCMA-Expressing Hematological Malignancies by Anti-BCMA CAR T Cells, Hum Gene Ther. 2018 Mar 8; Berdeja JG, et al. Durable clinical responses in heavily pretreated patients with relapsed/refractory multiple myeloma: updated results from a multicenter study of bb2121 anti-Bcma CAR T cell therapy. Blood. 2017;130:740; and Mouhieddine and Ghobrial, Immunotherapy in Multiple Myeloma: The Era of CAR T Cell Therapy, Hematologist, May-June 2018, Volume 15, issue 3); PSA (prostate-specific antigen); prostate-specific membrane antigen (PSMA); PSCA (Prostate stem cell antigen); Tyrosine-protein kinase transmembrane receptor ROR1; fibroblast activation protein (FAP); Tumor-associated glycoprotein 72 (TAG72); Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); Mesothelin; Human Epidermal growth factor Receptor 2 (ERBB2 (Her2/neu)); Prostase; Prostatic acid phosphatase (PAP); elongation factor 2 mutant (ELF2M); Insulin-like growth factor 1 receptor (IGF-1R); gplOO; BCR-ABL (breakpoint cluster region-Abelson); tyrosinase; New York esophageal squamous cell carcinoma 1 (NY-ESO-1); κ-light chain, LAGE (L antigen); MAGE (melanoma antigen); Melanoma-associated antigen 1 (MAGE-A1); MAGE A3; MAGE A6; legumain; Human papillomavirus (HPV) E6; HPV E7; prostein; survivin; PCTA1 (Galectin 8); Melan-A/MART-1; Ras mutant; TRP-1

(tyrosinase related protein 1, or gp75); Tyrosinase-related Protein 2 (TRP2); TRP-2/INT2 (TRP-2/intron 2); RAGE (renal antigen); receptor for advanced glycation end products 1 (RAGE1); Renal ubiquitous 1, 2 (RU1, RU2); intestinal carboxyl esterase (iCE); Heat shock protein 70-2 (HSP70-2) mutant; thyroid stimulating hormone receptor (TSHR); CD123; CD171; CD19; CD20; CD22; CD26; CD30; CD33; CD44v7/8 (cluster of differentiation 44, exons 7/8); CD53; CD92; CD100; CD148; CD150; CD200; CD261; CD262; CD362; CS-1 (CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1); ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); Tn antigen (Tn Ag); Fms-Like Tyrosine Kinase 3 (FLT3); CD38; CD138; CD44v6; B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2); Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); stage-specific embryonic antigen-4 (SSEA-4); Mucin 1, cell surface associated (MUC1); mucin 16 (MUC16); epidermal growth factor receptor (EGFR); epidermal growth factor receptor variant III (EGFRvIII); neural cell adhesion molecule (NCAM); carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); ephrin type-A receptor 2 (EphA2); Ephrin B2; Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); TGS5; high molecular weight-melanoma-associated antigen (HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor alpha; Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); G protein-coupled receptor class C group 5, member D (GPRC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide portion of globoH glycoceramide (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); CT (cancer/testis (antigen)); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; p53; p53 mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP);

ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; Cyclin D1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS); Squamous Cell Carcinoma Antigen Recognized By T Cells-1 or 3 (SART1, SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint-1, -2, -3 or -4 (SSX1, SSX2, SSX3, SSX4); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRL5); mouse double minute 2 homolog (MDM2); livin; alphafetoprotein (AFP); transmembrane activator and CAML Interactor (TACI); B-cell activating factor receptor (BAFF-R); V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS); immunoglobulin lambda-like polypeptide 1 (IGLL1); 707-AP (707 alanine proline); ART-4 (adenocarcinoma antigen recognized by T4 cells); BAGE (B antigen; b-catenin/m, b-catenin/mutated); CAMEL (CTL-recognized antigen on melanoma); CAP1 (carcinoembryonic antigen peptide 1); CASP-8 (caspase-8); CDC27m (cell-division cycle 27 mutated); CDK4/m (cycline-dependent kinase 4 mutated); Cyp-B (cyclophilin B); DAM (differentiation antigen melanoma); EGP-2 (epithelial glycoprotein 2); EGP-40 (epithelial glycoprotein 40); Erbb2, 3, 4 (erythroblastic leukemia viral oncogene homolog-2, -3, 4); FBP (folate binding protein); fAchR (Fetal acetylcholine receptor); G250 (glycoprotein 250); GAGE (G antigen); GnT-V (N-acetylglucosaminyltransferase V); HAGE (helicose antigen); ULA-A (human leukocyte antigen-A); HST2 (human signet ring tumor 2); KIAA0205; KDR (kinase insert domain receptor); LDLR/FUT (low density lipid receptor/GDP L-fucose: b-D-galactosidase 2-a-L fucosyltransferase); L1CAM (L1 cell adhesion molecule); MC1R (melanocortin 1 receptor); Myosin/m (myosin mutated); MUM-1, -2, -3 (melanoma ubiquitous mutated 1, 2, 3); NA88-A (NA cDNA clone of patient M88); KG2D (Natural killer group 2, member D) ligands; oncofetal antigen (h5T4); p190 minor bcr-abl (protein of 190KD bcr-abl); Pml/RARa (promyelocytic leukemia/retinoic acid receptor a); PRAME (preferentially expressed antigen of melanoma); SAGE (sarcoma antigen); TEL/AML1 (translocation Ets-family leukemia/acute

myeloid leukemia 1); TPI/m (triosephosphate isomerase mutated); CD70; and any combination thereof.

[0560] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-specific antigen (TSA).

[0561] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a neoantigen.

[0562] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-associated antigen (TAA).

[0563] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a universal tumor antigen. In certain preferred embodiments, the universal tumor antigen is selected from the group consisting of: a human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (Dl), and any combinations thereof.

[0564] In certain embodiments, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: CD19, BCMA, CD70, CLL-1, MAGE A3, MAGE A6, HPV E6, HPV E7, WT1, CD22, CD171, ROR1, MUC16, and SSX2. In certain preferred embodiments, the antigen may be CD19. For example, CD19 may be targeted in hematologic malignancies, such as in lymphomas, more particularly in B-cell lymphomas, such as without limitation in diffuse large B-cell lymphoma, primary mediastinal b-cell lymphoma, transformed follicular lymphoma, marginal zone lymphoma, mantle cell lymphoma, acute lymphoblastic leukemia including adult and pediatric ALL, non-Hodgkin lymphoma, indolent non-Hodgkin lymphoma, or chronic lymphocytic leukemia. For example, BCMA may be targeted in multiple myeloma or plasma cell leukemia (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic Chimeric Antigen Receptor T Cells Targeting B Cell Maturation Antigen). For example, CLL1 may be targeted in acute myeloid leukemia. For example, MAGE A3, MAGE A6, SSX2, and/or KRAS may be targeted in solid tumors. For example, HPV E6 and/or HPV E7 may be

targeted in cervical cancer or head and neck cancer. For example, WT1 may be targeted in acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), chronic myeloid leukemia (CML), non-small cell lung cancer, breast, pancreatic, ovarian or colorectal cancers, or mesothelioma. For example, CD22 may be targeted in B cell malignancies, including non-Hodgkin lymphoma, diffuse large B-cell lymphoma, or acute lymphoblastic leukemia. For example, CD171 may be targeted in neuroblastoma, glioblastoma, or lung, pancreatic, or ovarian cancers. For example, ROR1 may be targeted in ROR1+ malignancies, including non-small cell lung cancer, triple negative breast cancer, pancreatic cancer, prostate cancer, ALL, chronic lymphocytic leukemia, or mantle cell lymphoma. For example, MUC16 may be targeted in MUC16ecto+ epithelial ovarian, fallopian tube or primary peritoneal cancer. For example, CD70 may be targeted in both hematologic malignancies as well as in solid cancers such as renal cell carcinoma (RCC), gliomas (e.g., GBM), and head and neck cancers (HNSCC). CD70 is expressed in both hematologic malignancies as well as in solid cancers, while its expression in normal tissues is restricted to a subset of lymphoid cell types (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic CRISPR Engineered Anti-CD70 CAR-T Cells Demonstrate Potent Preclinical Activity Against Both Solid and Hematological Cancer Cells).

[0565] Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[0566] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO 9215322).

[0567] In general, CARs are comprised of an extracellular domain, a transmembrane domain, and an intracellular domain, wherein the extracellular domain comprises an antigen-binding domain that is specific for a predetermined target. While the antigen-binding domain of a CAR is often an antibody or antibody fragment (e.g., a single chain variable fragment, scFv), the binding domain is not particularly limited so long as it results in specific recognition

of a target. For example, in some embodiments, the antigen-binding domain may comprise a receptor, such that the CAR is capable of binding to the ligand of the receptor. Alternatively, the antigen-binding domain may comprise a ligand, such that the CAR is capable of binding the endogenous receptor of that ligand.

[0568] The antigen-binding domain of a CAR is generally separated from the transmembrane domain by a hinge or spacer. The spacer is also not particularly limited, and it is designed to provide the CAR with flexibility. For example, a spacer domain may comprise a portion of a human Fc domain, including a portion of the CH3 domain, or the hinge region of any immunoglobulin, such as IgA, IgD, IgE, IgG, or IgM, or variants thereof. Furthermore, the hinge region may be modified so as to prevent off-target binding by FcRs or other potential interfering objects. For example, the hinge may comprise an IgG4 Fc domain with or without a S228P, L235E, and/or N297Q mutation (according to Kabat numbering) in order to decrease binding to FcRs. Additional spacers/hinges include, but are not limited to, CD4, CD8, and CD28 hinge regions.

[0569] The transmembrane domain of a CAR may be derived either from a natural or from a synthetic source. Where the source is natural, the domain may be derived from any membrane bound or transmembrane protein. Transmembrane regions of particular use in this disclosure may be derived from CD8, CD28, CD3, CD45, CD4, CD5, CDS, CD9, CD 16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD137, CD 154, TCR. Alternatively, the transmembrane domain may be synthetic, in which case it will comprise predominantly hydrophobic residues such as leucine and valine. Preferably a triplet of phenylalanine, tryptophan and valine will be found at each end of a synthetic transmembrane domain. Optionally, a short oligo- or polypeptide linker, preferably between 2 and 10 amino acids in length may form the linkage between the transmembrane domain and the cytoplasmic signaling domain of the CAR. A glycine-serine doublet provides a particularly suitable linker.

[0570] Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3ζ; see U.S.

Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, CD2, CD7, LIGHT, LFA-1, NKG2C, B7-H3, CD30, CD40, PD-1, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO 2014/134165; PCT Publication No. WO 2012/079000). In certain embodiments, the primary signaling domain comprises a functional signaling domain of a protein selected from the group consisting of CD3 zeta, CD3 gamma, CD3 delta, CD3 epsilon, common FcR gamma (FCERIG), FcR beta (Fc Epsilon R1b), CD79a, CD79b, Fc gamma RIIa, DAP10, and DAP12. In certain preferred embodiments, the primary signaling domain comprises a functional signaling domain of CD3ζ or FcRγ. In certain embodiments, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, CDS, ICAM-1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), CD160, CD19, CD4, CD8 alpha, CD8 beta, IL2R beta, IL2R gamma, IL7R alpha, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CD11d, ITGAE, CD103, ITGAL, CD11a, LFA-1, ITGAM, CD11b, ITGAX, CD11c, ITGB1, CD29, ITGB2, CD18, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Lyl08), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, NKp44, NKp30, NKp46, and NKG2D. In certain embodiments, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: 4-1BB, CD27, and CD28. In certain embodiments, a chimeric antigen receptor may have the design as described in U.S. Patent No. 7,446,190, comprising an intracellular domain of CD3ζ chain (such as amino acid residues 52-163 of the human CD3 zeta chain, as shown in SEQ ID NO: 14 of US 7,446,190), a signaling region from CD28 and an antigen-binding element (or portion or domain; such as scFv). The CD28 portion, when between the zeta chain portion and the antigen-binding element, may suitably include the transmembrane and signaling domains of CD28 (such as amino acid residues 114-220 of SEQ ID NO: 10, full sequence shown in SEQ ID NO: 6 of US 7,446,190; these can include the following portion of CD28 as set forth in Genbank identifier NM_006139.

Alternatively, when the zeta sequence lies between the CD28 sequence and the antigen-binding element, intracellular domain of CD28 can be used alone (such as amino sequence set forth in SEQ ID NO: 9 of US 7,446,190). Hence, certain embodiments employ a CAR comprising (a) a zeta chain portion comprising the intracellular domain of human CD3ζ chain, (b) a costimulatory signaling region, and (c) an antigen-binding element (or portion or domain), wherein the costimulatory signaling region comprises the amino acid sequence encoded by SEQ ID NO: 6 of US 7,446,190.

[0571] Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native αβTCR, for example by antigen on professional antigen-presenting cells, with attendant costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects

[0572] By means of an example and without limitation, Kochenderfer et al., (2009) J Immunother. 32 (7): 689-702 described anti-CD19 chimeric antigen receptors (CAR). FMC63-28Z CAR contained a single chain variable region moiety (scFv) recognizing CD19 derived from the FMC63 mouse hybridoma (described in Nicholson et al., (1997) Molecular Immunology 34: 1157–1165), a portion of the human CD28 molecule, and the intracellular component of the human TCR-ζ molecule. FMC63-CD828BBZ CAR contained the FMC63 scFv, the hinge and transmembrane regions of the CD8 molecule, the cytoplasmic portions of CD28 and 4-1BB, and the cytoplasmic component of the TCR-ζ molecule. The exact sequence of the CD28 molecule included in the FMC63-28Z CAR corresponded to Genbank identifier NM_006139; the sequence included all amino acids starting with the amino acid sequence IEVMYPPPY (SEQ. I.D. No. 2) and continuing all the way to the carboxy-terminus of the protein. To encode the anti-CD19 scFv component of the vector, the authors designed a DNA sequence which was based on a portion of a previously published CAR (Cooper et al., (2003) Blood 101: 1637–1644). This sequence encoded the following components in frame from the 5' end to the 3' end: an XhoI site, the human granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor α-chain signal sequence, the FMC63 light chain variable region (as in Nicholson et al., supra), a linker peptide (as in Cooper et al., supra), the FMC63 heavy chain variable region (as in Nicholson et al., supra), and a NotI site. A plasmid encoding this sequence was digested with XhoI and NotI. To form the MSGV-FMC63-28Z retroviral vector, the XhoI and NotI-digested fragment encoding the FMC63 scFv was ligated into a second XhoI and NotI-digested fragment that encoded the MSGV retroviral backbone (as in Hughes et al.,

(2005) Human Gene Therapy 16: 457–472) as well as part of the extracellular portion of human CD28, the entire transmembrane and cytoplasmic portion of human CD28, and the cytoplasmic portion of the human TCR-ζ molecule (as in Maher et al., 2002) Nature Biotechnology 20: 70–75). The FMC63-28Z CAR is included in the KTE-C19 (axicabtagene ciloleucel) anti-CD19 CAR-T therapy product in development by Kite Pharma, Inc. for the treatment of inter alia patients with relapsed/refractory aggressive B-cell non-Hodgkin lymphoma (NHL). Accordingly, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may express the FMC63-28Z CAR as described by Kochenderfer et al. (supra). Hence, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element (or portion or domain; such as scFv) that specifically binds to an antigen, an intracellular signaling domain comprising an intracellular domain of a CD3ζ chain, and a costimulatory signaling region comprising a signaling domain of CD28. Preferably, the CD28 amino acid sequence is as set forth in Genbank identifier NM_006139 (sequence version 1, 2 or 3) starting with the amino acid sequence IEVMYPPPY and continuing all the way to the carboxy-terminus of the protein. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the anti-CD19 scFv as described by Kochenderfer et al. (supra).

[0573] Additional anti-CD19 CARs are further described in International Patent Publication No. WO 2015/187528. More particularly Example 1 and Table 1 of WO2015187528, incorporated by reference herein, demonstrate the generation of anti-CD19 CARs based on a fully human anti-CD19 monoclonal antibody (47G4, as described in US20100104509) and murine anti-CD19 monoclonal antibody (as described in Nicholson et al. and explained above). Various combinations of a signal sequence (human CD8-alpha or GM-CSF receptor), extracellular and transmembrane regions (human CD8-alpha) and intracellular T-cell signaling domains (CD28-CD3ζ; 4-1BB-CD3ζ; CD27-CD3ζ; CD28-CD27-CD3ζ, 4-1BB-CD27-CD3ζ; CD27-4-1BB-CD3ζ; CD28-CD27-FcεRI gamma chain; or CD28-FcεRI gamma chain) were disclosed. Hence, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element that specifically binds to an antigen, an extracellular and transmembrane region as set forth in Table 1 of WO2015187528 and an intracellular T-cell signaling domain as set forth in Table 1 of No. WO 2015/187528. Preferably, the antigen is CD19, more preferably the antigen-binding element is

an anti-CD19 scFv, even more preferably the mouse or human anti-CD19 scFv as described in Example 1 of. WO 2015/187528. In certain embodiments, the CAR comprises, consists essentially of or consists of an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, or SEQ ID NO: 13 as set forth in Table 1 of WO2015187528.

[0574] By means of an example and without limitation, chimeric antigen receptor that recognizes the CD70 antigen is described in WO2012058460A2 (see also, Park et al., CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma, Oral Oncol. 2018 Mar;78:145-150; and Jin et al., CD70, a novel target of CAR T-cell therapy for gliomas, Neuro Oncol. 2018 Jan 10;20(1):55-65). CD70 is expressed by diffuse large B-cell and follicular lymphoma and also by the malignant cells of Hodgkin's lymphoma, Waldenstrom's macroglobulinemia and multiple myeloma, and by HTLV-1- and EBV-associated malignancies. (Agathanggelou et al. Am.J.Pathol. 1995;147: 1152-1160; Hunter et al., Blood 2004; 104:4881. 26; Lens et al., J Immunol. 2005;174:6212-6219; Baba et al., J Virol. 2008;82:3843-3852.) In addition, CD70 is expressed by non-hematological malignancies such as renal cell carcinoma and glioblastoma. (Junker et al., J Urol. 2005;173:2150-2153; Chahlavi et al., Cancer Res 2005;65:5428-5438) Physiologically, CD70 expression is transient and restricted to a subset of highly activated T, B, and dendritic cells.

[0575] By means of an example and without limitation, chimeric antigen receptor that recognizes BCMA has been described (see, e.g., US20160046724A1; WO2016014789A2; WO2017211900A1; WO2015158671A1; US20180085444A1; WO2018028647A1; US20170283504A1; and WO2013154760A1).

[0576] In certain embodiments, the immune cell may, in addition to a CAR or exogenous TCR as described herein, further comprise a chimeric inhibitory receptor (inhibitory CAR) that specifically binds to a second target antigen and is capable of inducing an inhibitory or immunosuppressive or repressive signal to the cell upon recognition of the second target antigen. In certain embodiments, the chimeric inhibitory receptor comprises an extracellular antigen-binding element (or portion or domain) configured to specifically bind to a target antigen, a transmembrane domain, and an intracellular immunosuppressive or repressive signaling domain. In certain embodiments, the second target antigen is an antigen that is not expressed on the surface of a cancer cell or infected cell or the expression of which is downregulated on a cancer cell or an infected cell. In certain embodiments, the second target antigen is an MHC-class I molecule. In certain embodiments, the intracellular signaling domain

comprises a functional signaling portion of an immune checkpoint molecule, such as for example PD-1 or CTLA4. Advantageously, the inclusion of such inhibitory CAR reduces the chance of the engineered immune cells attacking non-target (e.g., non-cancer) tissues.

[0577] Alternatively, T-cells expressing CARs may be further modified to reduce or eliminate expression of endogenous TCRs in order to reduce off-target effects. Reduction or elimination of endogenous TCRs can reduce off-target effects and increase the effectiveness of the T cells (U.S. 9,181,527). T cells stably lacking expression of a functional TCR may be produced using a variety of approaches. T cells internalize, sort, and degrade the entire T cell receptor as a complex, with a half-life of about 10 hours in resting T cells and 3 hours in stimulated T cells (von Essen, M. et al. 2004. J. Immunol. 173:384-393). Proper functioning of the TCR complex requires the proper stoichiometric ratio of the proteins that compose the TCR complex. TCR function also requires two functioning TCR zeta proteins with ITAM motifs. The activation of the TCR upon engagement of its MHC-peptide ligand requires the engagement of several TCRs on the same T cell, which all must signal properly. Thus, if a TCR complex is destabilized with proteins that do not associate properly or cannot signal optimally, the T cell will not become activated sufficiently to begin a cellular response.

[0578] Accordingly, in some embodiments, TCR expression may eliminated using RNA interference (e.g., shRNA, siRNA, miRNA, etc.), CRISPR (e.g., without or without with functional domains), or other methods that target the nucleic acids encoding specific TCRs (e.g., TCR-α and TCR-β) and/or CD3 chains in primary T cells. By blocking expression of one or more of these proteins, the T cell will no longer produce one or more of the key components of the TCR complex, thereby destabilizing the TCR complex and preventing cell surface expression of a functional TCR.

[0579] In some instances, CAR may also comprise a switch mechanism for controlling expression and/or activation of the CAR. For example, a CAR may comprise an extracellular, transmembrane, and intracellular domain, in which the extracellular domain comprises a target-specific binding element that comprises a label, binding domain, or tag that is specific for a molecule other than the target antigen that is expressed on or by a target cell. In such embodiments, the specificity of the CAR is provided by a second construct that comprises a target antigen binding domain (e.g., an scFv or a bispecific antibody that is specific for both the target antigen and the label or tag on the CAR) and a domain that is recognized by or binds to the label, binding domain, or tag on the CAR. See, e.g., International Patent Publication Nos. WO 2013/044225, WO 2016/000304, WO 2015/057834, WO 2015/057852, and WO 2016/070061, US 9,233,125, and US 2016/0129109. In this way, a T-cell that expresses the

CAR can be administered to a subject, but the CAR cannot bind its target antigen until the second composition comprising an antigen-specific binding domain is administered.

[0580] Alternative switch mechanisms include CARs that require multimerization in order to activate their signaling function (see, e.g., US Patent Publication Nos. US 2015/0368342, US 2016/0175359, US 2015/0368360) and/or an exogenous signal, such as a small molecule drug (US 2016/0166613, Yung et al., Science, 2015), in order to elicit a T-cell response. Some CARs may also comprise a "suicide switch" to induce cell death of the CAR T-cells following treatment (Buddee et al., PLoS One, 2013) or to downregulate expression of the CAR following binding to the target antigen (International Patent Publication No. WO 2016/011210).

[0581] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[0582] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ-irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR+ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon-γ). CAR T cells of this kind may for example be used in animal models, for example to treat tumor xenografts.

[0583] In certain embodiments, ACT includes co-transferring CD4+ Th1 cells and CD8+ CTLs to induce a synergistic antitumor response (see, e.g., Li et al., Adoptive cell therapy with CD4+ T helper 1 cells and CD8+ cytotoxic T cells enhances complete rejection of an

established tumor, leading to generation of endogenous memory responses to non-targeted tumor epitopes. Clin Transl Immunology. 2017 Oct; 6(10): e160).

[0584] In certain embodiments, Th17 cells are transferred to a subject in need thereof. Th17 cells have been reported to directly eradicate melanoma tumors in mice to a greater extent than Th1 cells (Muranski P, et al., Tumor-specific Th17-polarized cells eradicate large established melanoma. Blood. 2008 Jul 15; 112(2):362-73; and Martin-Orozco N, et al., T helper 17 cells promote cytotoxic T cell activation in tumor immunity. Immunity. 2009 Nov 20; 31(5):787-98). Those studies involved an adoptive T cell transfer (ACT) therapy approach, which takes advantage of CD4+ T cells that express a TCR recognizing tyrosinase tumor antigen. Exploitation of the TCR leads to rapid expansion of Th17 populations to large numbers *ex vivo* for reinfusion into the autologous tumor-bearing hosts.

[0585] In certain embodiments, ACT may include autologous iPSC-based vaccines, such as irradiated iPSCs in autologous anti-tumor vaccines (see e.g., Kooreman, Nigel G. et al., Autologous iPSC-Based Vaccines Elicit Anti-tumor Responses *In Vivo*, Cell Stem Cell 22, 1–13, 2018, doi.org/10.1016/j.stem.2018.01.016).

[0586] Unlike T-cell receptors (TCRs) that are MHC restricted, CARs can potentially bind any cell surface-expressed antigen and can thus be more universally used to treat patients (see Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267). In certain embodiments, in the absence of endogenous T-cell infiltrate (e.g., due to aberrant antigen processing and presentation), which precludes the use of TIL therapy and immune checkpoint blockade, the transfer of CAR T-cells may be used to treat patients (see, e.g., Hinrichs CS, Rosenberg SA. Exploiting the curative potential of adoptive T-cell therapy for cancer. Immunol Rev (2014) 257(1):56–71. doi:10.1111/ imr.12132).

[0587] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoresponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction).

[0588] In certain embodiments, the treatment can be administered after lymphodepleting pretreatment in the form of chemotherapy (typically a combination of cyclophosphamide and fludarabine) or radiation therapy. Initial studies in ACT had short lived responses and the transferred cells did not persist *in vivo* for very long (Houot et al., T-cell-based immunotherapy:

adoptive cell transfer and checkpoint inhibition. Cancer Immunol Res (2015) 3(10):1115–22; and Kamta et al., Advancing Cancer Therapy with Present and Emerging Immuno-Oncology Approaches. Front. Oncol. (2017) 7:64). Immune suppressor cells like Tregs and MDSCs may attenuate the activity of transferred cells by outcompeting them for the necessary cytokines. Not being bound by a theory lymphodepleting pretreatment may eliminate the suppressor cells allowing the TILs to persist.

[0589] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment (e.g., glucocorticoid treatment). The cells or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. In certain embodiments, the immunosuppressive treatment provides for the selection and expansion of the immunoresponsive T cells within the patient.

[0590] In certain embodiments, the treatment can be administered before primary treatment (e.g., surgery or radiation therapy) to shrink a tumor before the primary treatment. In another embodiment, the treatment can be administered after primary treatment to remove any remaining cancer cells.

[0591] In certain embodiments, immunometabolic barriers can be targeted therapeutically prior to and/or during ACT to enhance responses to ACT or CAR T-cell therapy and to support endogenous immunity (see, e.g., Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267).

[0592] The administration of cells or population of cells, such as immune system cells or cell populations, such as more particularly immunoresponsive cells or cell populations, as disclosed herein may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, intrathecally, by intravenous or intralymphatic injection, or intraperitoneally. In some embodiments, the disclosed CARs may be delivered or administered into a cavity formed by the resection of tumor tissue (i.e. intracavity delivery) or directly into a tumor prior to resection (i.e. intratumoral delivery). In one embodiment, the cell compositions are administered by intravenous injection.

[0593] The administration of the cells or population of cells can consist of the administration of 104- 109 cells per kg body weight, preferably 105 to 106 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies

may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[0594] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[0595] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (see U.S. Patent Publication No. 20130071414; International Patent Publication WO 2011/146862; International Patent Publication WO 2014/011987; International Patent Publication WO 2013/040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[0596] In a further refinement of adoptive therapies, genome editing may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-

the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853; Ren et al., 2017, Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition, Clin Cancer Res. 2017 May 1;23(9):2255-2266. doi: 10.1158/1078-0432.CCR-16-1300. Epub 2016 Nov 4; Qasim et al., 2017, Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells, Sci Transl Med. 2017 Jan 25;9(374); Legut, et al., 2018, CRISPR-mediated TCR replacement generates superior anticancer transgenic T cells. Blood, 131(3), 311-322; and Georgiadis et al., Long Terminal Repeat CRISPR-CAR-Coupled "Universal" T Cells Mediate Potent Anti-leukemic Effects, Molecular Therapy, In Press, Corrected Proof, Available online 6 March 2018). Cells may be edited using any system and method of use thereof as described herein. The composition and systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited *ex vivo* and transferred to a subject in need thereof. Immunoresponsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed for example to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell (e.g. TRAC locus); to eliminate potential alloreactive T-cell receptors (TCR) or to prevent inappropriate pairing between endogenous and exogenous TCR chains, such as to knock-out or knock-down expression of an endogenous TCR in a cell; to disrupt the target of a chemotherapeutic agent in a cell; to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell; to knock-out or knock-down expression of other gene or genes in a cell, the reduced expression or lack of expression of which can enhance the efficacy of adoptive therapies using the cell; to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR; to knock-out or knock-down expression of one or more MHC constituent proteins in a cell; to activate a T cell; to modulate cells such that the cells are resistant to exhaustion or dysfunction; and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see International Patent Publication Nos. WO 2013/176915, WO 2014/059173, WO 2014/172606, WO 2014/184744, and WO 2014/191128).

[0597] In certain embodiments, editing may result in inactivation of a gene. By inactivating a gene, it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes

to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art. In certain embodiments, homology directed repair (HDR) is used to concurrently inactivate a gene (e.g., TRAC) and insert an endogenous TCR or CAR into the inactivated locus.

[0598] Hence, in certain embodiments, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell. Conventionally, nucleic acid molecules encoding CARs or TCRs are transfected or transduced to cells using randomly integrating vectors, which, depending on the site of integration, may lead to clonal expansion, oncogenic transformation, variegated transgene expression and/or transcriptional silencing of the transgene. Directing of transgene(s) to a specific locus in a cell can minimize or avoid such risks and advantageously provide for uniform expression of the transgene(s) by the cells. Without limitation, suitable 'safe harbor' loci for directed transgene integration include CCR5 or AAVS1. Homology-directed repair (HDR) strategies are known and described elsewhere in this specification allowing to insert transgenes into desired loci (e.g., TRAC locus).

[0599] Further suitable loci for insertion of transgenes, in particular CAR or exogenous TCR transgenes, include without limitation loci comprising genes coding for constituents of endogenous T-cell receptor, such as T-cell receptor alpha locus (TRA) or T-cell receptor beta locus (TRB), for example T-cell receptor alpha constant (TRAC) locus, T-cell receptor beta constant 1 (TRBC1) locus or T-cell receptor beta constant 2 (TRBC1) locus. Advantageously, insertion of a transgene into such locus can simultaneously achieve expression of the transgene, potentially controlled by the endogenous promoter, and knock-out expression of the endogenous TCR. This approach has been exemplified in Eyquem et al., (2017) Nature 543: 113-117, wherein the authors used CRISPR/Cas9 gene editing to knock-in a DNA molecule encoding a CD19-specific CAR into the TRAC locus downstream of the endogenous promoter; the CAR-T cells obtained by CRISPR were significantly superior in terms of reduced tonic CAR signaling and exhaustion.

[0600] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β, which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of

the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the potential development of graft versus host disease (GVHD). The inactivation of TCRα or TCRβ can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[0601] Hence, in certain embodiments, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous TCR in a cell. For example, NHEJ-based or HDR-based gene editing approaches can be employed to disrupt the endogenous TCR alpha and/or beta chain genes. For example, gene editing system or systems can be designed to target a sequence found within the TCR beta chain conserved between the beta 1 and beta 2 constant region genes (TRBC1 and TRBC2) and/or to target the constant region of the TCR alpha chain (TRAC) gene.

[0602] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present disclosure further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α-chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic

acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present disclosure allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[0603] In certain embodiments, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell. Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[0604] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[0605] International Patent Publication No. WO 2014/172606 relates to the use of MT1 and/or MT2 inhibitors to increase proliferation and/or activity of exhausted CD8+ T-cells and to decrease CD8+ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8+ immune cells). In certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[0606] In certain embodiments, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1, TIM-3, CEACAM-1, CEACAM-3, or CEACAM-5. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[0607] By means of an example and without limitation, International Patent Publication No. WO 2016/196388 concerns an engineered T cell comprising (a) a genetically engineered antigen receptor that specifically binds to an antigen, which receptor may be a CAR; and (b) a disrupted gene encoding a PD-L1, an agent for disruption of a gene encoding a PD- L1, and/or disruption of a gene encoding PD-L1, wherein the disruption of the gene may be mediated by a gene editing nuclease, a zinc finger nuclease (ZFN), CRISPR/Cas9 and/or TALEN. WO2015142675 relates to immune effector cells comprising a CAR in combination with an agent (such as the composition or system herein) that increases the efficacy of the immune effector cells in the treatment of cancer, wherein the agent may inhibit an immune inhibitory molecule, such as PD1, PD-L1, CTLA-4, TIM-3, LAG-3, VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, TGFR beta, CEACAM-1, CEACAM-3, or CEACAM-5. Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas9 mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[0608] In certain embodiments, cells may be engineered to express a CAR, wherein expression and/or function of methylcytosine dioxygenase genes (TET1, TET2 and/or TET3) in the cells has been reduced or eliminated, (such as the composition or system herein) (for example, as described in WO201704916).

[0609] In certain embodiments, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene

encoding an antigen targeted by an exogenous CAR or TCR, thereby reducing the likelihood of targeting of the engineered cells. In certain embodiments, the targeted antigen may be one or more antigen selected from the group consisting of CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, CD362, human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (D1), B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), and B-cell activating factor receptor (BAFF-R) (for example, as described in International Patent Publication Nos. WO 2016/011210 and WO 2017/011804).

[0610] In certain embodiments, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of one or more MHC constituent proteins, such as one or more HLA proteins and/or beta-2 microglobulin (B2M), in a cell, whereby rejection of non-autologous (e.g., allogeneic) cells by the recipient's immune system can be reduced or avoided. In preferred embodiments, one or more HLA class I proteins, such as HLA-A, B and/or C, and/or B2M may be knocked-out or knocked-down. Preferably, B2M may be knocked-out or knocked-down. By means of an example, Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[0611] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα, BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ, B2M and TCRα, B2M and TCRβ.

[0612] In certain embodiments, a cell may be multiplied edited (multiplex genome editing) as taught herein to (1) knock-out or knock-down expression of an endogenous TCR (for example, TRBC1, TRBC2 and/or TRAC), (2) knock-out or knock-down expression of an immune checkpoint protein or receptor (for example PD1, PD-L1 and/or CTLA4); and (3)

knock-out or knock-down expression of one or more MHC constituent proteins (for example, HLA-A, B and/or C, and/or B2M, preferably B2M).

[0613] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patent Nos. 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded *in vitro* or *in vivo*.

[0614] Immune cells may be obtained using any method known in the art. In one embodiment, allogenic T cells may be obtained from healthy subjects. In one embodiment T cells that have infiltrated a tumor are isolated. T cells may be removed during surgery. T cells may be isolated after removal of tumor tissue by biopsy. T cells may be isolated by any means known in the art. In one embodiment, T cells are obtained by apheresis. In one embodiment, the method may comprise obtaining a bulk population of T cells from a tumor sample by any suitable method known in the art. For example, a bulk population of T cells can be obtained from a tumor sample by dissociating the tumor sample into a cell suspension from which specific cell populations can be selected. Suitable methods of obtaining a bulk population of T cells may include, but are not limited to, any one or more of mechanically dissociating (e.g., mincing) the tumor, enzymatically dissociating (e.g., digesting) the tumor, and aspiration (e.g., as with a needle).

[0615] The bulk population of T cells obtained from a tumor sample may comprise any suitable type of T cell. Preferably, the bulk population of T cells obtained from a tumor sample comprises tumor infiltrating lymphocytes (TILs).

[0616] The tumor sample may be obtained from any mammal. Unless stated otherwise, as used herein, the term "mammal" refers to any mammal including, but not limited to, mammals of the order Logomorpha, such as rabbits; the order Carnivora, including Felines (cats) and Canines (dogs); the order Artiodactyla, including Bovines (cows) and Swines (pigs); or of the order Perssodactyla, including Equines (horses). The mammals may be non-human primates, e.g., of the order Primates, Ceboids, or Simoids (monkeys) or of the order Anthropoids (humans and apes). In some embodiments, the mammal may be a mammal of the order Rodentia, such as mice and hamsters. Preferably, the mammal is a non-human primate or a human. An especially preferred mammal is the human.

[0617] T cells can be obtained from a number of sources, including peripheral blood mononuclear cells (PBMC), bone marrow, lymph node tissue, spleen tissue, and tumors. In certain embodiments of the present disclosure, T cells can be obtained from a unit of blood

collected from a subject using any number of techniques known to the skilled artisan, such as Ficoll separation. In one preferred embodiment, cells from the circulating blood of an individual are obtained by apheresis or leukapheresis. The apheresis product typically contains lymphocytes, including T cells, monocytes, granulocytes, B cells, other nucleated white blood cells, red blood cells, and platelets. In one embodiment, the cells collected by apheresis may be washed to remove the plasma fraction and to place the cells in an appropriate buffer or media for subsequent processing steps. In one embodiment, the cells are washed with phosphate buffered saline (PBS). In an alternative embodiment, the wash solution lacks calcium and may lack magnesium or may lack many if not all divalent cations. Initial activation steps in the absence of calcium lead to magnified activation. As those of ordinary skill in the art would readily appreciate a washing step may be accomplished by methods known to those in the art, such as by using a semi-automated "flow-through" centrifuge (for example, the Cobe 2991 cell processor) according to the manufacturer's instructions. After washing, the cells may be resuspended in a variety of biocompatible buffers, such as, for example, Ca-free, Mg-free PBS. Alternatively, the undesirable components of the apheresis sample may be removed and the cells directly resuspended in culture media.

[0618] In another embodiment, T cells are isolated from peripheral blood lymphocytes by lysing the red blood cells and depleting the monocytes, for example, by centrifugation through a PERCOLL™ gradient. A specific subpopulation of T cells, such as CD28+, CD4+, CDC, CD45RA+, and CD45RO+ T cells, can be further isolated by positive or negative selection techniques. For example, in one preferred embodiment, T cells are isolated by incubation with anti-CD3/anti-CD28 (i.e., 3×28)-conjugated beads, such as DYNABEADS® M-450 CD3/CD28 T, or XCYTE DYNABEADS™ for a time period sufficient for positive selection of the desired T cells. In one embodiment, the time period is about 30 minutes. In a further embodiment, the time period ranges from 30 minutes to 36 hours or longer and all integer values there between. In a further embodiment, the time period is at least 1, 2, 3, 4, 5, or 6 hours. In yet another preferred embodiment, the time period is 10 to 24 hours. In one preferred embodiment, the incubation time period is 24 hours. For isolation of T cells from patients with leukemia, use of longer incubation times, such as 24 hours, can increase cell yield. Longer incubation times may be used to isolate T cells in any situation where there are few T cells as compared to other cell types, such in isolating tumor infiltrating lymphocytes (TIL) from tumor tissue or from immunocompromised individuals. Further, use of longer incubation times can increase the efficiency of capture of CD8+ T cells.

[0619] Enrichment of a T cell population by negative selection can be accomplished with a combination of antibodies directed to surface markers unique to the negatively selected cells. A preferred method is cell sorting and/or selection via negative magnetic immunoadherence or flow cytometry that uses a cocktail of monoclonal antibodies directed to cell surface markers present on the cells negatively selected. For example, to enrich for CD4+ cells by negative selection, a monoclonal antibody cocktail typically includes antibodies to CD14, CD20, CD11b, CD16, HLA-DR, and CD8.

[0620] Further, monocyte populations (e.g., CD14+ cells) may be depleted from blood preparations by a variety of methodologies, including anti-CD14 coated beads or columns, or utilization of the phagocytotic activity of these cells to facilitate removal. Accordingly, in one embodiment, paramagnetic particles of a size sufficient to be engulfed by phagocytotic monocytes is used. In certain embodiments, the paramagnetic particles are commercially available beads, for example, those produced by Life Technologies under the trade name Dynabeads™. In one embodiment, other non-specific cells are removed by coating the paramagnetic particles with "irrelevant" proteins (e.g., serum proteins or antibodies). Irrelevant proteins and antibodies include those proteins and antibodies or fragments thereof that do not specifically target the T cells to be isolated. In certain embodiments, the irrelevant beads include beads coated with sheep anti-mouse antibodies, goat anti-mouse antibodies, and human serum albumin.

[0621] In brief, such depletion of monocytes is performed by preincubating T cells isolated from whole blood, apheresed peripheral blood, or tumors with one or more varieties of irrelevant or non-antibody coupled paramagnetic particles at any amount that allows for removal of monocytes (approximately a 20:1 bead:cell ratio) for about 30 minutes to 2 hours at 22 to 37 degrees C., followed by magnetic removal of cells which have attached to or engulfed the paramagnetic particles. Such separation can be performed using standard methods available in the art. For example, any magnetic separation methodology may be used including a variety of which are commercially available, (e.g., DYNAL® Magnetic Particle Concentrator (DYNAL MPC®)). Assurance of requisite depletion can be monitored by a variety of methodologies known to those of ordinary skill in the art, including flow cytometric analysis of CD14 positive cells, before and after depletion.

[0622] For isolation of a desired population of cells by positive or negative selection, the concentration of cells and surface (e.g., particles such as beads) can be varied. In certain embodiments, it may be desirable to significantly decrease the volume in which beads and cells are mixed together (i.e., increase the concentration of cells), to ensure maximum contact of

cells and beads. For example, in one embodiment, a concentration of 2 billion cells/ml is used. In one embodiment, a concentration of 1 billion cells/ml is used. In a further embodiment, greater than 100 million cells/ml is used. In a further embodiment, a concentration of cells of 10, 15, 20, 25, 30, 35, 40, 45, or 50 million cells/ml is used. In yet another embodiment, a concentration of cells from 75, 80, 85, 90, 95, or 100 million cells/ml is used. In further embodiments, concentrations of 125 or 150 million cells/ml can be used. Using high concentrations can result in increased cell yield, cell activation, and cell expansion. Further, use of high cell concentrations allows more efficient capture of cells that may weakly express target antigens of interest, such as CD28-negative T cells, or from samples where there are many tumor cells present (i.e., leukemic blood, tumor tissue, etc). Such populations of cells may have therapeutic value and would be desirable to obtain. For example, using high concentration of cells allows more efficient selection of CD8+ T cells that normally have weaker CD28 expression.

[0623] In a related embodiment, it may be desirable to use lower concentrations of cells. By significantly diluting the mixture of T cells and surface (e.g., particles such as beads), interactions between the particles and cells is minimized. This selects for cells that express high amounts of desired antigens to be bound to the particles. For example, CD4+ T cells express higher levels of CD28 and are more efficiently captured than CD8+ T cells in dilute concentrations. In one embodiment, the concentration of cells used is 5×106/ml. In other embodiments, the concentration used can be from about 1×105/ml to 1×106/ml, and any integer value in between.

[0624] T cells can also be frozen. Wishing not to be bound by theory, the freeze and subsequent thaw step provides a more uniform product by removing granulocytes and to some extent monocytes in the cell population. After a washing step to remove plasma and platelets, the cells may be suspended in a freezing solution. While many freezing solutions and parameters are known in the art and will be useful in this context, one method involves using PBS containing 20% DMSO and 8% human serum albumin, or other suitable cell freezing media, the cells then are frozen to −80° C at a rate of 1° per minute and stored in the vapor phase of a liquid nitrogen storage tank. Other methods of controlled freezing may be used as well as uncontrolled freezing immediately at −20° C. or in liquid nitrogen.

[0625] T cells may also be antigen-specific T cells. For example, tumor-specific T cells can be used. In certain embodiments, antigen-specific T cells can be isolated from a patient of interest, such as a patient afflicted with a cancer or an infectious disease. In one embodiment, neoepitopes are determined for a subject and T cells specific to these antigens are isolated.

Antigen-specific cells for use in expansion may also be generated *in vitro* using any number of methods known in the art, for example, as described in U.S. Patent Publication No. US 20040224402 entitled, Generation and Isolation of Antigen-Specific T Cells, or in U.S. Pat. No. 6,040,177. Antigen-specific cells for use herein may also be generated using any number of methods known in the art, for example, as described in Current Protocols in Immunology, or Current Protocols in Cell Biology, both published by John Wiley & Sons, Inc., Boston, Mass.

[0626] In a related embodiment, it may be desirable to sort or otherwise positively select (e.g. via magnetic selection) the antigen specific cells prior to or following one or two rounds of expansion. Sorting or positively selecting antigen-specific cells can be carried out using peptide-MHC tetramers (Altman, et al., Science. 1996 Oct. 4; 274(5284):94-6). In another embodiment, the adaptable tetramer technology approach is used (Andersen et al., 2012 Nat Protoc. 7:891-902). Tetramers are limited by the need to utilize predicted binding peptides based on prior hypotheses, and the restriction to specific HLAs. Peptide-MHC tetramers can be generated using techniques known in the art and can be made with any MHC molecule of interest and any antigen of interest as described herein. Specific epitopes to be used in this context can be identified using numerous assays known in the art. For example, the ability of a polypeptide to bind to MHC class I may be evaluated indirectly by monitoring the ability to promote incorporation of 125I labeled β2-microglobulin (β2m) into MHC class I/β2m/peptide heterotrimeric complexes (see Parker et al., J. Immunol. 152:163, 1994).

[0627] In one embodiment cells are directly labeled with an epitope-specific reagent for isolation by flow cytometry followed by characterization of phenotype and TCRs. In one embodiment, T cells are isolated by contacting with T cell specific antibodies. Sorting of antigen-specific T cells, or generally any cells, can be carried out using any of a variety of commercially available cell sorters, including, but not limited to, MoFlo sorter (DakoCytomation, Fort Collins, Colo.), FACSAria™, FACSArray™, FACSVantage™, BD™ LSR II, and FACSCalibur™ (BD Biosciences, San Jose, Calif.).

[0628] In a preferred embodiment, the method comprises selecting cells that also express CD3. The method may comprise specifically selecting the cells in any suitable manner. Preferably, the selecting is carried out using flow cytometry. The flow cytometry may be carried out using any suitable method known in the art. The flow cytometry may employ any suitable antibodies and stains. Preferably, the antibody is chosen such that it specifically recognizes and binds to the particular biomarker being selected. For example, the specific selection of CD3, CD8, TIM-3, LAG-3, 4-1BB, or PD-1 may be carried out using anti-CD3,

anti-CD8, anti-TIM-3, anti-LAG-3, anti-4-lBB, or anti-PD-1 antibodies, respectively. The antibody or antibodies may be conjugated to a bead (e.g., a magnetic bead) or to a fluorochrome. Preferably, the flow cytometry is fluorescence-activated cell sorting (FACS). TCRs expressed on T cells can be selected based on reactivity to autologous tumors. Additionally, T cells that are reactive to tumors can be selected for based on markers using the methods described in patent publication Nos. WO2014133567 and WO2014133568, herein incorporated by reference in their entirety. Additionally, activated T cells can be selected for based on surface expression of CD107a.

[0629] In one embodiment, the method further comprises expanding the numbers of T cells in the enriched cell population. Such methods are described in U.S. Patent No. 8,637,307 and is herein incorporated by reference in its entirety. The numbers of T cells may be increased at least about 3-fold (or 4-, 5-, 6-, 7-, 8-, or 9-fold), more preferably at least about 10-fold (or 20-, 30-, 40-, 50-, 60-, 70-, 80-, or 90-fold), more preferably at least about 100-fold, more preferably at least about 1,000 fold, or most preferably at least about 100,000-fold. The numbers of T cells may be expanded using any suitable method known in the art. Exemplary methods of expanding the numbers of cells are described in patent publication No. WO 2003/057171, U.S. Patent No. 8,034,334, and U.S. Patent Publication No. 2012/0244133, each of which is incorporated herein by reference.

[0630] In one embodiment, *ex vivo* T cell expansion can be performed by isolation of T cells and subsequent stimulation or activation followed by further expansion. In one embodiment, the T cells may be stimulated or activated by a single agent. In another embodiment, T cells are stimulated or activated with two agents, one that induces a primary signal and a second that is a co-stimulatory signal. Ligands useful for stimulating a single signal or stimulating a primary signal and an accessory molecule that stimulates a second signal may be used in soluble form. Ligands may be attached to the surface of a cell, to an Engineered Multivalent Signaling Platform (EMSP), or immobilized on a surface. In a preferred embodiment both primary and secondary agents are co-immobilized on a surface, for example a bead or a cell. In one embodiment, the molecule providing the primary activation signal may be a CD3 ligand, and the co-stimulatory molecule may be a CD28 ligand or 4-1BB ligand.

[0631] In certain embodiments, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in International Patent Publication No. WO 2015/120096, by a method comprising enriching a population of lymphocytes obtained from a donor subject; stimulating the population of lymphocytes with one or more T-cell stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed

system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using a single cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells for a predetermined time to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. In certain embodiments, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in WO 2015/120096, by a method comprising: obtaining a population of lymphocytes; stimulating the population of lymphocytes with one or more stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using at least one cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. The predetermined time for expanding the population of transduced T cells may be 3 days. The time from enriching the population of lymphocytes to producing the engineered T cells may be 6 days. The closed system may be a closed bag system. Further provided is population of T cells comprising a CAR or an exogenous TCR obtainable or obtained by said method, and a pharmaceutical composition comprising such cells.

[0632] In certain embodiments, T cell maturation or differentiation *in vitro* may be delayed or inhibited by the method as described in International Patent Publication No. WO 2017/070395, comprising contacting one or more T cells from a subject in need of a T cell therapy with an AKT inhibitor (such as, e.g., one or a combination of two or more AKT inhibitors disclosed in claim 8 of WO2017070395) and at least one of exogenous Interleukin-7 (IL-7) and exogenous Interleukin-15 (IL-15), wherein the resulting T cells exhibit delayed maturation or differentiation, and/or wherein the resulting T cells exhibit improved T cell function (such as, e.g., increased T cell proliferation; increased cytokine production; and/or increased cytolytic activity) relative to a T cell function of a T cell cultured in the absence of an AKT inhibitor.

[0633] In certain embodiments, a patient in need of a T cell therapy may be conditioned by a method as described in International Patent Publication No. WO 2016/191756 comprising

administering to the patient a dose of cyclophosphamide between 200 mg/m2/day and 2000 mg/m2/day and a dose of fludarabine between 20 mg/m2/day and 900 mg/m^2/day.

Diseases

Genetic Diseases and Diseases with a Genetic and/or Epigenetic Aspect

[0634] The compositions, systems, or components thereof can be used to treat and/or prevent a genetic disease or a disease with a genetic and/or epigenetic aspect. The genes and conditions exemplified herein are not exhaustive. In some embodiments, a method of treating and/or preventing a genetic disease can include administering a composition, system, and/or one or more components thereof to a subject, where the composition, system, and/or one or more components thereof is capable of modifying one or more copies of one or more genes associated with the genetic disease or a disease with a genetic and/or epigenetic aspect in one or more cells of the subject. In some embodiments, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can eliminate a genetic disease or a symptom thereof in the subject. In some embodiments, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can decrease the severity of a genetic disease or a symptom thereof in the subject. In some embodiments, the compositions, systems, or components thereof can modify one or more genes or polynucleotides associated with one or more diseases, including genetic diseases and/or those having a genetic aspect and/or epigenetic aspect, including but not limited to, any one or more set forth in Table 4. It will be appreciated that those diseases and associated genes listed herein are non-exhaustive and non-limiting. Further some genes play roles in the development of multiple diseases.

[0635] Table 2

Table 2. Exemplary Genetic and Other Diseases and Associated Genes

Disease Name	Primary Tissues or System Affected	Additional Tissues/Systems Affected	Genes
Achondroplasia	Bone and Muscle		fibroblast growth factor receptor 3 (FGFR3)
Achromatopsia	eye		CNGA3, CNGB3, GNAT2, PDE6C, PDE6H, ACHM2, ACHM3,
Acute Renal Injury	kidney		NFkappaB, AATF, p85alpha, FAS, Apoptosis cascade elements (e.g. FASR, Caspase 2, 3, 4, 6, 7, 8, 9, 10, AKT, TNF alpha, IGF1, IGF1R, RIPK1), p53

Age Related Macular Degeneration	eye		Abcr; CCL2; CC2; CP (ceruloplasmin); Timp3; cathepsinD; VLDLR, CCR2
AIDS	Immune System		KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1
Albinism (including oculocutaneous albinism (types 1-7) and ocular albinism)	Skin, hair, eyes,		TYR, OCA2, TYRP1, and SLC45A2, SLC24A5 and C10orf11
Alkaptonuria	Metabolism of amino acids	Tissues/organs where homogentisic acid accumulates, particularly cartilage (joints), heart valves, kidneys	HGD
alpha-1 antitrypsin deficiency (AATD or A1AD)	Lung	Liver, skin, vascular system, kidneys, GI	SERPINA1, those set forth in WO2017165862, PiZ allele
ALS	CNS		SOD1; ALS2; ALS3; ALS5; ALS7;STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c); DPP6; NEFH, PTGS1, SLC1A2, TNFRSF10B, PRPH, HSP90AA1, CRIA2, IFNG, AMPA2 S100B, FGF2, AOX1, CS, TXN, RAPHJ1, MAP3K5, NBEAL1, GPX1, ICA1L, RAC1, MAPT, ITPR2, ALS2CR4, GLS, ALS2CR8, CNTFR, ALS2CR11, FOLH1, FAM117B, P4HB, CNTF, SQSTM1, STRADB, NAIP, NLR, YWHAQ, SLC33A1, TRAK2, SCA1, NIF3L1, NIF3, PARD3B, COX8A, CDK15, HECW1, HECT, C2, WW 15, NOS1, MET, SOD2, HSPB1, NEFL, CTSB, ANG, HSPA8, RNase A, VAPB, VAMP, SNCA, alpha HGF, CAT, ACTB, NEFM, TH, BCL2, FAS, CASP3, CLU, SMN1, G6PD, BAX, HSF1, RNF19A, JUN, ALS2CR12, HSPA5, MAPK14, APEX1, TXNRD1, NOS2, TIMP1, CASP9, XIAP, GLG1, EPO, VEGFA, ELN, GDNF, NFE2L2, SLC6A3, HSPA4, APOE, PSMB8, DCTN2, TIMP3, KIFAP3, SLC1A1, SMN2, CCNC, STUB1, ALS2, PRDX6, SYP, CABIN1, CASP1, GART, CDK5, ATXN3, RTN4, C1QB, VEGFC, HTT, PARK7, XDH, GFAP, MAP2, CYCS, FCGR3B, CCS, UBL5, MMP9m SLC18A3, TRPM7, HSPB2, AKT1, DEERL1, CCL2, NGRN, GSR, TPPP3, APAF1, BTBD10, GLUD1, CXCR4, S:C1A3, FLT1, PON1, AR, LIF, ERBB3, :GA:S1, CD44, TP53, TLR3, GRIA1, GAPDH, AMPA, GRIK1, DES, CHAT, FLT4, CHMP2B, BAG1,

			CHRNA4, GSS, BAK1, KDR, GSTP1, OGG1, IL6
Alzheimer's Disease	Brain		E1; CHIP; UCH; UBB; Tau; LRP; PICALM; CLU; PS1; SORL1; CR1; VLDLR; UBA1; UBA3; CHIP28; AQP1; UCHL1; UCHL3; APP, AAA, CVAP, AD1, APOE, AD2, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3, ALAS2, ABCA1, BIN1, BDNF, BTNL8, C1ORF49, CDH4, CHRNB2, CKLFSF2, CLEC4E, CR1L, CSF3R, CST3, CYP2C, DAPK1, ESR1, FCAR, FCGR3B, FFA2, FGA, GAB2, GALP, GAPDHS, GMPB, HP, HTR7, IDE, IF127, IFI6, IFIT2, IL1RN, IL-1RA, IL8RA, IL8RB, JAG1, KCNJ15, LRP6, MAPT, MARK4, MPHOSPH1, MTHFR, NBN, NCSTN, NIACR2, NMNAT3, NTM, ORM1, P2RY13, PBEF1, PCK1, PICALM, PLAU, PLXNC1, PRNP, PSEN1, PSEN2, PTPRA, RALGPS2, RGSL2, SELENBP1, SLC25A37, SORL1, Mitoferrin-1, TF, TFAM, TNF, TNFRSF10C, UBE1C
Amyloidosis			APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB
Amyloid neuropathy			TTR, PALB
Anemia	Blood		CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT
Angelman Syndrome	Nervous system, brain		UBE3A
Attention Deficit Hyperactivity Disorder (ADHD)	Brain		PTCHD1
Autoimmune lymphoproliferative syndrome	Immune system		TNFRSF6, APT1, FAS, CD95, ALPS1A
Autism, Autism spectrum disorders (ASDs), including Asperger's and a general diagnostic category called Pervasive Developmental Disorders (PDDs)	Brain		PTCHD1; Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; GLO1, RTT, PPMX, MRX16, RX79, NLGN3, NLGN4, KIAA1260, AUTSX2, FMR1, FMR2; FXR1; FXR2; MGLUR5, ATP10C, CDH10, GRM6, MGLUR6, CDH9, CNTN4, NLGN2, CNTNAP2, SEMA5A, DHCR7, NLGN4X, NLGN4Y, DPP6, NLGN5, EN2, NRCAM, MDGA2, NRXN1, FMR2, AFF2, FOXP2, OR4M2, OXTR, FXR1, FXR2, PAH, GABRA1, PTEN, GABRA5, PTPRZ1, GABRB3, GABRG1, HIRIP3, SEZ6L2, HOXA1, SHANK3, IL6, SHBZRAP1, LAMB1, SLC6A4, SERT, MAPK3, TAS2R1, MAZ, TSC1, MDGA2, TSC2, MECP2, UBE3A, WNT2, see also 20110023145

autosomal dominant polycystic kidney disease (ADPKD) - (includes diseases such as von Hippel-Lindau disease and tuberous sclerosis complex disease)	kidney	liver	PKD1, PKD2
Autosomal Recessive Polycystic Kidney Disease (ARPKD)	kidney	liver	PKDH1
Ataxia-Telangiectasia (a.k.a Louis Bar syndrome)	Nervous system, immune system	various	ATM
B-Cell Non-Hodgkin Lymphoma			BCL7A, BCL7
Bardet-Biedl syndrome	Eye, musculoskeletal system, kidney, reproductive organs	Liver, ear, gastrointestinal system, brain	ARL6, BBS1, BBS2, BBS4, BBS5, BBS7, BBS9, BBS10, BBS12, CEP290, INPP5E, LZTFL1, MKKS, MKS1, SDCCAG8, TRIM32, TTC8
Bare Lymphocyte Syndrome	blood		TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5
Bartter's Syndrome (types I, II, III, IVA and B, and V)	kidney		SLC12A1 (type I), KCNJ1 (type II), CLCNKB (type III), BSND (type IV A), or both the CLCNKA CLCNKB genes (type IV B), CASR (type V).
Becker muscular dystrophy	Muscle		DMD, BMD, MYF6
Best Disease (Vitelliform Macular Dystrophy type 2)	eye		VMD2
Bleeding Disorders	blood		TBXA2R, P2RX1, P2X1
Blue Cone Monochromacy	eye		OPN1LW, OPN1MW, and LCR
Breast Cancer	Breast tissue		BRCA1, BRCA2, COX-2
Bruton's Disease (aka X-linked Agammglobulinemia)	Immune system, specifically B cells		BTK
Cancers (e.g., lymphoma, chronic lymphocytic leukemia (CLL), B cell acute lymphocytic leukemia (B-ALL), acute lymphoblastic leukemia, acute myeloid leukemia, non-Hodgkin's lymphoma (NHL), diffuse large cell lymphoma (DLCL), multiple myeloma, renal cell carcinoma (RCC), neuroblastoma, colorectal cancer, breast cancer, ovarian cancer, melanoma, sarcoma, prostate cancer, lung cancer, esophageal cancer, hepatocellular carcinoma, pancreatic cancer, astrocytoma, mesothelioma, head and neck cancer, and medulloblastoma	Various		FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC, TRBC, those described in WO2015048577
Cardiovascular Diseases	heart	Vascular system	IL1B, XDH, TP53, PTGS, MB, IL4, ANGPT1, ABCGu8, CTSK, PTGIR, KCNJ11, INS, CRP, PDGFRB, CCNA2, PDGFB, KCNJ5, KCNN3, CAPN10, ADRA2B, ABCG5, PRDX2, CPAN5, PARP14, MEX3C, ACE, RNF, IL6, TNF, STN, SERPINE1, ALB, ADIPOQ, APOB, APOE, LEP, MTHFR, APOA1,

			EDN1, NPPB, NOS3, PPARG, PLAT, PTGS2, CETP, AGTR1, HMGCR, IGF1, SELE, REN, PPARA, PON1, KNG1, CCL2, LPL, VWF, F2, ICAM1, TGFB, NPPA, IL10, EPO, SOD1, VCAM1, IFNG, LPA, MPO, ESR1, MAPK, HP, F3, CST3, COG2, MMP9, SERPINC1, F8, HMOX1, APOC3, IL8, PROL1, CBS, NOS2, TLR4, SELP, ABCA1, AGT, LDLR, GPT, VEGFA, NR3C2, IL18, NOS1, NR3C1, FGB, HGF, IL1A, AKT1, LIPC, HSPD1, MAPK14, SPP1, ITGB3, CAT, UTS2, THBD, F10, CP, TNFRSF11B, EGFR, MMP2, PLG, NPY, RHOD, MAPK8, MYC, FN1, CMA1, PLAU, GNB3, ADRB2, SOD2, F5, VDR, ALOX5, HLA-DRB1, PARP1, CD40LG, PON2, AGER, IRS1, PTGS1, ECE1, F7, IRMN, EPHX2, IGFBP1, MAPK10, FAS, ABCB1, JUN, IGFBP3, CD14, PDE5A, AGTR2, CD40, LCAT, CCR5, MMP1, TIMP1, ADM, DYT10, STAT3, MMP3, ELN, USF1, CFH, HSPA4, MMP12, MME, F2R, SELL, CTSB, ANXA5, ADRB1, CYBA, FGA, GGT1, LIPG, HIF1A, CXCR4, PROC, SCARB1, CD79A, PLTP, ADD1, FGG, SAA1, KCNH2, DPP4, NPR1, VTN, KIAA0101, FOS, TLR2, PPIG, IL1R1, AR, CYP1A1, SERPINA1, MTR, RBP4, APOA4, CDKN2A, FGF2, EDNRB, ITGA2, VLA-2, CABIN1, SHBG, HMGB1, HSP90B2P, CYP3A4, GJA1, CAV1, ESR2, LTA, GDF15, BDNF, CYP2D6, NGF, SP1, TGIF1, SRC, EGF, PIK3CG, HLA-A, KCNQ1, CNR1, FBN1, CHKA, BEST1, CTNNB1, IL2, CD36, PRKAB1, TPO, ALDH7A1, CX3CR1, TH, F9, CH1, TF, HFE, IL17A, PTEN, GSTM1, DMD, GATA4, F13A1, TTR, FABP4, PON3, APOC1, INSR, TNFRSF1B, HTR2A, CSF3, CYP2C9, TXN, CYP11B2, PTH, CSF2, KDR, PLA2G2A, THBS1, GCG, RHOA, ALDH2, TCF7L2, NFE2L2, NOTCH1, UGT1A1, IFNA1, PPARD, SIRT1, GNHR1, PAPPA, ARR3, NPPC, AHSP, PTK2, IL13, MTOR, ITGB2, GSTT1, IL6ST, CPB2, CYP1A2, HNF4A, SLC64A, PLA2G6, TNFSF11, SLC8A1, F2RL1, AKR1A1, ALDH9A1, BGLAP, MTTP, MTRR, SULT1A3, RAGE, C4B, P2RY12, RNLS, CREB1, POMC, RAC1, LMNA, CD59, SCM5A, CYP1B1, MIF, MMP13,

			TIMP2, CYP19A1, CUP21A2, PTPN22, MYH14, MBL2, SELPLG, AOC3, CTSL1, PCNA, IGF2, ITGB1, CAST, CXCL12, IGHE, KCNE1, TFRC, COL1A1, COL1A2, IL2RB, PLA2G10, ANGPT2, PROCR, NOX4, HAMP, PTPN11, SLCA1, IL2RA, CCL5, IRF1, CF:AR, CA:CA, EIF4E, GSTP1, JAK2, CYP3A5, HSPG2, CCL3, MYD88, VIP, SOAT1, ADRBK1, NR4A2, MMP8, NPR2, GCH1, EPRS, PPARGC1A, F12, PECAM1, CCL4, CERPINA34, CASR, FABP2, TTF2, PROS1, CTF1, SGCB, YME1L1, CAMP, ZC3H12A, AKR1B1, MMP7, AHR, CSF1, HDAC9, CTGF, KCNMA1, UGT1A, PRKCA, COMT, S100B, EGR1, PRL, IL15, DRD4, CAMK2G, SLC22A2, CCL11, PGF, THPO, GP6, TACR1, NTS, HNF1A, SST, KCDN1, LOC646627, TBXAS1, CUP2J2, TBXA2R, ADH1C, ALOX12, AHSG, BHMT, GJA4, SLC25A4, ACLY, ALOX5AP, NUMA1, CYP27B1, CYSLTR2, SOD3, LTC4S, UCN, GHRL, APOC2, CLEC4A, KBTBD10, TNC, TYMS, SHC1, LRP1, SOCS3, ADH1B, KLK3, HSD11B1, VKORC1, SERPINB2, TNS1, RNF19A, EPOR, ITGAM, PITX2, MAPK7, FCGR3A, LEEPR, ENG, GPX1, GOT2, HRH1, NR112, CRH, HTR1A, VDAC1, HPSE, SFTPD, TAP2, RMF123, PTK2Bm NTRK2, IL6R, ACHE, GLP1R, GHR, GSR, NQO1, NR5A1, GJB2, SLC9A1, MAOA, PCSK9, FCGR2A, SERPINF1, EDN3, UCP2, TFAP2A, C4BPA, SERPINF2, TYMP, ALPP, CXCR2, SLC3A3, ABCG2, ADA, JAK3, HSPA1A, FASN, FGF1, F11, ATP7A, CR1, GFPA, ROCK1, MECP2, MYLK, BCHE, LIPE, ADORA1, WRN, CXCR3, CD81, SMAD7, LAMC2, MAP3K5, CHGA, IAPP, RHO, ENPP1, PTHLH, NRG1, VEGFC, ENPEP, CEBPB, NAGLU,. F2RL3, CX3CL1, BDKRB1, ADAMTS13, ELANE, ENPP2, CISH, GAST, MYOC, ATP1A2, NF1, GJB1, MEF2A, VCL, BMPR2, TUBB, CDC42, KRT18, HSF1, MYB, PRKAA2, ROCK2, TFP1, PRKG1, BMP2, CTNND1, CTH, CTSS, VAV2, NPY2R, IGFBP2, CD28, GSTA1, PPIA, APOH, S100A8, IL11, ALOX15, FBLN1, NR1H3, SCD, GIP, CHGB, PRKCB, SRD5A1,HSD11B2, CALCRL, GALNT2, ANGPTL4,

			KCNN4, PIK3C2A, HBEGF, CYP7A1, HLA-DRB5, BNIP3, GCKR, S100A12, PADI4, HSPA14, CXCR1, H19, KRTAP19-3, IDDM2, RAC2, YRY1, CLOCK, NGFR, DBH, CHRNA4, CACNA1C, PRKAG2, CHAT, PTGDS, NR1H2, TEK, VEGFB, MEF2C, MAPKAPK2, TNFRSF11A, HSPA9, CYSLTR1, MAT1A, OPRL1, IMPA1, CLCN2, DLD, PSMA6, PSMB8, CHI3L1, ALDH1B1, PARP2,STAR, LBP, ABCC6, RGS2, EFNB2, GJB6, APOA2, AMPD1, DYSF, FDFT1,EMD2, CCR6, GJB3, IL1RL1, ENTPD1, BBS4, CELSR2, F11R, RAPGEF3, HYAL1, ZNF259, ATOX1, ATF6, KHK, SAT1, GGH, TIMP4, SLC4A4, PDE2A, PDE3B, FADS1, FADS2, TMSB4X, TXNIP, LIMS1, RHOB, LY96, FOXO1, PNPLA2,TRH, GJC1, S:C17A5, FTO, GJD2, PRSC1, CASP12, GPBAR1, PXK, IL33, TRIB1, PBX4, NUPR1, 15-SEP, CILP2, TERC, GGT2, MTCO1, UOX, AVP
Cataract	eye		CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1
CDKL-5 Deficiencies or Mediated Diseases	Brain, CNS		CDKL5
Charcot-Marie-Tooth (CMT) disease (Types 1, 2, 3, 4,)	Nervous system	Muscles (dystrophy)	PMP22 (CMT1A and E), MPZ (CMT1B), LITAF (CMT1C), EGR2 (CMT1D), NEFL (CMT1F), GJB1 (CMT1X), MFN2 (CMT2A), KIF1B (CMT2A2B), RAB7A (CMT2B), TRPV4 (CMT2C), GARS (CMT2D), NEFL (CMT2E), GAPD1 (CMT2K), HSPB8 (CMT2L), DYNC1H1, CMT2O), LRSAM1 (CMT2P), IGHMBP2 (CMT2S), MORC2 (CMT2Z), GDAP1 (CMT4A), MTMR2 or SBF2/MTMR13 (CMT4B), SH3TC2 (CMT4C), NDRG1 (CMT4D), PRX (CMT4F), FIG4 (CMT4J), NT-3
Chédiak-Higashi Syndrome	Immune system	Skin, hair, eyes, neurons	LYST
Choroidermia			CHM, REP1,
Chorioretinal atrophy	eye		PRDM13, RGR, TEAD1
Chronic Granulomatous Disease	Immune system		CYBA, CYBB, NCF1, NCF2, NCF4

Chronic Mucocutaneous Candidiasis	Immune system		AIRE, CARD9, CLEC7A IL12B, IL12B1, IL1F, IL17RA, IL17RC, RORC, STAT1, STAT3, TRAF31P2
Cirrhosis	liver		KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988
Colon cancer (Familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC))	Gastrointestinal		FAP: APC HNPCC: MSH2, MLH1, PMS2, SH6, PMS1
Combined Immunodeficiency	Immune System		IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228
Cone(-rod) dystrophy	eye		AIPL1, CRX, GUA1A, GUCY2D, PITPM3, PROM1, PRPH2, RIMS1, SEMA4A, ABCA4, ADAM9, ATF6, C21ORF2, C8ORF37, CACNA2D4, CDHR1, CERKL, CNGA3, CNGB3, CNNM4, CNAT2, IFT81, KCNV2, PDE6C, PDE6H, POC1B, RAX2, RDH5, RPGRIP1, TTLL5, RetCG1, GUCY2E
Congenital Stationary Night Blindness	eye		CABP4, CACNA1F, CACNA2D4, GNAT1, CPR179, GRK1, GRM6, LRIT3, NYX, PDE6B, RDH5, RHO, RLBP1, RPE65, SAG, SLC24A1, TRPM1,
Congenital Fructose Intolerance	Metabolism		ALDOB
Cori's Disease (Glycogen Storage Disease Type III)	Various-wherever glycogen accumulates, particularly liver, heart, skeletal muscle		AGL
Corneal clouding and dystrophy	eye		APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD
Cornea plana congenital			KERA, CNA2
Cri du chat Syndrome, also known as 5p syndrome and cat cry syndrome			Deletions involving only band 5p15.2 to the entire short arm of chromosome 5, e.g. CTNND2, TERT,
Cystic Fibrosis (CF)	Lungs and respiratory system	Pancreas, liver, digestive system, reproductive system, exocrine, glands,	CTFR, ABCC7, CF, MRP7, SCNN1A, those described in WO2015157070
Diabetic nephropathy	kidney		Gremlin, 12/15- lipoxygenase, TIM44,
Dent Disease (Types 1 and 2)	Kidney		Type 1: CLCN5, Type 2: ORCL
Dentatorubro-Pallidoluysian Atrophy (DRPLA) (aka Haw River and Naito-Oyanagi Disease)	CNS, brain, muscle		Atrophin-1 and Atn1
Down Syndrome	various		Chromosome 21 trisomy
Drug Addiction	Brain		Prkce; Drd2; Drd4; ABAT; GRIA2;Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1

Duane syndrome (Types 1, 2, and 3, including subgroups A, B and C). Other names for this condition include: Duane's Retraction Syndrome (or DR syndrome), Eye Retraction Syndrome, Retraction Syndrome, Congenital retraction syndrome and Stilling-Turk-Duane Syndrome	eye		CHN1, indels on chromosomes 4 and 8
Duchenne muscular dystrophy (DMD)	muscle	Cardiovascular, respiratory	DMD, BMD, dystrophin gene, intron flanking exon 51 of DMD gene, exon 51 mutations in DMD gene, see also WO2013163628 and US Pat. Pub. 20130145487
Edward's Syndrome (Trisomy 18)			Complete or partial trisomy of chromosome 18
Ehlers-Danlos Syndrome (Types I-VI)	Various depending on type: including musculoskeletal, eye, vasculature, immune, and skin		COL5A1, COL5A2, COL1A1, COL3A1, TNXB, PLOD1, COL1A2, FKBP14 and ADAMTS2
Emery-Dreifuss muscular dystrophy	muscle		LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A
Enhanced S-Cone Syndrome	eye		NR2E3, NRL
Fabry's Disease	Various – including skin, eyes, and gastrointestinal system, kidney, heart, brain, nervous system		GLA
Facioscapulohumeral muscular dystrophy	muscles		FSHMD1A, FSHD1A, FRG1,
Factor H and Factor H-like 1	blood		HF1, CFH, HUS
Factor V Leiden thrombophilia and Factor V deficiency	blood		Factor V (F5)
Factor V and Factor VII deficiency	blood		MCFD2
Factor VII deficiency	blood		F7
Factor X deficiency	blood		F10
Factor XI deficiency	blood		F11
Factor XII deficiency	blood		F12, HAF
Factor XIIIA deficiency	blood		F13A1, F13A
Factor XIIIB deficiency	blood		F13B
Familial Hypercholestereolemia	Cardiovascular system		APOB, LDLR, PCSK9
Familial Mediterranean Fever (FMF) also called recurrent polyserositis or familial paroxysmal polyserositis	Various-organs/tissues with serous or synovial membranes, skin, joints	Heart, kidney, brain/CNS, reproductive organs	MEFV
Fanconi Anemia	Various – blood (anemia), immune system,		FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCC, FANCG, RAD51, BRCA1,

	cognitive, kidneys, eyes, musculoskeletal		BRCA2, BRIP1, BACH1, FANCJ, FANCB, FANCD1, FANCD2, FANCD, FAD, FANCE, FACE, FANCF, FANCI, ERCC4, FANCL, FANCM, PALB2, RAD51C, SLX4, UBE2T, FANCB, XRCC9, PHF9, KIAA1596
Fanconi Syndrome Types I (Childhood onset) and II (Adult Onset)	kidneys		FRTS1, GATM
Fragile X syndrome and related disorders	brain		FMR1, FMR2; FXR1; FXR2; mGLUR5
Fragile XE Mental Retardation (aka Martin Bell syndrome)	Brain, nervous system		FMR1
Friedreich Ataxia (FRDA)	Brain, nervous system	heart	FXN/X25
Fuchs endothelial corneal dystrophy	Eye		TCF4; COL8A2
Galactosemia	Carbohydrate metabolism disorder	Various-where galactose accumulates – liver, brain, eyes	GALT, GALK1, and GALE
Gastrointestinal Epithelial Cancer, GI cancer			CISH
Gaucher Disease (Types 1, 2, and 3, as well as other unusual forms that may not fit into these types)	Fat metabolism disorder	Various-liver, spleen, blood, CNS, skeletal system	GBA
Griscelli syndrome			
Glaucoma	eye		MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A, those described in WO2015153780
Glomerulo sclerosis	kidney		CC chemokine ligand 2
Glycogen Storage Diseases Types I-VI -See also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease	Metabolism Diseases		SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM, see also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease
RBC Glycolytic enzyme deficiency	blood		any mutations in a gene for an enzyme in the glycolysis pathway including mutations in genes for hexokinases I and II, glucokinase, phosphoglucose isomerase, phosphofructokinase, aldolase Bm triosephosphate isomerease, glyceraldehydee-3-phosphate dehydrogenase, phosphoglycerokinase, phosphoglycerate mutase, enolase I, pyruvate kinase
Hartnup's disease	Malabsorption disease	Various- brain, gastrointestinal, skin,	SLC6A19
Hearing Loss	ear		NOX3, Hes5, BDNF,

Hemochromatosis (HH)	Iron absorption regulation disease	Various- wherever iron accumulates, liver, heart, pancreas, joints, pituitary gland	HFE and H63D
Hemophagocytic lymphohistiocytosis disorders	blood		PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3
Hemorrhagic disorders	blood		PI, ATT, F5
Hers disease (Glycogen storage disease Type VI)	liver	muscle	PYGL
Hereditary angioedema (HAE)			kalikrein B1
Hereditary Hemorrhagic Telangiectasia (Osler-Weber-Rendu Syndrome)	Skin and mucous membranes		ACVRL1, ENG and SMAD4
Hereditary Spherocytosis	blood		NK1, EPB42, SLC4A1, SPTA1, and SPTB
Hereditary Persistence of Fetal Hemoglobin	blood		HBG1, HBG2, BCL11A, promoter region of HBG 1 and/or 2 (in the CCAAT box)
Hemophilia (hemophilia A (Classic) a B (aka Christmas disease) and C)	blood		A: FVIII, F8C, HEMA B: FVIX, HEMB C: F9, F11
Hepatic adenoma	liver		TCF1, HNF1A, MODY3
Hepatic failure, early onset, and neurologic disorder	liver		SCOD1, SCO1
Hepatic lipase deficiency	liver		LIPC
Hepatoblastoma, cancer and carcinomas	liver		CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5
Hermansky-Pudlak syndrome	Skin, eyes, blood, lung, kidneys, intestine		HPS1, HPS3, HPS4, HPS5, HPS6, HPS7, DTNBP1, BLOC1, BLOC1S2, BLOC3
HIV susceptibility or infection	Immune system		IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5), those in WO2015148670A1
Holoprosencephaly (HPE) (Alobar, Semilobar, and Lobar)	brain		ACVRL1, ENG, SMAD4
Homocystinuria	Metabolic disease	Various- connective tissue, muscles, CNS, cardiovascular system	CBS, MTHFR, MTR, MTRR, and MMADHC
HPV			HPV16 and HPV18 E6/E7
HSV1, HSV2, and related keratitis	eye		HSV1 genes (immediate early and late HSV-1 genes (UL1, 1.5, 5, 6, 8, 9, 12, 15, 16, 18, 19, 22, 23, 26, 26.5, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 42, 48, 49.5, 50, 52, 54, S6, RL2, RS1, those described in WO2015153789, WO2015153791
Hunter's Syndrome (aka Mucopolysaccharidosis type II)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDS

Huntington's disease (HD) and HD-like disorders	Brain, nervous system		HD, HTT, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17, PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2, and those described in WO2013130824, WO2015089354
Hurler's Syndrome (aka mucopolysaccharidosis type I H, MPS IH)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α-L-iduronidase
Hurler-Scheie syndrome (aka mucopolysaccharidosis type I H-S, MPS I H-S)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α-L-iduronidase
hyaluronidase deficiency (aka MPS IX)	Soft and connective tissues		HYAL1
Hyper IgM syndrome	Immune system		CD40L
Hyper- tension caused renal damage	kidney		Mineral corticoid receptor
Immunodeficiencies	Immune System		CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI
Inborn errors of metabolism: including urea cycle disorders, organic acidemias), fatty acid oxidation defects, amino acidopathies, carbohydrate disorders, mitochondrial disorders	Metabolism diseases, liver	Various organs and cells	See also: Carbohydrate metabolism disorders (e.g. galactosemia), Amino acid Metabolism disorders (e.g. phenylketonuria), Fatty acid metabolism (e.g. MCAD deficiency), Urea Cycle disorders (e.g. Citrullinemia), Organic acidemias (e.g. Maple Syrup Urine disease), Mitochondrial disorders (e.g. MELAS), peroxisomal disorders (e.g. Zellweger syndrome)
Inflammation	Various		IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
Inflammatory Bowel Diseases (e.g. Ulcerative Colitis and Chron's Disease)	Gastrointestinal	Joints, skin	NOD2, IRGM, LRRK2, ATG5, ATG16L1, IRGM, GATM, ECM1, CDH1, LAMB1, HNF4A, GNA12, IL10, CARD9/15. CCR6, IL2RA, MST1, TNFSF15, REL, STAT3, IL23R, IL12B, FUT2
Interstitial renal fibrosis	kidney		TGF-β type II receptor
Job's Syndrome (aka Hyper IgE Syndrome)	Immune System		STAT3, DOCK8
Juvenile Retinoschisis	eye		RS1, XLRS1
Kabuki Syndrome 1			MLL4, KMT2D
Kennedy Disease (aka Spinobulbar Muscular Atrophy)	Muscles, brain, nervous system		SBMA/SMAX1/AR
Klinefelter syndrome	Various- particularly those involved		Extra X chromosome in males

	in development of male characteristics		
Lafora Disease	Brain, CNS		EMP2A and EMP2B
Leber Congenital Amaurosis	eye		CRB1, RP12, CORD2, CRD, CRX, IMPDH1, OTX2, AIPL1, CABP4, CCT2, CEP290, CLUAP1, CRB1, CRX, DTHD1, GDF6, GUCY2D, IFT140, IQCB1, KCNJ13, LCA5, LRAT, NMNAT1, PRPH2, RD3, RDH12, RPE65, RP20, RPGRIP1, SPATA7, TULP1, LCA1, LCA4, GUC2D, CORD6, LCA3,
Lesch-Nyhan Syndrome	Metabolism disease	Various – joints, cognitive, brain, nervous system	HPRT1
Leukocyte deficiencies and disorders	blood		ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4
Leukemia	Blood		TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN
Limb-girdle muscular dystrophy diseases	muscle		LGMD
Lowe syndrome	brain, eyes, kidneys		OCRL
Lupus glomerulo- nephritis	kidney		MAPK1
Machado-Joseph's Disease (also known as Spinocerebellar ataxia Type 3)	Brain, CNS, muscle		ATX3
Macular degeneration	eye		ABC4, CBC1, CHM1, APOE, C1QTNF5, C2, C3, CCL2, CCR2, CD36, CFB, CFH, CFHR1, CFHR3, CNGB3, CP, CRP, CST3, CTSD, CX3CR1, ELOVL4, ERCC6, FBLN5, FBLN6, FSCN2, HMCN1, HTRA1, IL6, IL8, PLEKHA1, PROM1, PRPH2, RPGR, SERPING1, TCOF1, TIMP3, TLR3
Macular Dystrophy	eye		BEST1, C1QTNF5, CTNNA1, EFEMP1, ELOVL4, FSCN2, GUCA1B, HMCN1, IMPG1, OTX2,

			PRDM13, PROM1, PRPH2, RP1L1, TIMP3, ABCA4, CFH, DRAM2, IMG1, MFSD8, ADMD, STGD2, STGD3, RDS, RP7, PRPH, AVMD, AOFMD, VMD2
Malattia Leventinesse	eye		EFEMP1, FBLN3
Maple Syrup Urine Disease	Metabolism disease		BCKDHA, BCKDHB, and DBT
Marfan syndrome	Connective tissue	Musculoskeletal	FBN1
Maroteaux-Lamy Syndrome (aka MPS VI)	Musculoskeletal system, nervous system	Liver, spleen	ARSB
McArdle's Disease (Glycogen Storage Disease Type V)	Glycogen storage disease	muscle	PYGM
Medullary cystic kidney disease	kidney		UMOD, HNFJ, FJHN, MCKD2, ADMCKD2
Metachromatic leukodystrophy	Lysosomal storage disease	Nervous system	ARSA
Methylmalonic acidemia (MMA)	Metabolism disease		MMAA, MMAB, MUT, MMACHC, MMADHC, LMBRD1
Morquio Syndrome (aka MPS IV A and B)	Connective tissue, skin, bone, eyes	heart	GALNS
Mucopolysaccharidosis diseases (Types I H/S, I H, II, III A B and C, I S, IVA and B, IX, VII, and VI)	Lysosomal storage disease – affects various organs/tissues		See also Hurler/Scheie syndrome, Hurler disease, Sanfillipo syndrome, Scheie syndrome, Morquio syndrome, hyaluronidase deficiency, Sly syndrome, and Maroteaux-Lamy syndrome
Muscular Atrophy	muscle		VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1
Muscular dystrophy	muscle		FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1
Myotonic dystrophy (Type 1 and Type 2)	Muscles	Eyes, heart, endocrine	CNBP (Type 2) and DMPK (Type 1)
Neoplasia			PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family

			members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bcl2; caspases family (9 members: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Neurofibromatosis (NF) (NF1, formerly Recklinghausen's NF, and NF2)	brain, spinal cord, nerves, and skin		NF1, NF2
Niemann-Pick Lipidosis (Types A, B, and C)	Lysosomal Storage Disease	Various- where sphingomyelin accumulates, particularly spleen, liver, blood, CNS	Types A and B: SMPD1; Type C: NPC1 or NPC2
Noonan Syndrome	Various – musculoskeletal, heart, eyes, reproductive organs, blood		PTPN11, SOS1, RAF1 and KRAS
Norrie Disease or X-linked Familial Exudative Vitreoretinopathy	eye		NDP
North Carolina Macular Dystrophy	eye		MCDR1
Osteogenesis imperfecta (OI) (Types I, II, III, IV, V, VI, VII)	bones, musculoskeletal		COL1A1, COL1A2, CRTAP, P3H
Osteopetrosis	bones		LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1
Patau's Syndrome (Trisomy 13)	Brain, heart, skeletal system		Additional copy of chromosome 13
Parkinson's disease (PD)	Brain, nervous system		SNCA (PARK1), UCHL1 (PARK 5), and LRRK2 (PARK8), (PARK3), PARK2, PARK4, PARK7 (PARK7), PINK1 (PARK6); x-Synuclein, DJ-1, Parkin, NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, NCAP, PRKN, PDJ, DBH, NDUFV2
Pattern Dystrophy of the RPE	eye		RDS/peripherin
Phenylketonuria (PKU)	Metabolism disorder	Various due to build-up of phenylalanine, phenyl ketones in tissues and CNS	PAH, PKU1, QDPR, DHPR, PTS
Polycystic kidney and hepatic disease	Kidney, liver		FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63
Pompe's Disease	Glycogen storage disease	Various – heart, liver, spleen	GAA
Porphyria (actually refers to a group of different diseases all	Various- wherever heme		ALAD, ALAS2, CPOX, FECH, HMBS, PPOX, UROD, or UROS

having a specific heme production process abnormality)	precursors accumulate		
posterior polymorphous corneal dystrophy	eyes		TCF4; COL8A2
Primary Hyperoxaluria (e.g. type 1)	Various – eyes, heart, kidneys, skeletal system		LDHA (lactate dehydrogenase A) and hydroxyacid oxidase 1 (HAO1)
Primary Open Angle Glaucoma (POAG)	eyes		MYOC
Primary sclerosing cholangitis	Liver, gallbladder		TCF4; COL8A2
Progeria (also called Hutchinson-Gilford progeria syndrome)	All		LMNA
Prader-Willi Syndrome	Musculoskeletal system, brain, reproductive and endocrine system		Deletion of region of short arm of chromosome 15, including UBE3A
Prostate Cancer	prostate		HOXB13, MSMB, GPRC6A, TP53
Pyruvate Dehydrogenase Deficiency	Brain, nervous system		PDHA1
Kidney/Renal carcinoma	kidney		RLIP76, VEGF
Rett Syndrome	Brain		MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1
Retinitis pigmentosa (RP)	eye		ADIPOR1, ABCA4, AGBL5, ARHGEF18, ARL2BP, ARL3, ARL6, BEST1, BBS1, BBS2, C2ORF71, C8ORF37, CA4, CERKL, CLRN1, CNGA1, CMGB1, CRB1, CRX, CYP4V2, DHDDS, DHX38, EMC1, EYS, FAM161A, FSCN2, GPR125, GUCA1B, HK1, HPRPF3, HGSNAT, IDH3B, IMPDH1, IMPG2, IFT140, IFT172, KLHL7, KIAA1549, KIZ, LRAT, MAK, MERTK, MVK, NEK2, NUROD1, NR2E3, NRL, OFD1, PDE6A, PDE6B, PDE6G, POMGNT1, PRCD, PROM1, PRPF3, PRPF4, PRPF6, PRPF8, PRPF31, PRPH2, RPB3, RDH12, REEP6, RP39, RGR, RHO, RLBP1, ROM1, RP1, RP1L1, RPY, RP2, RP9, RPE65, RPGR, SAMD11, SAG, SEMA4A, SLC7A14, SNRNP200, SPP2, SPATA7, TRNT1, TOPORS, TTC8, TULP1, USH2A, ZFN408, ZNF513, see also 20120204282
Scheie syndrome (also known as mucopolysaccharidosis type I S(MPS I-S))	Various- liver, spleen, eye, joint, heart, brain, skeletal		IDUA, α-L-iduronidase
Schizophrenia	Brain		Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin); Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a;

			GSK3b; 5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1); TCF4; COL8A2
Secretase Related Disorders	Various		APH-1 (alpha and beta); PSEN1; NCSTN; PEN-2; Nos1, Parp1, Nat1, Nat2, CTSB, APP, APH1B, PSEN2, PSENEN, BACE1, ITM2B, CTSD, NOTCH1, TNF, INS, DYT10, ADAM17, APOE, ACE, STN, TP53, IL6, NGFR, IL1B, ACHE, CTNNB1, IGF1, IFNG, NRG1, CASP3, MAPK1, CDH1, APBB1, HMGCR, CREB1, PTGS2, HES1, CAT, TGFB1, ENO2, ERBB4, TRAPPC10, MAOB, NGF, MMP12, JAG1, CD40LG, PPARG, FGF2, LRP1, NOTCH4, MAPK8, PREP, NOTCH3, PRNP, CTSG, EGF, REN, CD44, SELP, GHR, ADCYAP1, INSR, GFAP, MMP3, MAPK10, SP1, MYC, CTSE, PPARA, JUN, TIMP1, IL5, IL1A, MMP9, HTR4, HSPG2, KRAS, CYCS, SMG1, IL1R1, PROK1, MAPK3, NTRK1, IL13, MME, TKT, CXCR2, CHRM1, ATXN1,PAWR, NOTCJ2, M6PR, CYP46A1, CSNK1D, MAPK14, PRG2, PRKCA, L1 CAM, CD40, NR1I2, JAG2, CTNND1, CMA1, SORT1, DLK1, THEM4, JUP, CD46, CCL11, CAV3, RNASE3, HSPA8, CASP9, CYP3A4, CCR3, TFAP2A, SCP2, CDK4, JOF1A, TCF7L2, B3GALTL, MDM2, RELA, CASP7, IDE, FANP4, CASK, ADCYAP1R1, ATF4, PDGFA, C21ORF33, SCG5, RMF123, NKFB1, ERBB2, CAV1, MMP7, TGFA, RXRA, STX1A, PSMC4, P2RY2, TNFRSF21, DLG1, NUMBL, SPN, PLSCR1, UBQLN2, UBQLN1, PCSK7, SPON1, SILV, QPCT, HESS, GCC1
Selective IgA Deficiency	Immune system		Type 1: MSH5; Type 2: TNFRSF13B
Severe Combined Immunodeficiency (SCID) and SCID-X1, and ADA-SCID	Immune system		JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4, those identified in US Pat. App. Pub. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937;
Sickle cell disease	blood		HBB, BCL11A, BCL11Ae, cis-regulatory elements of the B-globin locus, HBG 1/2 promoter, HBG distal CCAAT box region between -92 and -130 of the HBG Transcription Start Site, those described in WO2015148863, WO 2013/126794, US Pat. Pub. 20110182867
Sly Syndrome (aka MPS VII)			GUSB

Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, 8, 12 and 17)			ATXN1, ATXN2, ATX3
Sorsby Fundus Dystrophy	eye		TIMP3
Stargardt disease	eye		ABCR, ELOVL4, ABCA4, PROM1
Tay-Sachs Disease	Lysosomal Storage disease	Various – CNS, brain, eye	HEX-A
Thalassemia (Alpha, Beta, Delta)	blood		HBA1, HBA2 (Alpha), HBB (Beta), HBB and HBD (delta), LCRB, BCL11A, BCL11Ae, cis-regulatory elements of the B-globin locus, HBG 1/2 promoter, those described in WO2015148860, US Pat. Pub. 20110182867, 2015/148860
Thymic Aplasia (DiGeorge Syndrome;22q11.2 deletion syndrome)	Immune system, thymus		deletion of 30 to 40 genes in the middle of chromosome 22 at a location known as 22q11.2, including TBX1, DGCR8
Transthyretin amyloidosis (ATTR)	liver		TTR (transthyretin)
trimethylaminuria	Metabolism disease		FMO3
Trinucleotide Repeat Disorders (generally)	Various		HTT; SBMA/SMAX1/AR; FXN/X25 ATX3; ATXN1; ATXN2; DMPK; Atrophin-1 and Atn1 (DRPLA Dx); CBP (Creb-BP - global instability); VLDLR; Atxn7; Atxn10; FEN1, TNRC6A, PABPN1, JPH3, MED15, ATXN1, ATXN3, TBP, CACNA1A, ATXN80S, PPP2R2B, ATXN7, TNRC6B, TNRC6C, CELF3, MAB21L1, MSH2, TMEM185A, SIX5, CNPY3, RAXE, GNB2, RPL14, ATXN8, ISR, TTR, EP400, GIGYF2, OGG1, STC1, CNDP1, C10ORF2, MAML3, DKC1, PAXIP1, CASK, MAPT, SP1, POLG, AFF2, THBS1, TP53, ESR1, CGGBP1, ABT1, KLK3, PRNP, JUN, KCNN3, BAX, FRAXA, KBTBD10, MBNL1, RAD51, NCOA3, ERDA1, TSC1, COMP, GGLC, RRAD, MSH3, DRD2, CD44, CTCF, CCND1, CLSPN, MEF2A, PTPRU, GAPDH, TRIM22, WT1, AHR, GPX1, TPMT, NDP, ARX, TYR, EGR1, UNG, NUMBL, FABP2, EN2, CRYGC, SRP14, CRYGB, PDCD1,HOXA1, ATXN2L, PMS2, GLA, CBL, FTH1, IL12RB2, OTX2, HOXA5, POLG2, DLX2, AHRR, MANF, RMEM158, see also 20110016540
Turner's Syndrome (XO)	Various – reproductive organs, and sex characteristics, vasculature		Monosomy X
Tuberous Sclerosis	CNS, heart, kidneys		TSC1, TSC2

Usher syndrome (Types I, II, and III)	Ears, eyes		ABHD12, CDH23, CIB2, CLRN1, DFNB31, GPR98, HARS, MYO7A, PCDH15, USH1C, USH1G, USH2A, USH11A, those described in WO2015134812A1
Velocardiofacial syndrome (aka 22q11.2 deletion syndrome, DiGeorge syndrome, conotruncal anomaly face syndrome (CTAF), autosomal dominant Opitz G/BB syndrome or Cayler cardiofacial syndrome)	Various – skeletal, heart, kidney, immune system, brain		Many genes are deleted, COM, TBX1, and other are associated with symptoms
Von Gierke's Disease (Glycogen Storage Disease type I)	Glycogen Storage disease	Various – liver, kidney	G6PC and SLC37A4
Von Hippel-Lindau Syndrome	Various – cell growth regulation disorder	CNS, Kidney, Eye, visceral organs	VHL
Von Willebrand Disease (Types I, II and III)	blood		VWF
Wilson Disease	Various - Copper Storage Disease	Liver, brains, eyes, other tissues where copper builds up	ATP7B
Wiskott-Aldrich Syndrome	Immune System		WAS
Xeroderma Pigmentosum	Skin	Nervous system	POLH
XXX Syndrome	Endocrine, brain		X chromosome trisomy

[0636] In some embodiments, the compositions, systems, or components thereof can be used treat or prevent a disease in a subject by modifying one or more genes associated with one or more cellular functions, such as any one or more of those in **Table 3**. In some embodiments, the disease is a genetic disease or disorder. In some of embodiments, the composition, system, or component thereof can modify one or more genes or polynucleotides associated with one or more genetic diseases such as any set forth in **Table 3**.

Table 3. Exemplary Genes controlling Cellular Functions

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1

ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor Signaling	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA
Ephrin Receptor Signaling Actin Cytoskeleton Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK
	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3

Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation Signaling	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
Acute Phase Response Signaling	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 Signaling Aryl Hydrocarbon Receptor Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3
	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1

Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & Beta catenin Signaling	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6

Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative Stress Response	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic Stellate Cell Activation	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA
G-Protein Coupled Receptor Signaling	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA

Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S Checkpoint Regulation	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
Amyotrophic Lateral Sclerosis Signaling	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1
Nicotinate and Nicotinamide Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term Depression	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2

Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1

Cell Cycle: G2/M DNA Damage Checkpoint Regulation	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the Cardiovascular System	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction Notch Signaling	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3 HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum Stress Pathway Pyrimidine Metabolism	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3 NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
Cardiac & Beta Adrenergic Signaling	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid Metabolism	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair Pathway	ERCC5; ERCC4; XPA; XPC; ERCC1
Starch and Sucrose Metabolism	UCHL1; HK2; GCK; GPI; HK1
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor Signaling	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline Metabolism	ALDH1A1; NOS3; NOS2A
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose Metabolism	HK2; GCK; HK1
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR

Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid Metabolism	PRDX6; CHKA
Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics by Cytochrome p450	GSTP1; CYP1B1
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate Metabolism	PRMT5
Androgen and Estrogen Metabolism	PRMT5
Ascorbate and Aldarate Metabolism	ALDH1A1
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor Signaling	GNB2L1
NRF2-mediated Oxidative Stress Response	PRDX1
Pentose Phosphate Pathway	GPI
Pentose and Glucuronate Interconversions	UCHL1
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and Isoleucine Degradation	ALDH1A1
Glycine, Serine and Threonine Metabolism	CHKA
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[0637] In an aspect, the disclosure provides a method of individualized or personalized treatment of a genetic disease in a subject in need of such treatment comprising: (a) introducing one or more mutations *ex vivo* in a tissue, organ or a cell line, or *in vivo* in a transgenic non-human mammal, comprising delivering to cell(s) of the tissue, organ, cell or mammal a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment, wherein the specific mutations or precise sequence substitutions are or have been correlated to the genetic disease; (b) testing treatment(s) for the genetic disease on the cells to which the vector has been delivered that have the specific mutations or precise sequence substitutions correlated to the genetic disease; and (c) treating the subject based on results from the testing of treatment(s) of step (b).

Infectious Diseases

[0638] In some embodiments, the composition, system(s) or component(s) thereof can be used to diagnose, prognose, treat, and/or prevent an infectious disease caused by a microorganism, such as bacteria, virus, fungi, parasites, or combinations thereof.

[0639] In some embodiments, the system(s) or component(s) thereof can be capable of targeting specific microorganism within a mixed population. Exemplary methods of such techniques are described in e.g. Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL. 2014. Programmable removal of bacterial strains by use of genome-targeting composition, systems, mBio 5:e00928-13; Citorik RJ, Mimee M, Lu TK. 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32:1141–1145, the teachings of which can be adapted for use with the compositions, systems, and components thereof described herein.

[0640] In some embodiments, the composition, system(s) and/or components thereof can be capable of targeting pathogenic and/or drug-resistant microorganisms, such as bacteria, virus, parasites, and fungi. In some embodiments, the composition, system,(s) and/or components thereof can be capable of targeting and modifying one or more polynucleotides in a pathogenic microorganism such that the microorganism is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host cell.

[0641] In some embodiments, the pathogenic bacteria that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, those of the genus Actinomyces (e.g. *A. israelii*), Bacillus (e.g. *B. anthracis*, *B.*

cereus), Bactereoides (e.g. *B. fragilis*), Bartonella (*B. henselae, B. quintana*), Bordetella (*B. pertussis*), Borrelia (e.g. *B. burgdorferi, B. garinii, B. afzelii, and B. recurreentis*), Brucella (e.g. *B. abortus, B. canis, B. melitensis, and B. suis*), Campylobacter (e.g. *C. jejuni),* Chlamydia (e.g. *C. pneumoniae and C. trachomatis*), Chlamydophila (e.g. *C. psittaci*), Clostridium (e.g. *C. botulinum, C. difficile, C. perfringens. C. tetani*), Corynebacterium (e.g. *C. diphtheriae*), Enterococcus (e.g. *E. Faecalis, E. faecium*), Ehrlichia (*E. canis and E. chaffensis*) Escherichia (e.g. *E. coli),* Francisella (e.g. *F. tularensis*), Haemophilus (e.g. *H. influenzae*), Helicobacter (*H. pylori*), Klebsiella (E.g. *K. pneumoniae*), Legionella (e.g. *L. pneumophila*), Leptospira (e.g. *L. interrogans, L. santarosai, L. weilii, L. noguchii*), Listereia (e.g. *L. monocytogeenes*), Mycobacterium (e.g. *M. leprae, M. tuberculosis, M. ulcerans*), Mycoplasma (*M. pneumoniae*), Neisseria (*N. gonorrhoeae and N. menigitidis*), Nocardia (e.g. N. asteeroides), Pseudomonas (P. aeruginosa), Rickettsia (R. rickettsia), Salmonella (S. typhi and S. typhimurium), Shigella (S. sonnei and S. dysenteriae), Staphylococcus (S. aureus, S. epidermidis, and S. saprophyticus), Streeptococcus (S. agalactiaee, S. pneumoniae, S. pyogenes), Treponema (T. pallidum), Ureeaplasma (e.g. U. urealyticum), Vibrio (e.g. V. cholerae), Yersinia (e.g. Y. pestis, Y. enteerocolitica, and Y. pseudotuberculosis).

[0642] In some embodiments, the pathogenic virus that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, a double-stranded DNA virus, a partly double-stranded DNA virus, a single-stranded DNA virus, a positive single-stranded RNA virus, a negative single-stranded RNA virus, or a double stranded RNA virus. In some embodiments, the pathogenic virus can be from the family *Adenoviridae* (e.g. Adenovirus), *Herpeesviridae* (e.g. Herpes simplex, type 1, Herpes simplex, type 2, Varicella-zoster virus, Epstein–Barr virus, Human cytomegalovirus, Human herpesvirus, type 8), P*apillomaviridae* (e.g. Human papillomavirus), *Polyomaviridae* (e.g. BK virus, JC virus), *Poxviridae* (e.g. smallpox), *Hepadnaviridae* (e.g. Hepatitis B), *Parvoviridae* (e.g. Parvovirus B19), *Astroviridae* (e.g. Human astrovirus), *Caliciviridae* (e.g. Norwalk virus), *Picornaviridae* (e.g. coxsackievirus, hepatitis A virus, poliovirus, rhinovirus), *Coronaviridae* (e.g. Severe acute respiratory syndrome-related coronavirus, strains: Severe acute respiratory syndrome virus, Severe acute respiratory syndrome coronavirus 2 (COVID-19)), *Flaviviridae* (e.g. Hepatitis C virus, yellow fever virus, dengue virus, West Nile virus,TBE virus), Togaviridae (e.g. Rubella virus), *Hepeviridae* (e.g. Hepatitis E virus), *Retroviridae* (Human immunodeficiency virus (HIV)), *Orthomyxoviridae* (e.g. Influenza virus), *Arenaviridae* (e.g. Lassa virus), *Bunyaviridae* (e.g. Crimean-Congo hemorrhagic fever virus, Hantaan virus), *Filoviridae* (e.g. Ebola virus and Marburg virus), *Paramyxoviridae* (e.g.

Measles virus, Mumps virus, Parainfluenza virus, Respiratory syncytial virus), *Rhabdoviridae* (Rabies virus), Hepatits D virus, Reoviridae (e.g. Rotavirus, Orbivirus, Coltivirus, Banna virus).

[0643] In some embodiments, the pathogenic fungi that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, those of the genus *Candida* (e.g. *C. albicans*), *Aspergillus* (e.g. *A. fumigatus*, A. flavus, *A. clavatus*), *Cryptococcus* (e.g. *C. neoformans*, *C. gattii*), *Histoplasma* (e.g., *H. capsulatum*), *Pneumocystis* (e.g. *P. jiroveecii*), *Stachybotrys* (e.g. *S. chartarum*).

[0644] In some embodiments, the pathogenic parasites that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof described herein include, but are not limited to, protozoa, helminths, and ectoparasites. In some embodiments, the pathogenic protozoa that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, those from the groups Sarcodina (e.g. ameba such as Entamoeba), Mastigophora (e.g. flagellates such as Giardia and Leishmania), Cilophora (e.g. ciliates such as Balantidum), and sporozoa (e.g. plasmodium and cryptosporidium). In some embodiments, the pathogenic helminths that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof described herein include, but are not limited to, flatworms (platyhelminths), thorny-headed worms (acanthoceephalins), and roundworms (nematodes). In some embodiments, the pathogenic ectoparasites that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof described herein include, but are not limited to, ticks, fleas, lice, and mites.

[0645] In some embodiments, the pathogenic parasite that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, *Acanthamoeba spp., Balamuthia mandrillaris, Babesiosis spp. (e.g. Babesia B. divergens, B. bigemina, B. equi, B. microfti, B. duncani), Balantidiasis spp. (e.g. Balantidium coli), Blastocystis spp., Cryptosporidium spp., Cyclosporiasis spp. (e.g. Cyclospora cayetanensis), Dientamoebiasis spp. (e.g. Dientamoeba fragilis), Amoebiasis spp. (e.g. Entamoeba histolytica), Giardiasis spp. (e.g. Giardia lamblia), Isosporiasis spp. (e.g. Isospora belli), Leishmania spp., Naegleria spp. (e.g. Naegleria fowleri), Plasmodium spp. (e.g. Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale curtisi, Plasmodium ovale wallikeri, Plasmodium malariae, Plasmodium knowlesi), Rhinosporidiosis spp. (e.g. Rhinosporidium seeberi), Sarcocystosis spp. (e.g. Sarcocystis bovihominis, Sarcocystis suihominis), Toxoplasma spp. (e.g. Toxoplasma gondii), Trichomonas spp. (e.g. Trichomonas vaginalis), Trypanosoma spp. (e.g. Trypanosoma brucei), Trypanosoma spp. (e.g.*

Trypanosoma cruzi), Tapeworm (e.g. Cestoda, Taenia multiceps, Taenia saginata, Taenia solium), Diphyllobothrium latum spp., Echinococcus spp. (e.g. Echinococcus granulosus, Echinococcus multilocularis, E. vogeli, E. oligarthrus), Hymenolepis spp. (e.g. Hymenolepis nana, Hymenolepis diminuta), Bertiella spp. (e.g. Bertiella mucronata, Bertiella studeri), Spirometra (e.g. Spirometra erinaceieuropaei), Clonorchis spp. (e.g. Clonorchis sinensis; Clonorchis viverrini), Dicrocoelium spp. (e.g. Dicrocoelium dendriticum), Fasciola spp. (e.g. Fasciola hepatica, Fasciola gigantica), Fasciolopsis spp. (e.g. Fasciolopsis buski), Metagonimus spp. (e.g. Metagonimus yokogawai), Metorchis spp. (e.g. Metorchis conjunctus), Opisthorchis spp. (e.g. Opisthorchis viverrini, Opisthorchis felineus), Clonorchis spp. (e.g. Clonorchis sinensis), Paragonimus spp. (e.g. Paragonimus westermani; Paragonimus africanus; Paragonimus caliensis; Paragonimus kellicotti; Paragonimus skrjabini; Paragonimus uterobilateralis), Schistosoma sp., Schistosoma spp. (e.g. Schistosoma mansoni, Schistosoma haematobium, Schistosoma japonicum, Schistosoma mekongi, and Schistosoma intercalatum), Echinostoma spp. (e.g. E. echinatum), Trichobilharzia spp. (e.g. Trichobilharzia regent), Ancylostoma spp. (e.g. Ancylostoma duodenale), Necator spp. (e.g. Necator americanus), Angiostrongylus spp., Anisakis spp., Ascaris spp. (e.g. Ascaris lumbricoides), Baylisascaris spp. (e.g. Baylisascaris procyonis), Brugia spp. (e.g. Brugia malayi, Brugia timori), Dioctophyme spp. (e.g. Dioctophyme renale), Dracunculus spp. (e.g. Dracunculus medinensis), Enterobius spp. (e.g. Enterobius vermicularis, Enterobius gregorii), Gnathostoma spp. (e.g. Gnathostoma spinigerum, Gnathostoma hispidum), Halicephalobus spp. (e.g. Halicephalobus gingivalis), Loa loa spp. (e.g. Loa loa filaria), Mansonella spp. (e.g. Mansonella streptocerca), Onchocerca spp. (e.g. Onchocerca volvulus), Strongyloides spp. (e.g. Strongyloides stercoralis), Thelazia spp. (e.g. Thelazia californiensis, Thelazia callipaeda), Toxocara spp. (e.g. Toxocara canis, Toxocara cati, Toxascaris leonine), Trichinella spp. (e.g. Trichinella spiralis, Trichinella britovi, Trichinella nelsoni, Trichinella nativa), Trichuris spp. (e.g. Trichuris trichiura, Trichuris vulpis), Wuchereria spp. (e.g. Wuchereria bancrofti), Dermatobia spp. (e.g. Dermatobia hominis), Tunga spp. (e.g. Tunga penetrans), Cochliomyia spp. (e.g. Cochliomyia hominivorax), Linguatula spp. (e.g. Linguatula serrata), Archiacanthocephala sp., Moniliformis sp. (e.g. Moniliformis moniliformis), Pediculus spp. (e.g. Pediculus humanus capitis, Pediculus humanus humanus), Pthirus spp. (e.g. Pthirus pubis), Arachnida spp. (e.g. Trombiculidae, Ixodidae, Argaside), Siphonaptera spp (e.g. Siphonaptera: Pulicinae), Cimicidae spp. (e.g. Cimex lectularius and Cimex hemipterus), Diptera spp., Demodex spp. (e.g. Demodex folliculorum/brevis/canis), Sarcoptes spp. (e.g. Sarcoptes scabiei), Dermanyssus spp. (e.g. Dermanyssus gallinae),

Ornithonyssus spp. (e.g. Ornithonyssus sylviarum, Ornithonyssus bursa, Ornithonyssus bacoti), Laelaps spp. (e.g. Laelaps echidnina), Liponyssoides spp. (e.g. Liponyssoides sanguineus).

[0646] In some embodiments, the gene targets can be any of those as set forth in Table 1 of Strich and Chertow. 2019. J. Clin. Microbio. 57:4 e01307-18, which is incorporated herein as if expressed in its entirety herein.

[0647] In some embodiments, the method can include delivering a composition, system, and/or component thereof to a pathogenic organism described herein, allowing the composition, system, and/or component thereof to specifically bind and modify one or more targets in the pathogenic organism, whereby the modification kills, inhibits, reduces the pathogenicity of the pathogenic organism, or otherwise renders the pathogenic organism non-pathogenic. In some embodiments, delivery of the composition, system, occurs *in vivo* (i.e. in the subject being treated). In some embodiments, delivery occurs by an intermediary, such as microorganism or phage that is non-pathogenic to the subject but is capable of transferring polynucleotides and/or infecting the pathogenic microorganism. In some embodiments, the intermediary microorganism can be an engineered bacteria, virus, or phage that contains the composition, system(s) and/or component(s) thereof and/or vectors and/or vector systems. The method can include administering an intermediary microorganism containing the composition, system(s) and/or component(s) thereof and/or vectors and/or vector systems to the subject to be treated. The intermediary microorganism can then produce the system and/or component thereof or transfer a composition, system, polynucleotide to the pathogenic organism. In embodiments, where the system and/or component thereof, vector, or vector system is transferred to the pathogenic microorganism, the composition, system, or component thereof is then produced in the pathogenic microorganism and modifies the pathogenic microorganism such that it is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host or cell thereof.

[0648] In some embodiments, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g. a virus), the composition, system can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA cannot be replicated by the host cell's machinery into a functional virus. In some embodiments, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g. a virus), the composition, system can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA is deleted from the host cell's genome.

[0649] It will be appreciated that inhibiting or killing the pathogenic microorganism, the disease and/or condition that its infection causes in the subject can be treated or prevented. Thus, also provided herein are methods of treating and/or preventing one or more diseases or symptoms thereof caused by any one or more pathogenic microorganisms, such as any of those described herein.

Mitochondrial Diseases

[0650] Some of the most challenging mitochondrial disorders arise from mutations in mitochondrial DNA (mtDNA), a high copy number genome that is maternally inherited. In some embodiments, mtDNA mutations can be modified using a composition, system, described herein. In some embodiments, the mitochondrial disease that can be diagnosed, prognosed, treated, and/or prevented can be MELAS (mitochondrial myopathy encephalopathy, and lactic acidosis and stroke-like episodes), CPEO/PEO (chronic progressive external ophthalmoplegia syndrome/progressive external ophthalmoplegia), KSS (Kearns-Sayre syndrome), MIDD (maternally inherited diabetes and deafness), MERRF (myoclonic epilepsy associated with ragged red fibers), NIDDM (noninsulin-dependent diabetes mellitus), LHON (Leber hereditary optic neuropathy), LS (Leigh Syndrome) an aminoglycoside induced hearing disorder, NARP (neuropathy, ataxia, and pigmentary retinopathy), Extrapyramidal disorder with akinesia-rigidity, psychosis and SNHL, Nonsyndromic hearing loss a cardiomyopathy, an encephalomyopathy, Pearson's syndrome, or a combination thereof.

[0651] In some embodiments, the mtDNA of a subject can be modified *in vivo* or *ex vivo*. In some embodiments, where the mtDNA is modified *ex vivo*, after modification the cells containing the modified mitochondria can be administered back to the subject. In some embodiments, the composition, system, or component thereof can be capable of correcting an mtDNA mutation, or a combination thereof.

[0652] In some embodiments, at least one of the one or more mtDNA mutations is selected from the group consisting of: A3243G, C3256T, T3271C, G1019A, A1304T, A15533G, C1494T, C4467A, T1658C, G12315A, A3421G, A8344G, T8356C, G8363A, A13042T, T3200C, G3242A, A3252G, T3264C, G3316A, T3394C, T14577C, A4833G, G3460A, G9804A, G11778A, G14459A, A14484G, G15257A, T8993C, T8993G, G10197A, G13513A, T1095C, C1494T, A1555G, G1541A, C1634T, A3260G, A4269G, T7587C, A8296G, A8348G, G8363A, T9957C, T9997C, G12192A, C12297T, A14484G, G15059A, duplication of CCCCCTCCCC-tandem repeats at positions 305-314 and/or 956-965, deletion at positions from 8,469-13,447, 4,308-14,874, and/or 4,398-14,822, 961ins/delC, the mitochondrial common deletion (e.g. mtDNA 4,977 bp deletion), and combinations thereof.

[0653] In some embodiments, the mitochondrial mutation can be any mutation as set forth in or as identified by use of one or more bioinformatic tools available at Mitomap available at mitomap.org. Such tools include, but are not limited to, "Variant Search, aka Market Finder", Find Sequences for Any Haplogroup, aka "Sequence Finder", "Variant Info", "POLG Pathogenicity Prediction Server", "MITOMASTER", "Allele Search", "Sequence and Variant Downloads", "Data Downloads". MitoMap contains reports of mutations in mtDNA that can be associated with disease and maintains a database of reported mitochondrial DNA Base Substitution Diseases: rRNA/tRNA mutations.

[0654] In some embodiments, the method includes delivering a composition, system, and/or a component thereof to a cell, and more specifically one or more mitochondria in a cell, allowing the composition, system, and/or component thereof to modify one or more target polynucleotides in the cell, and more specifically one or more mitochondria in the cell. The target polynucleotides can correspond to a mutation in the mtDNA, such as any one or more of those described herein. In some embodiments, the modification can alter a function of the mitochondria such that the mitochondria functions normally or at least is/are less dysfunctional as compared to an unmodified mitochondria. Modification can occur *in vivo* or *ex vivo*. Where modification is performed *ex vivo*, cells containing modified mitochondria can be administered to a subject in need thereof in an autologous or allogenic manner.

Microbiome Modification

[0655] Microbiomes play important roles in health and disease. For example, the gut microbiome can play a role in health by controlling digestion, preventing growth of pathogenic microorganisms and have been suggested to influence mood and emotion. Imbalanced microbiomes can promote disease and are suggested to contribute to weight gain, unregulated blood sugar, high cholesterol, cancer, and other disorders. A healthy microbiome has a series of joint characteristics that can be distinguished from non-healthy individuals; thus detection and identification of the disease-associated microbiome can be used to diagnose and detect disease in an individual. The compositions, systems, and components thereof can be used to screen the microbiome cell population and be used to identify a disease associated microbiome. Cell screening methods utilizing compositions, systems, and components thereof are described elsewhere herein and can be applied to screening a microbiome, such as a gut, skin, vagina, and/or oral microbiome, of a subject.

[0656] In some embodiments, the microbe population of a microbiome in a subject can be modified using a composition, system, and/or component thereof described herein. In some embodiments, the composition, system, and/or component thereof can be used to identify and

select one or more cell types in the microbiome and remove them from the microbiome population. Exemplary methods of selecting cells using a composition, system, and/or component thereof are described elsewhere herein. In this way, the make-up or microorganism profile of the microbiome can be altered. In some embodiments, the alteration causes a change from a diseased microbiome composition to a healthy microbiome composition. In this way the ratio of one type or species of microorganism to another can be modified, such as going from a diseased ratio to a healthy ratio. In some embodiments, the cells selected are pathogenic microorganisms.

[0657] In some embodiments, the compositions and systems described herein can be used to modify a polynucleotide in a microorganism of a microbiome in a subject. In some embodiments, the microorganism is a pathogenic microorganism. In some embodiments, the microorganism is a commensal and non-pathogenic microorganism. Methods of modifying polynucleotides in a cell in the subject are described elsewhere herein and can be applied to these embodiments.

Models of Diseases and Conditions

[0658] In an aspect, the disclosure provides a method of modeling a disease associated with a genomic locus in a eukaryotic organism or a non-human organism comprising manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus comprising delivering a non- naturally occurring or engineered composition comprising a viral vector system comprising one or more viral vectors operably encoding a composition for expression thereof, wherein the composition comprises particle delivery system or the delivery system or the virus particle of any one of the above embodiments or the cell of any one of the above embodiment.

[0659] In one aspect, the disclosure provides a method of generating a model eukaryotic cell that can include one or more a mutated disease genes and/or infectious microorganisms. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method includes (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors comprise a composition, system, and/or component thereof and/or a vector or vector system that is capable of driving expression of a composition, system, and/or component thereof.

[0660] The disease modeled can be any disease with a genetic or epigenetic component. In some embodiments, the disease modeled can be any as discussed elsewhere herein, including but not limited to any as set forth in Tables 4 and 5 herein.

In situ Disease Detection

[0661] The compositions, systems, and/or components thereof can be used for diagnostic methods of detection such as in CASFISH (see e.g. Deng et al. 2015. PNAS USA 112(38): 11870-11875), CRISPR-Live FISH (see e.g. Wang et al. 2020. Science; 365(6459):1301-1305), sm-FISH (Lee and Jefcoate. 2017. Front. Endocrinol. doi.org/10.3389/fendo.2017.00289), sequential FISH CRISPRainbow (Ma et al. Nat Biotechnol, 34 (2016), pp. 528-530), CRISPR-Sirius (Nat Methods, 15 (2018), pp. 928-931), Casilio (Cheng et al. Cell Res, 26 (2016), pp. 254-257), Halo-Tag based genomic loci visualization techniques (e.g. Deng et al. 2015. PNAS USA 112(38): 11870-11875; Knight et al., Science, 350 (2015), pp. 823-826), RNA-aptamer based methods (e.g. Ma et al., J Cell Biol, 214 (2016), pp. 529-537), molecular beacon-based methods (e.g. Zhao et al. Biomaterials, 100 (2016), pp. 172-183; Wu et al. Nucleic Acids Res (2018)), Quantum Dot-based systems (e.g. Ma et al. Anal Chem, 89 (2017), pp. 12896-12901), multiplexed methods (e.g. Ma et al., Proc Natl Acad Sci U S A, 112 (2015), pp. 3002-3007; Fu et al. Nat Commun, 7 (2016), p. 11707; Ma et al. Nat Biotechnol, 34 (2016), pp. 528-530; Shao et al. Nucleic Acids Res, 44 (2016), Article e86); Wang et al. Sci Rep, 6 (2016), p. 26857), ç, and other *in situ* CRISPR-hybridization based methods (e.g. Chen et al. Cell, 155 (2013), pp. 1479-1491; Gu et al. Science, 359 (2018), pp. 1050-1055; Tanebaum et al. Cell, 159 (2014), pp. 635-646; Ye et al. Protein Cell, 8 (2017), pp. 853-855; Chen et al. Nat Commun, 9 (2018), p. 5065; Shao et al. ACS Synth Biol (2017); Fu et al. Nat Commun, 7 (2016), p. 11707; Shao et al. Nucleic Acids Res, 44 (2016), Article e86; Wang et al., Sci Rep, 6 (2016), p. 26857), all of which are incorporated by reference herein as if expressed in their entirety and whose teachings can be adapted to the compositions, systems, and components thereof described herein in view of the description herein.

[0662] In some embodiments, the composition, system, or component thereof can be used in a detection method, such as an *in situ* detection method described herein. In some embodiments, the composition, system, or component thereof can include a catalytically inactivate Cas effector described herein and use this system in detection methods such as fluorescence *in situ* hybridization (FISH) or any other described herein. In some embodiments, the inactivated Cas effector, which lacks the ability to produce DNA double-strand breaks may be fused with a marker, such as fluorescent protein, such as the enhanced green fluorescent protein (eEGFP) and co-expressed with small guide RNAs to target pericentric, centric and telomeric repeats *in vivo*. The dCas effector or system thereof can be used to visualize both repetitive sequences and individual genes in the human genome. Such new applications of

labelled dCas effector and compositions, systems thereof can be important in imaging cells and studying the functional nuclear architecture, especially in cases with a small nucleus volume or complex 3-D structures.

Cell Selection

[0663] In some embodiments, the compositions, systems, and/or components thereof described herein can be used in a method to screen and/or select cells. In some embodiments, composition, system-based screening/selection method can be used to identify diseased cells in a cell population. In some embodiments, selection of the cells results in a modification in the cells such that the selected cells die. In this way, diseased cells can be identified and removed from the healthy cell population. In some embodiments, the diseased cells can be a cancer cell, pre-cancerous cell, a virus or other pathogenic organism infected cells, or otherwise abnormal cell. In some embodiments, the modification can impart another detectable change in the cells to be selected (e.g. a functional change and/or genomic barcode) that facilitates selection of the desired cells. In some embodiments a negative selection scheme can be used to obtain a desired cell population. In these embodiments, the cells to be selected against are modified, thus can be removed from the cell population based on their death or identification or sorting based the detectable change imparted on the cells. Thus, in these embodiments, the remaining cells after selection are the desired cell population.

[0664] In some embodiments, a method of selecting one or more cell(s) containing a polynucleotide modification can include introducing one or more composition, system,(s) and/or components thereof, and/or vectors or vector systems into the cell(s), wherein the composition, system,(s) and/or components thereof, and/or vectors or vector systems Therapeutic Agent Development

[0665] In some embodiments, the method involves developing a therapeutic based on the composition, system, described herein. In particular embodiments, the therapeutic comprises a Cas effector and/or a guide RNA capable of hybridizing to a target sequence of interest. In particular embodiments, the therapeutic is a vector or vector system that can contain a) a first regulatory element operably linked to a nucleotide sequence encoding the Cas effector protein(s); and b) a second regulatory element operably linked to one or more nucleotide sequences encoding one or more nucleic acid molecules comprising a guide RNA comprising a guide sequence, a direct repeat sequence; wherein components (a) and (b) are located on same or different vectors. In particular embodiments, the biologically active agent is a composition comprising a delivery system operably configured to deliver composition, system, or components thereof, and/or or one or more polynucleotide sequences, vectors, or vector

systems containing or encoding said components into a cell and capable of forming a complex with the components of the composition and system herein, and wherein said complex is operable in the cell. In some embodiments, the complex can include the Cas effector protein(s) as described herein, guide RNA comprising the guide sequence, and a direct repeat sequence. In any such compositions, the delivery system can be a yeast system, a lipofection system, a microinjection system, a biolistic system, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates or artificial virions, or any other system as described herein. In particular embodiments, the delivery is via a particle, a nanoparticle, a lipid or a cell penetrating peptide (CPP).

[0666] Also described herein are methods for developing or designing a composition, system, optionally a composition, system, based therapy or therapeutic, comprising (a) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, and optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more gRNA recognizing one or more of said (sub)selected target sites.

[0667] In some embodiments, the method for developing or designing a gRNA for use in a composition, system, optionally a composition, system, based therapy or therapeutic, can include (a) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of

the (sub)selected target sites for an individual subject, optionally designing one or more gRNA recognizing one or more of said (sub)selected target sites.

[0668] In some embodiments, the method for developing or designing a composition, system, optionally a composition, system, based therapy or therapeutic in a population can include (a) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more gRNA recognizing one or more of said (sub)selected target sites.

[0669] In some embodiments the method for developing or designing a gRNA for use in a composition, system, optionally a composition, system, based therapy or therapeutic in a population, can include (a) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more gRNA recognizing one or more of said (sub)selected target sites.

[0670] In some embodiments, the method for developing or designing a composition, system, such as a composition, system, based therapy or therapeutic, optionally in a population; or for developing or designing a gRNA for use in a composition, system, optionally a composition, system, based therapy or therapeutic, optionally in a population, can include selecting a set of target sequences for one or more loci in a target population, wherein the target sequences do not contain variants occurring above a threshold allele frequency in the target

population (i.e. platinum target sequences); removing from said selected (platinum) target sequences any target sequences having high frequency off-target candidates (relative to other (platinum) targets in the set) to define a final target sequence set; preparing one or more, such as a set of compositions, systems, based on the final target sequence set, optionally wherein a number of CRISP-Cas systems prepared is based (at least in part) on the size of a target population.

[0671] In certain embodiments, off-target candidates/off-targets, PAM restrictiveness, target cleavage efficiency, or effector protein specificity is identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as described herein elsewhere. In certain embodiments, off-target candidates/off-targets are identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as described herein elsewhere. In certain embodiments, off-targets, or off target candidates have at least 1, preferably 1-3, mismatches or (distal) PAM mismatches, such as 1 or more, such as 1, 2, 3, or more (distal) PAM mismatches. In certain embodiments, sequencing-based DSB detection assay comprises labeling a site of a DSB with an adapter comprising a primer binding site, labeling a site of a DSB with a barcode or unique molecular identifier, or combination thereof, as described herein elsewhere.

[0672] It will be understood that the guide sequence of the gRNA is 100% complementary to the target site, i.e. does not comprise any mismatch with the target site. It will be further understood that "recognition" of an (off-)target site by a gRNA presupposes composition, system, functionality, i.e. an (off-)target site is only recognized by a gRNA if binding of the gRNA to the (off-)target site leads to composition, system, activity (such as induction of single or double strand DNA cleavage, transcriptional modulation, etc.).

[0673] In certain embodiments, the target sites having minimal sequence variation across a population are characterized by absence of sequence variation in at least 99%, preferably at least 99.9%, more preferably at least 99.99% of the population. In certain embodiments, optimizing target location comprises selecting target sequences or loci having an absence of sequence variation in at least 99%, %, preferably at least 99.9%, more preferably at least 99.99% of a population. These targets are referred to herein elsewhere also as "platinum targets". In certain embodiments, said population comprises at least 1,000 individuals, such as at least 5,000 individuals, such as at least 10,000 individuals, such as at least 50,000 individuals.

[0674] In certain embodiments, the off-target sites are characterized by at least one mismatch between the off-target site and the gRNA. In certain embodiments, the off-target

sites are characterized by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the gRNA. In certain embodiments, the off-target sites are characterized by at least one mismatch between the off-target site and the gRNA and by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the gRNA.

[0675] In certain embodiments, said minimal number of off-target sites across said population is determined for high-frequency haplotypes in said population. In certain embodiments, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the off-target site locus in said population. In certain embodiments, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the target site locus in said population. In certain embodiments, the high-frequency haplotypes are characterized by occurrence in at least 0.1% of the population.

[0676] In certain embodiments, the number of (sub)selected target sites needed to treat a population is estimated based on based low frequency sequence variation, such as low frequency sequence variation captured in large scale sequencing datasets. In certain embodiments, the number of (sub)selected target sites needed to treat a population of a given size is estimated.

[0677] In certain embodiments, the method further comprises obtaining genome sequencing data of a subject to be treated; and treating the subject with a composition, system, selected from the set of compositions, systems, wherein the composition, system, selected is based (at least in part) on the genome sequencing data of the individual. In certain embodiments, the ((sub)selected) target is validated by genome sequencing, preferably whole genome sequencing.

[0678] In certain embodiments, target sequences or loci as described herein are (further) selected based on optimization of one or more parameters, such as PAM type (natural or modified), PAM nucleotide content, PAM length, target sequence length, PAM restrictiveness, target cleavage efficiency, and target sequence position within a gene, a locus or other genomic region. Methods of optimization are discussed in greater detail elsewhere herein.

[0679] In certain embodiments, target sequences or loci as described herein are (further) selected based on optimization of one or more of target loci location, target length, target specificity, and PAM characteristics. As used herein, PAM characteristics may comprise for instance PAM sequence, PAM length, and/or PAM GC contents. In certain embodiments, optimizing PAM characteristics comprises optimizing nucleotide content of a PAM. In certain

embodiments, optimizing nucleotide content of PAM is selecting a PAM with a motif that maximizes abundance in the one or more target loci, minimizes mutation frequency, or both. Minimizing mutation frequency can for instance be achieved by selecting PAM sequences devoid of or having low or minimal CpG.

[0680] In certain embodiments, the effector protein for each composition and system, in the set of compositions, systems, is selected based on optimization of one or more parameters selected from the group consisting of; effector protein size, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, effector protein specificity, effector protein stability or half-life, effector protein immunogenicity or toxicity. Methods of optimization are discussed in greater detail elsewhere herein.

Optimization of the Systems

[0681] The methods of the present disclosure can involve optimization of selected parameters or variables associated with the composition, system, and/or its functionality, as described herein further elsewhere. Optimization of the composition, system, in the methods as described herein may depend on the target(s), such as the therapeutic target or therapeutic targets, the mode or type of composition, system, modulation, such as composition, system, based therapeutic target(s) modulation, modification, or manipulation, as well as the delivery of the composition, system, components. One or more targets may be selected, depending on the genotypic and/or phenotypic outcome. For instance, one or more therapeutic targets may be selected, depending on (genetic) disease etiology or the desired therapeutic outcome. The (therapeutic) target(s) may be a single gene, locus, or other genomic site, or may be multiple genes, loci or other genomic sites. As is known in the art, a single gene, locus, or other genomic site may be targeted more than once, such as by use of multiple gRNAs.

[0682] The activity of the composition and/or system, such as therapy or therapeutics may involve target disruption, such as target mutation, such as leading to gene knockout. The activity of the composition and/or system, such as therapy or therapeutics may involve replacement of particular target sites, such as leading to target correction. Therapy or therapeutics may involve removal of particular target sites, such as leading to target deletion. The activity of the composition and/or system, such as therapy or therapeutics may involve modulation of target site functionality, such as target site activity or accessibility, leading for instance to (transcriptional and/or epigenetic) gene or genomic region activation or gene or genomic region silencing.

[0683] Accordingly, in an aspect, the disclosure relates to a method as described herein, comprising selection of one or more (therapeutic) target, selecting one or more functionality of the composition and/or system, and optimization of selected parameters or variables associated with the system and/or its functionality. In a related aspect, the disclosure relates to a method as described herein, comprising (a) selecting one or more (therapeutic) target loci, (b) selecting one or more system functionalities, (c) optionally selecting one or more modes of delivery, and preparing, developing, or designing a system selected based on steps (a)-(c).

[0684] In certain embodiments, the functionality of the composition and/or system comprises genomic mutation. In certain embodiments, the functionality of the composition and/or system comprises single genomic mutation. In certain embodiments, the functionality of the composition and/or system functionality comprises multiple genomic mutation. In certain embodiments, the functionality of the composition and/or system comprises gene knockout. In certain embodiments, the functionality of the composition and/or system comprises single gene knockout. In certain embodiments, the functionality of the composition and/or system comprises multiple gene knockout. In certain embodiments, the functionality of the composition and/or system comprises gene correction. In certain embodiments, the functionality of the composition and/or system comprises single gene correction. In certain embodiments, the functionality of the composition and/or system comprises multiple gene correction. In certain embodiments, the functionality of the composition and/or system comprises genomic region correction. In certain embodiments, the functionality of the composition and/or system comprises single genomic region correction. In certain embodiments, the functionality of the composition and/or system comprises multiple genomic region correction. In certain embodiments, the functionality of the composition and/or system comprises gene deletion. In certain embodiments, the functionality of the composition and/or system comprises single gene deletion. In certain embodiments, the functionality of the composition and/or system comprises multiple gene deletion. In certain embodiments, the functionality of the composition and/or system comprises genomic region deletion. In certain embodiments, the functionality of the composition and/or system comprises single genomic region deletion. In certain embodiments, the functionality of the composition and/or system comprises multiple genomic region deletion. In certain embodiments, the functionality of the composition and/or system comprises modulation of gene or genomic region functionality. In certain embodiments, the functionality of the composition and/or system comprises modulation of single gene or genomic region functionality. In certain embodiments, the functionality of the composition and/or system comprises modulation of multiple gene or genomic region

functionality. In certain embodiments, the functionality of the composition and/or system comprises gene or genomic region functionality, such as gene or genomic region activity. In certain embodiments, the functionality of the composition and/or system comprises single gene or genomic region functionality, such as gene or genomic region activity. In certain embodiments, the functionality of the composition and/or system comprises multiple gene or genomic region functionality, such as gene or genomic region activity. In certain embodiments, the functionality of the composition and/or system comprises modulation gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing. In certain embodiments, the functionality of the composition and/or system comprises modulation single gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing. In certain embodiments, the functionality of the composition and/or system comprises modulation multiple gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing.

[0685] Optimization of selected parameters or variables in the methods as described herein may result in optimized or improved the system, such as CRISPR-Cas system-based therapy or therapeutic, specificity, efficacy, and/or safety. In certain embodiments, one or more of the following parameters or variables are taken into account, are selected, or are optimized in the methods of the disclosure as described herein: Cas protein allosteric interactions, Cas protein functional domains and functional domain interactions, CRISPR effector specificity, gRNA specificity, CRISPR-Cas complex specificity, PAM restrictiveness, PAM type (natural or modified), PAM nucleotide content, PAM length, CRISPR effector activity, gRNA activity, CRISPR-Cas complex activity, target cleavage efficiency, target site selection, target sequence length, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, CRISPR effector stability, CRISPR effector mRNA stability, gRNA stability, CRISPR-Cas complex stability, CRISPR effector protein or mRNA immunogenicity or toxicity, gRNA immunogenicity or toxicity, CRISPR-Cas complex immunogenicity or toxicity, CRISPR effector protein or mRNA dose or titer, gRNA dose or titer, CRISPR-Cas complex dose or titer, CRISPR effector protein size, CRISPR effector expression level, gRNA expression level, CRISPR-Cas complex expression level, CRISPR effector spatiotemporal expression, gRNA spatiotemporal expression, CRISPR-Cas complex spatiotemporal expression.

[0686] By means of example, and without limitation, parameter or variable optimization may be achieved as follows. CRISPR effector specificity may be optimized by selecting the most specific CRISPR effector. This may be achieved for instance by selecting the most specific CRISPR effector orthologue or by specific CRISPR effector mutations which increase specificity. gRNA specificity may be optimized by selecting the most specific gRNA. This can be achieved for instance by selecting gRNA having low homology, i.e. at least one or preferably more, such as at least 2, or preferably at least 3, mismatches to off-target sites. CRISPR-Cas complex specificity may be optimized by increasing CRISPR effector specificity and/or gRNA specificity as above. PAM restrictiveness may be optimized by selecting a CRISPR effector having to most restrictive PAM recognition. This can be achieved for instance by selecting a CRISPR effector orthologue having more restrictive PAM recognition or by specific CRISPR effector mutations which increase or alter PAM restrictiveness. PAM type may be optimized for instance by selecting the appropriate CRISPR effector, such as the appropriate CRISPR effector recognizing a desired PAM type. The CRISPR effector or PAM type may be naturally occurring or may for instance be optimized based on CRISPR effector mutants having an altered PAM recognition, or PAM recognition repertoire. PAM nucleotide content may for instance be optimized by selecting the appropriate CRISPR effector, such as the appropriate CRISPR effector recognizing a desired PAM nucleotide content. The CRISPR effector or PAM type may be naturally occurring or may for instance be optimized based on CRISPR effector mutants having an altered PAM recognition, or PAM recognition repertoire. PAM length may for instance be optimized by selecting the appropriate CRISPR effector, such as the appropriate CRISPR effector recognizing a desired PAM nucleotide length. The CRISPR effector or PAM type may be naturally occurring or may for instance be optimized based on CRISPR effector mutants having an altered PAM recognition, or PAM recognition repertoire.

[0687] Target length or target sequence length may be optimized, for instance, by selecting the appropriate CRISPR effector, such as the appropriate CRISPR effector recognizing a desired target or target sequence nucleotide length. Alternatively, or in addition, the target (sequence) length may be optimized by providing a target having a length deviating from the target (sequence) length typically associated with the CRISPR effector, such as the naturally occurring CRISPR effector. The CRISPR effector or target (sequence) length may be naturally occurring or may for instance be optimized based on CRISPR effector mutants having an altered target (sequence) length recognition, or target (sequence) length recognition repertoire. For instance, increasing or decreasing target (sequence) length may influence target recognition and/or off-target recognition. CRISPR effector activity may be optimized by selecting the most

active CRISPR effector. This may be achieved for instance by selecting the most active CRISPR effector orthologue or by specific CRISPR effector mutations which increase activity. The ability of the CRISPR effector protein to access regions of high chromatin accessibility, may be optimized by selecting the appropriate CRISPR effector or mutant thereof, and can consider the size of the CRISPR effector, charge, or other dimensional variables etc. The degree of uniform CRISPR effector activity may be optimized by selecting the appropriate CRISPR effector or mutant thereof, and can consider CRISPR effector specificity and/or activity, PAM specificity, target length, mismatch tolerance, epigenetic tolerance, CRISPR effector and/or gRNA stability and/or half-life, CRISPR effector and/or gRNA immunogenicity and/or toxicity, etc. gRNA activity may be optimized by selecting the most active gRNA. In some embodiments, this can be achieved by increasing gRNA stability through RNA modification. CRISPR-Cas complex activity may be optimized by increasing CRISPR effector activity and/or gRNA activity as above.

[0688] The target site selection may be optimized by selecting the optimal position of the target site within a gene, locus or other genomic region. The target site selection may be optimized by optimizing target location comprises selecting a target sequence with a gene, locus, or other genomic region having low variability. This may be achieved for instance by selecting a target site in an early and/or conserved exon or domain (i.e. having low variability, such as polymorphisms, within a population).

[0689] In certain embodiments, optimizing target (sequence) length comprises selecting a target sequence within one or more target loci between 5 and 25 nucleotides. In certain embodiments, a target sequence is 20 nucleotides.

[0690] In certain embodiments, optimizing target specificity comprises selecting targets loci that minimize off-target candidates.

[0691] In some embodiments, the target site may be selected by minimization of off-target effects (e.g. off-targets qualified as having 1-5, 1-4, or preferably 1-3 mismatches compared to target and/or having one or more PAM mismatches, such as distal PAM mismatches), preferably also considering variability within a population. CRISPR effector stability may be optimized by selecting CRISPR effector having appropriate half-life, such as preferably a short half-life while still capable of maintaining sufficient activity. In some embodiments, this can be achieved by selecting an appropriate CRISPR effector orthologue having a specific half-life or by specific CRISPR effector mutations or modifications which affect half-life or stability, such as inclusion (e.g. fusion) of stabilizing or destabilizing domains or sequences. CRISPR effector mRNA stability may be optimized by increasing or decreasing CRISPR effector

mRNA stability. In some embodiments, this can be achieved by increasing or decreasing CRISPR effector mRNA stability through mRNA modification. gRNA stability may be optimized by increasing or decreasing gRNA stability. In some embodiments, this can be achieved by increasing or decreasing gRNA stability through RNA modification. CRISPR-Cas complex stability may be optimized by increasing or decreasing CRISPR effector stability and/or gRNA stability as above. CRISPR effector protein or mRNA immunogenicity or toxicity may be optimized by decreasing CRISPR effector protein or mRNA immunogenicity or toxicity. In some embodiments, this can be achieved by mRNA or protein modifications. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. gRNA immunogenicity or toxicity may be optimized by decreasing gRNA immunogenicity or toxicity. In some embodiments, this can be achieved by gRNA modifications. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. CRISPR-Cas complex immunogenicity or toxicity may be optimized by decreasing CRISPR effector immunogenicity or toxicity and/or gRNA immunogenicity or toxicity as above, or by selecting the least immunogenic or toxic CRISPR effector/gRNA combination. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. CRISPR effector protein or mRNA dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. gRNA dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. CRISPR-Cas complex dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. CRISPR effector protein size may be optimized by selecting minimal protein size to increase efficiency of delivery, in particular for virus mediated delivery. CRISPR effector, gRNA, or CRISPR-Cas complex expression level may be optimized by limiting (or extending) the duration of expression and/or limiting (or increasing) expression level. This may be achieved for instance by using self-inactivating compositions, systems,, such as including a self-targeting (e.g. CRISPR effector targeting) gRNA, by using viral vectors having limited expression duration, by using appropriate promoters for low (or high) expression levels, by combining different delivery methods for individual CRISP-Cas system components, such as virus mediated delivery of CRISPR-effector encoding nucleic acid combined with non-virus mediated delivery of gRNA, or virus mediated delivery of gRNA combined with non-virus mediated delivery of CRISPR effector protein or mRNA. CRISPR effector, gRNA, or CRISPR-Cas complex spatiotemporal expression may be optimized by appropriate choice of conditional and/or inducible expression systems, including controllable CRISPR effector

activity optionally a destabilized CRISPR effector and/or a split CRISPR effector, and/or cell- or tissue-specific expression systems.

[0692] In an aspect, the disclosure relates to a method as described herein, comprising selection of one or more (therapeutic) target, selecting the functionality of the composition and/or system, selecting mode of delivery, selecting delivery vehicle or expression system, and optimization of selected parameters or variables associated with the system and/or its functionality, optionally wherein the parameters or variables are one or more selected from CRISPR effector specificity, gRNA specificity, CRISPR-Cas complex specificity, PAM restrictiveness, PAM type (natural or modified), PAM nucleotide content, PAM length, CRISPR effector activity, gRNA activity, CRISPR-Cas complex activity, target cleavage efficiency, target site selection, target sequence length, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, CRISPR effector stability, CRISPR effector mRNA stability, gRNA stability, CRISPR-Cas complex stability, CRISPR effector protein or mRNA immunogenicity or toxicity, gRNA immunogenicity or toxicity, CRISPR-Cas complex immunogenicity or toxicity, CRISPR effector protein or mRNA dose or titer, gRNA dose or titer, CRISPR-Cas complex dose or titer, CRISPR effector protein size, CRISPR effector expression level, gRNA expression level, CRISPR-Cas complex expression level, CRISPR effector spatiotemporal expression, gRNA spatiotemporal expression, CRISPR-Cas complex spatiotemporal expression.

[0693] It will be understood that the parameters or variables to be optimized as well as the nature of optimization may depend on the (therapeutic) target, the functionality of the composition and/or system, the system mode of delivery, and/or the delivery vehicle or expression system.

[0694] In an aspect, the disclosure relates to a method as described herein, comprising optimization of gRNA specificity at the population level. Preferably, said optimization of gRNA specificity comprises minimizing gRNA target site sequence variation across a population and/or minimizing gRNA off-target incidence across a population.

[0695] In some embodiments, optimization can result in selection of a CRISPR-Cas effector that is naturally occurring or is modified. In some embodiments, optimization can result in selection of a CRISPR-Cas effector that has nuclease, nickase, deaminase, transposase, and/or has one or more effector functionalities deactivated or eliminated. In some embodiments, optimizing a PAM specificity can include selecting a CRISPR-Cas effector with a modified PAM specificity. In some embodiments, optimizing can include selecting a

CRISPR-Cas effector having a minimal size. In certain embodiments, optimizing effector protein stability comprises selecting an effector protein having a short half-life while maintaining sufficient activity, such as by selecting an appropriate CRISPR effector orthologue having a specific half-life or stability. In certain embodiments, optimizing immunogenicity or toxicity comprises minimizing effector protein immunogenicity or toxicity by protein modifications. In certain embodiments, optimizing functional specific comprises selecting a protein effector with reduced tolerance of mismatches and/or bulges between the guide RNA and one or more target loci.

[0696] In certain embodiments, optimizing efficacy comprises optimizing overall efficiency, epigenetic tolerance, or both. In certain embodiments, maximizing overall efficiency comprises selecting an effector protein with uniform enzyme activity across target loci with varying chromatin complexity, selecting an effector protein with enzyme activity limited to areas of open chromatin accessibility. In certain embodiments, chromatin accessibility is measured using one or more of ATAC-seq, or a DNA-proximity ligation assay. In certain embodiments, optimizing epigenetic tolerance comprises optimizing methylation tolerance, epigenetic mark competition, or both. In certain embodiments, optimizing methylation tolerance comprises selecting an effector protein that modify methylated DNA. In certain embodiments, optimizing epigenetic tolerance comprises selecting an effector protein unable to modify silenced regions of a chromosome, selecting an effector protein able to modify silenced regions of a chromosome, or selecting target loci not enriched for epigenetic markers

[0697] In certain embodiments, selecting an optimized guide RNA comprises optimizing gRNA stability, gRNA immunogenicity, or both, or other gRNA associated parameters or variables as described herein elsewhere.

[0698] In certain embodiments, optimizing gRNA stability and/or gRNA immunogenicity comprises RNA modification, or other gRNA associated parameters or variables as described herein elsewhere. In certain embodiments, the modification comprises removing 1-3 nucleotides form the 3' end of a target complementarity region of the gRNA. In certain embodiments, modification comprises an extended gRNA and/or trans RNA/DNA element that create stable structures in the gRNA that compete with gRNA base pairing at a target of off-target loci, or extended complimentary nucleotides between the gRNA and target sequence, or both.

[0699] In certain embodiments, the mode of delivery comprises delivering gRNA and/or CRISPR effector protein, delivering gRNA and/or CRISPR effector mRNA, or delivery gRNA

and/or CRISPR effector as a DNA based expression system. In certain embodiments, the mode of delivery further comprises selecting a delivery vehicle and/or expression systems from the group consisting of liposomes, lipid particles, nanoparticles, biolistics, or viral-based expression/delivery systems. In certain embodiments, expression is spatiotemporal expression is optimized by choice of conditional and/or inducible expression systems, including controllable CRISPR effector activity optionally a destabilized CRISPR effector and/or a split CRISPR effector, and/or cell- or tissue-specific expression system.

[0700] The methods as described herein may further involve selection of the mode of delivery. In certain embodiments, gRNA (and tracr, if and where needed, optionally provided as a sgRNA) and/or CRISPR effector protein are or are to be delivered. In certain embodiments, gRNA (and tracr, if and where needed, optionally provided as a sgRNA) and/or CRISPR effector mRNA are or are to be delivered. In certain embodiments, gRNA (and tracr, if and where needed, optionally provided as a sgRNA), CRISPR effector, and/or transposase provided in a DNA-based expression system are or are to be delivered. In certain embodiments, delivery of the individual system components comprises a combination of the above modes of delivery. In certain embodiments, delivery comprises delivering gRNA, CRISPR effector protein, and/or transposase, delivering gRNA and/or CRISPR effector mRNA, or delivering gRNA and/or CRISPR effector and/or transposase as a DNA based expression system.

[0701] The methods as described herein may further involve selection of the composition, system delivery vehicle and/or expression system. Delivery vehicles and expression systems are described herein elsewhere. By means of example, delivery vehicles of nucleic acids and/or proteins include nanoparticles, liposomes, etc. Delivery vehicles for DNA, such as DNA-based expression systems include, for instance, biolistics, viral based vector systems (e.g. adenoviral, AAV, lentiviral), etc. The skilled person will understand that selection of the mode of delivery, as well as delivery vehicle or expression system, may depend on for instance the cell or tissues to be targeted. In certain embodiments, the delivery vehicle and/or expression system for delivering the compositions, systems, or components thereof comprises liposomes, lipid particles, nanoparticles, biolistics, or viral-based expression/delivery systems.

Considerations for Therapeutic Applications

[0702] A consideration in genome editing therapy is the choice of sequence-specific nuclease, such as a variant of a Cas nuclease. Each nuclease variant may possess its own unique set of strengths and weaknesses, many of which must be balanced in the context of treatment to maximize therapeutic benefit. For a specific editing therapy to be efficacious, a sufficiently high level of modification must be achieved in target cell populations to reverse

disease symptoms. This therapeutic modification 'threshold' is determined by the fitness of edited cells following treatment and the amount of gene product necessary to reverse symptoms. With regard to fitness, editing creates three potential outcomes for treated cells relative to their unedited counterparts: increased, neutral, or decreased fitness. In the case of increased fitness, corrected cells may be able and expand relative to their diseased counterparts to mediate therapy. In this case, where edited cells possess a selective advantage, even low numbers of edited cells can be amplified through expansion, providing a therapeutic benefit to the patient. Where the edited cells possess no change in fitness, an increase the therapeutic modification threshold can be warranted. As such, significantly greater levels of editing may be needed to treat diseases, where editing creates a neutral fitness advantage, relative to diseases where editing creates increased fitness for target cells. If editing imposes a fitness disadvantage, as would be the case for restoring function to a tumor suppressor gene in cancer cells, modified cells would be outcompeted by their diseased counterparts, causing the benefit of treatment to be low relative to editing rates. This may be overcome with supplemental therapies to increase the potency and/or fitness of the edited cells relative to the diseased counterparts.

[0703] In addition to cell fitness, the amount of gene product necessary to treat disease can also influence the minimal level of therapeutic genome editing that can treat or prevent a disease or a symptom thereof. In cases where a small change in the gene product levels can result in significant changes in clinical outcome, the minimal level of therapeutic genome editing is less relative to cases where a larger change in the gene product levels are needed to gain a clinically relevant response. In some embodiments, the minimal level of therapeutic genome editing can range from 0.1 to 1 %, 1-5%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, 30-35%, 35-40%, 40-45%. 45-50%, or 50-55%. Thus, where a small change in gene product levels can influence clinical outcomes and diseases where there is a fitness advantage for edited cells, are ideal targets for genome editing therapy, as the therapeutic modification threshold is low enough to permit a high chance of success.

[0704] The activity of NHEJ and HDR DSB repair can vary by cell type and cell state. NHEJ is not highly regulated by the cell cycle and is efficient across cell types, allowing for high levels of gene disruption in accessible target cell populations. In contrast, HDR acts primarily during S/G2 phase, and is therefore restricted to cells that are actively dividing, limiting treatments that require precise genome modifications to mitotic cells [Ciccia, A. & Elledge, S.J. Molecular cell 40, 179-204 (2010); Chapman, J.R., et al. Molecular cell 47, 497-510 (2012)].

[0705] The efficiency of correction via HDR may be controlled by the epigenetic state or sequence of the targeted locus, or the specific repair template configuration (single vs. double stranded, long vs. short homology arms) used [Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004); Beumer, K.J., et al. G3 (2013)]. The relative activity of NHEJ and HDR machineries in target cells may also affect gene correction efficiency, as these pathways may compete to resolve DSBs [Beumer, K.J., et al. Proceedings of the National Academy of Sciences of the United States of America 105, 19821-19826 (2008)]. HDR also imposes a delivery challenge not seen with NHEJ strategies, as it uses the concurrent delivery of nucleases and repair templates. Thus, these differences can be kept in mind when designing, optimizing, and/or selecting therapeutic as described in greater detail elsewhere herein.

[0706] Polynucleotide modification application can include combinations of proteins, small RNA molecules, and/or repair templates, and can make, in some embodiments, delivery of these multiple parts substantially more challenging than, for example, traditional small molecule therapeutics. Two main strategies for delivery of compositions, systems, and components thereof have been developed: *ex vivo* and *in vivo*. In some embodiments of *ex vivo* treatments, diseased cells are removed from a subject, edited and then transplanted back into the patient. In other embodiments, cells from a healthy allogeneic donor are collected, modified using a composition, system or component thereof, to impart various functionalities and/or reduce immunogenicity, and administered to an allogeneic recipient in need of treatment. *Ex vivo* editing has the advantage of allowing the target cell population to be well defined and the specific dosage of therapeutic molecules delivered to cells to be specified. The latter consideration may be particularly important when off-target modifications are a concern, as titrating the amount of nuclease may decrease such mutations (Hsu et al., 2013). Another advantage of *ex vivo* approaches is the typically high editing rates that can be achieved, due to the development of efficient delivery systems for proteins and nucleic acids into cells in culture for research and gene therapy applications.

[0707] *In vivo* polynucleotide modification via compositions, systems, and/or components thereof involves direct delivery of the compositions, systems, and/or components thereof to cell types in their native tissues. *In vivo* polynucleotide modification via compositions, systems, and/or components thereof allows diseases in which the affected cell population is not amenable to *ex vivo* manipulation to be treated. Furthermore, delivering compositions, systems, and/or components thereof to cells in situ allows for the treatment of multiple tissue and cell types.

[0708] In some embodiments, such as those where viral vector systems are used to generate viral particles to deliver the composition, system and/or component thereof to a cell, the total cargo size of the composition, system and/or component thereof should be considered as vector systems can have limits on the size of a polynucleotide that can be expressed therefrom and/or packaged into cargo inside of a viral particle. In some embodiments, the tropism of a vector system, such as a viral vector system, should be considered as it can impact the cell type to which the composition, system or component thereof can be efficiently and/or effectively delivered.

[0709] When delivering a system or component thereof via a viral-based system, it can be important to consider the amount of viral particles that will be needed to achieve a therapeutic effect so as to account for the potential immune response that can be elicited by the viral particles when delivered to a subject or cell(s). When delivering a system or component thereof via a viral based system, it can be important to consider mechanisms of controlling the distribution and/or dosage of the system *in vivo*. Generally, to reduce the potential for off-target effects, it is optimal but not necessarily required, that the amount of the system be as close to the minimum or least effective dose.

[0710] In some embodiments, it can be important to consider the immunogenicity of the system or component thereof. In embodiments, where the immunogenicity of the system or component thereof is of concern, the immunogenicity system or component thereof can be reduced. By way of example only, the immunogenicity of the system or component thereof can be reduced using the approach set out in Tangri et al. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the CRISPR enzyme and/or transposase in the host species (human or other species).

Xenotransplantation

[0711] The present disclosure also contemplates use of the compositions and systems described hereinto provide RNA-guided DNA nucleases adapted to be used to provide modified tissues for transplantation. For example, RNA-guided DNA nucleases may be used to knockout, knockdown or disrupt selected genes in an animal, such as a transgenic pig (such as the human heme oxygenase-1 transgenic pig line), for example by disrupting expression of genes that encode epitopes recognized by the human immune system, i.e. xenoantigen genes. Candidate porcine genes for disruption may for example include α(1,3)-galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase genes (see International Patent Publication WO 2014/066505). In addition, genes encoding endogenous retroviruses may be disrupted, for example the genes encoding all porcine endogenous retroviruses (see

Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), Science 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). In addition, RNA-guided DNA nucleases may be used to target a site for integration of additional genes in xenotransplant donor animals, such as a human CD55 gene to improve protection against hyperacute rejection.

[0712] Embodiments herein also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The present effector protein systems may be harnessed to correct these defects of genomic instability.

[0713] Several further aspects herein relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

[0714] In some embodiments, the systems or complexes can target nucleic acid molecules, can target and cleave or nick or simply sit upon a target DNA molecule (depending if the effector has mutations that render it a nickase or "dead"). Such systems or complexes are amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include but are not limited to genes involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders. Accordingly, target sequences for such systems or complexes can be in candidate disease genes, e.g.:

Table 4– Diseases and Targets					
Disease	**GENE**	**SPACER**	**PAM**	**Mechanism**	**References**
Hypercholesterolemia	HMG-CR	GCCAAATTGGACGACCCTCG (SEQ ID NO: 103)	CGG	Knockout	Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia. (Plosker GL et al. Drugs 1996, 51(3):433-459)
Hypercholesterolemia	SQLE	CGAGGAGACCCCCGTTTCGG (SEQ ID NO: 104)	TGG	Knockout	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964–971, November 2011)
Hyperlipidemia	DGAT1	CCCGCCGCCGCCGTGGCTCG (SEQ ID NO: 105)	AGG	Knockout	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development [2010, 13(4):489-496)
Leukemia	BCR-ABL	TGAGCTCTACGAGATCCACA (SEQ ID NO: 106)	AGG	Knockout	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi). (Fuchs et al. Oncogene 2002, 21(37):5716-5724)

KITS

[0715] In another aspect, the present disclosure provides kit and kit of parts. The terms "kit of parts" and "kit" as used throughout this specification refer to a product containing components necessary for carrying out the specified methods (e.g., methods for detecting, quantifying or isolating immune cells as taught herein), packed so as to allow their transport and storage. Materials suitable for packing the components comprised in a kit include crystal, plastic (e.g., polyethylene, polypropylene, polycarbonate), bottles, flasks, vials, ampules, paper, envelopes, or other types of containers, carriers or supports. Where a kit comprises a plurality of components, at least a subset of the components (e.g., two or more of the plurality of components) or all of the components may be physically separated, e.g., comprised in or on separate containers, carriers or supports. The components comprised in a kit may be sufficient or may not be sufficient for carrying out the specified methods, such that external reagents or substances may not be necessary or may be necessary for performing the methods, respectively. Typically, kits are employed in conjunction with standard laboratory equipment, such as liquid

handling equipment, environment (e.g., temperature) controlling equipment, analytical instruments, etc. In addition to the recited binding agents(s) as taught herein, such as for example, antibodies, hybridization probes, amplification and/or sequencing primers, optionally provided on arrays or microarrays, the present kits may also include some or all of solvents, buffers (such as for example but without limitation histidine-buffers, citrate-buffers, succinate-buffers, acetate-buffers, phosphate-buffers, formate buffers, benzoate buffers, TRIS (Tris(hydroxymethyl)-aminomethan) buffers or maleate buffers, or mixtures thereof), enzymes (such as for example but without limitation thermostable DNA polymerase), detectable labels, detection reagents, and control formulations (positive and/or negative), useful in the specified methods. Typically, the kits may also include instructions for use thereof, such as on a printed insert or on a computer readable medium. The terms may be used interchangeably with the term "article of manufacture", which broadly encompasses any man-made tangible structural product, when used in the present context.

[0716] The present application also provides aspects and embodiments as set forth in the following numbered Statements:

[0717] Statement 1. An engineered or non-naturally occurring composition comprising: a. a site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof; b. a non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof; c. a guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide; and d. a polynucleotide encoding a retrotransposon RNA, wherein the retrotransposon RNA comprises or encodes a donor polynucleotide.

[0718] Statement 2. An engineered or non-naturally occurring composition comprising: a. two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof; b. two non-LTR retrotransposon polypeptides, each connected to or otherwise capable of forming a complex with one of the two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof; c. two guide molecules, each capable of forming a complex with one of the site-specific nuclease polypeptides and directing site-specific binding to a target sequence of a target polynucleotide; and d. a polynucleotide encoding a retrotransposon RNA comprising or encoding a donor polynucleotide.

[0719] Statement 3. The composition of Statement 1 or 2, wherein the retrotransposon RNA capable of forming a complex with the non-LTR retrotransposon polypeptide.

[0720] Statement 4. The composition of Statement 1 or 2, wherein the retrotransposon RNA comprises an binding element capable of binding to the non-LTR retrotransposon polypeptide.

[0721] Statement 5. The composition of Statement 4, wherein the binding element comprises a hairpin structure.

[0722] Statement 6. The composition of any one of the preceding Statements, wherein the donor polynucleotide is for insertion at, or adjacent to, the target sequence.

[0723] Statement 7. The composition of any one of the preceding Statements, wherein the site-specific nuclease is a nickase.

[0724] Statement 8. The composition of any one of the preceding Statements, wherein the site-specific nuclease lacks nuclease activity.

[0725] Statement 9. The composition of any one of the preceding Statements, wherein the non-LTR retrotransposon polypeptide is a dimer, wherein the dimer subunits are connected or form a tandem fusion.

[0726] Statement 10. The composition of any one of the preceding Statements, wherein the non-LTR retrotransposon polypeptide comprises a first retrotransposon polypeptide and a second retrotransposon polypeptide, wherein the second retrotransposon polypeptide comprises nuclease or nickase activity.

[0727] Statement 11. The composition of Statement 10, wherein the site-specific nuclease is connected to the second retrotransposon polypeptide.

[0728] Statement 12. The composition of any of the preceding Statements, wherein the nuclease domain(s), and/or homing domain of the retrotransposon polypeptide is inactivated.

[0729] Statement 13. The composition of any one of the preceding Statements, wherein the non-LTR retrotransposon polypeptide is R2.

[0730] Statement 14. The composition of Statement 13, wherein the R2 is from *Bombyx mori, Clonorchis sinensis*, or *Zonotrichia albicollis*.

[0731] Statement 15. The composition of any one of Statements 1 to 12, wherein the non-LTR retrotransposon polypeptide is L1.

[0732] Statement 16. The composition of any one of proceeding Statements, wherein the site-specific nuclease polypeptide comprises a nuclear localization signal sequence.

[0733] Statement 17. The composition of any one of proceeding Statements, wherein the non-LTR retrotransposon polypeptide comprises a nuclear localization signal.

[0734] Statement 18. The composition of any one of proceeding Statements, wherein the polynucleotide encoding a retrotransposon RNA comprises a poly-A tail.

[0735] Statement 19. An engineered composition for non-native, targeted transposition of donor sequence into targeted nucleic acids, comprising: a. a fusion protein comprising a site-specific nuclease fused to the N-terminus of a non-LTR retrotransposon polypeptide, or a polynucleotide comprising a coding sequence thereof; and b. a donor construct comprising a donor polynucleotide sequence located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide.

[0736] Statement 20. The composition of Statement 19, wherein the donor polynucleotide further comprises a polymerase processing element to facilitate 3' end processing of the donor polynucleotide sequence.

[0737] Statement 21. The composition of any one of Statements 19 to 20, wherein the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both.

[0738] Statement 22. The composition of Statement 21, wherein the homology region is between 8 and 25 base pairs.

[0739] Statement 23. The composition of Statement 21 or 22, wherein the homology region is on the 3' end of the donor polynucleotide only.

[0740] Statement 24. The composition of any one of Statements 19 to 23, wherein the donor polynucleotide sequence is between 5bp and 50 kb in length.

[0741] Statement 25. The composition of any one of Statements 19 to 24, wherein non-LTR retrotransposon polypeptide is a wild-type non-LTR retrotransposon polypeptide.

[0742] Statement 26. The composition of any one of Statements 19 to 24, wherein the non-LTR retrotransposon polypeptide comprises one or more modification or truncations.

[0743] Statement 27. The composition of Statement 26, wherein the one or more modifications or one or more truncations are in an endonuclease domain or reverse transcriptase domain.

[0744]

[0745] Statement 28. The composition of Statement 26, wherein the one or more modifications or truncations are truncations are in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain.

[0746] Statement 29. The composition of any one of Statements 19 to 28, wherein the fusion protein comprises a nuclear localization signal.

[0747] Statement 30. The composition of any one of the preceding Statements, wherein the site-specific nuclease polypeptide is a Cas polypeptide.

[0748] Statement 31. The composition of Statement 30, further comprising a guide molecule capable of forming a CRISPR-Cas complex with the Cas polypeptide and directing site-specific binding to a target sequence of a target polynucleotide.

[0749] Statement 32. The composition of Statement 30, wherein the guide directs the fusion protein to a target sequence 5' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

[0750] Statement 33. The composition of Statement 30, wherein the guide directs the fusion protein to a target sequence 3' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

[0751] Statement 34. The composition of any one of Statements 30-32, wherein the Cas polypeptide is a Class 2, Type II Cas or a Type V Cas.

[0752] Statement 35. The composition of Statement 34, wherein the Cas polypeptide is a Class 2, Type II Cas.

[0753] Statement 36. The composition of Statement 35, wherein the Type II Cas is a Cas9.

[0754] Statement 37. The composition of Statement 36, wherein the Cas9 has an HNH domain that is inactivated.

[0755] Statement 38. The composition of Statement 36, wherein the Cas9 is Cas9D10A or Cas9H840A.

[0756] Statement 39. The composition of Statement 34, wherein the Cas polypeptide is a Class 2, Type V Cas.

[0757] Statement 40. The composition of Statement 39, wherein the Type V Cas is Cas12a or Cas12b.

[0758] Statement 41. The composition of any one of Statements 1 to 29, where the site specific nuclease polypeptide is a IscB or a TnpB.

[0759] Statement 42. The composition of any one of the preceding Statements, wherein the polynucleotide encoding a retrotransposon RNA comprises a pol2 promoter, a pol3 promoter, or a T7 promoter.

[0760] Statement 43. The composition of any one or the preceding Statements, wherein a 3' end of the retrotransposon RNA is complementary to the target sequence, specifically to a portion of a nicked target sequence.

[0761] Statement 44. The composition of any one of the preceding Statements, further comprising an RNaseH.

[0762] Statement 45. The composition any one of the preceding Statements, wherein the two site-specific nuclease polypeptides bind to two target sites on the target polynucleotide, and the donor polynucleotide is inserted to a position between the two target sites.

[0763] Statement 46. The composition of any one of the preceding Statements, wherein the retrotransposon RNA comprises a region capable of hybridizing with an overhang of the target polynucleotide.

[0764] Statement 47. The composition of any one of the proceeding Statements, wherein the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA.

[0765] Statement 48. The composition of any one of the proceeding Statements, wherein the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA.

[0766] Statement 49. The composition of Statement 48, wherein the mRNA comprises a poly-A tail.

[0767] Statement 50. The composition of any one of the proceeding Statements, wherein the polynucleotide comprising the coding sequence of non-LTR retrotransposon polypeptide is an mRNA.

[0768] Statement 51. The composition of Statement 50, wherein the mRNA comprises a poly-A tail.

[0769] Statement 52. The composition of any one of the proceeding Statements, wherein the donor polynucleotide comprises a homology sequence of the target sequence.

[0770] Statement 53. The composition of Statement 52, wherein the homology sequence is of a region on a strand of the target sequence that contains a PAM of the site-specific nuclease polypeptide.

[0771] Statement 54. The composition of Statement 53, wherein the region comprises the PAM sequence.

[0772] Statement 55. The composition of Statement 53 or 54, wherein the region is at 3' side of a cleavage site of the site-specific nuclease polypeptide.

[0773] Statement 56. The composition of any one of Statements 52-55, wherein the homology sequence comprises from 1 to 30, from 4 to 10, or from 10 to 25 nucleotides in length.

[0774] Statement 57. The composition of Statement 52, wherein the homology sequence is of a region on a strand that binds to the guide.

[0775] Statement 58. The composition of Statement 57, wherein the region comprises at least a portion of a guide-binding sequence.

[0776] Statement 59. The composition of Statement 57 or 58, where the region comprises a sequence at 3' side of the guide-binding sequence.

[0777] Statement 60. The composition of Statement 59, wherein the guide molecule forms a RNA-DNA duplex with the target sequence, and the region comprises a sequence of 5 to 15 nucleotides from 3' of the RNA-DNA duplex.

[0778] Statement 61. The composition of Statement 60, wherein the region comprises a sequence of 10 nucleotides from 3' side of the RNA-DNA duplex.

[0779] Statement 62. The composition of any one of the proceeding Statements, wherein the donor polynucleotide is an RNA comprising a poly-A tail.

[0780] Statement 63. A vector system comprising one or more vectors, the one or more vectors comprising one or more polynucleotides encoding the polypeptides and/or polynucleotides of Statements 1 to 62, or a combination thereof.

[0781] Statement 64. The system according to Statement 63, wherein the one or more polynucleotides comprise one or more regulatory elements operably configures to express the polypeptide(s) and/or the nucleic acid component(s), optionally wherein the one or more regulatory elements comprise inducible promoters.

[0782] Statement 65. The system of Statement 63 or 64, wherein the polynucleotide molecule encoding the Cas polypeptide is codon optimized for expression in a eukaryotic cell.

[0783] Statement 66. A cell or progeny thereof transiently or non-transiently transfected with the vector system of any one of Statements 63 to 65.

[0784] Statement 67. An organism comprising the cell of Statement 66.

[0785] Statement 68. A method of inserting a donor polynucleotide sequence into a target polynucleotide comprising: introducing the engineered or non-naturally occurring composition of any one of Statements 1 to 62 to a cell or population of cells, wherein the complex of the site-specific nuclease polypeptide and the guide directs the non-LTR retrotransposon polypeptide to the target sequence, and wherein the non-LTR retrotransposon polypeptide inserts the donor polynucleotide encoded by the retrotransposon RNA at or adjacent to the target sequence.

[0786] Statement 69. The method of Statement 68, wherein the donor polynucleotide: a. introduces one or more mutations to the target polynucleotide, b. inserts a functional gene or gene fragment at the target polynucleotide, c. corrects or introduces a premature stop codon in

the target polynucleotide, d. disrupts or restores a splice cite in the target polynucleotide, e. causes a shift in the open reading frame of the target polynucleotide, or f. a combination thereof.

[0787] Statement 70. The method of Statement 69, wherein the one or more mutations include substitutions, deletions, and insertions.

[0788] Statement 71. The method of Statement 68, wherein the donor polynucleotide is between 100 bases and 30 kb in length.

[0789] Statement 72. The method of any one of Statements 68 to 71, wherein the polypeptide and/or nucleic acid components are provided via one or more polynucleotide encoding the polypeptides and/or nucleic acid component(s), and wherein the one or more polynucleotides are operably configured to express the polypeptides and/or nucleic acid component(s).

[0790] Statement 73. The method of any one of Statements 68 to 72, wherein the composition is delivered via liposomes, nanoparticles, exosomes, microvesicles, microinjection, a gene-gun, or one or more viral vectors.

[0791] Statement 74. The method of any one of Statements 68 to 72, wherein the donor polynucleotide is inserted to a region on the target sequence that is 3' of a PAM-containing strand.

[0792] Statement 75. The method of any one of Statements 68 to 72, wherein the donor polynucleotide is inserted to a region on the target sequence that is 3' of a sequence complementary to the guide molecule.

EXAMPLES

Example 1 –

[0793] FIG. 33 shows detection of insertion products by amplifying junction between 3' UTR of donor and target site. DNA transfections (in HEK 293 cells) showed beginning of target-primed reverse transcription (TPRT), but no defined insertion products Expected defined band so not getting insertion but a variety of insertion products.

Example 2 -

[0794] Transfection of mRNA with coding sequences of R2 and mRNA donors in HEK 293 cells. The mRNA constructs are shown in FIG. 34. Insertion frequency of donors constructs with (4, 10, 25, 50, 75, and 100 bp homology sequences) with or without poly-A tails were tested (FIG. 35).

[0795] mRNA constructs designed for inserting donor polynucleotides to the 3' side of the PAM-containing strand or 3' side of the guide-binding sequence were tested. The constructs and their homology positions are shown in FIG. 36.

[0796] 3' insertion was tested by PCR. Results and primer locations are shown in FIG. 37. Sequencing was performed to confirm the insertions (FIGs. 38A-38F). 5' insertion was also tested by PCR. Results and primer locations are shown in FIG. 39. Sequencing was performed to confirm the insertions (FIGs. 40A-40B). Guides used for targets 6 and 7 are below (PAMs are in bold):

[0797] Target 6: TCAGTCCAGCCCCTTCAGTC**TGG** (SEQ ID NO: 107)

[0798] Target 7: ACACAACAAGGCAGTGACAG**TGG** (SEQ ID NO: 108)

[0799] FIG. 41 shows that R2 orthologs (as indicated on the x-axis) was transfected into HEK293FT cells with or without supplementation of associated human codon optimized ORFs. Insertions were detected at the human 28S rRNA using ddPCR. R2NS-1_CSi is R2 from *Clonorchis sinensis*. R2-1_ZA is R2 from *Zonotrichia albicollis*. The sequences of the orthologs and human codon optimized versions are in Table 1.

* * *

[0800] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. 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 come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is

1. An engineered or non-naturally occurring composition comprising:

a. a site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;

b. a non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;

c. a guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide; and

d. a polynucleotide encoding a retrotransposon RNA, wherein the retrotransposon RNA comprises or encodes a donor polynucleotide.

2. An engineered or non-naturally occurring composition comprising:

a. two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof;

b. two non-LTR retrotransposon polypeptides, each connected to or otherwise capable of forming a complex with one of the two site-specific nuclease polypeptides, or one or more polynucleotides comprising coding sequences thereof;

c. two guide molecules, each capable of forming a complex with one of the site-specific nuclease polypeptides and directing site-specific binding to a target sequence of a target polynucleotide; and

d. a polynucleotide encoding a retrotransposon RNA comprising or encoding a donor polynucleotide.

3. The composition of claim 1, wherein the retrotransposon RNA capable of forming a complex with the non-LTR retrotransposon polypeptide.

4. The composition of claim 1, wherein the retrotransposon RNA comprises an binding element capable of binding to the non-LTR retrotransposon polypeptide.

5. The composition of claim 4, wherein the binding element comprises a hairpin structure.

6. The composition of claim 1, wherein the donor polynucleotide is for insertion at, or adjacent to, the target sequence.

7. The composition of claim 1, wherein the site-specific nuclease is a nickase.

8. The composition of claim 1, wherein the site-specific nuclease lacks nuclease activity.

9. The composition of claim 1, wherein the non-LTR retrotransposon polypeptide is a dimer, wherein the dimer subunits are connected or form a tandem fusion.

10. The composition of claim 1, wherein the non-LTR retrotransposon polypeptide comprises a first retrotransposon polypeptide and a second retrotransposon polypeptide, wherein the second retrotransposon polypeptide comprises nuclease or nickase activity.

11. The composition of claim 10, wherein the site-specific nuclease is connected to the second retrotransposon polypeptide.

12. The composition of claim 1, wherein the nuclease domain(s), and/or homing domain of the retrotransposon polypeptide is inactivated.

13. The composition of claim 1, wherein the non-LTR retrotransposon polypeptide is R2.

14. The composition of claim 13, wherein the R2 is from *Bombyx mori*, *Clonorchis sinensis*, or *Zonotrichia albicollis*.

15. The composition of claim 1, wherein the non-LTR retrotransposon polypeptide is L1.

16. The composition of claim 1, wherein the site-specific nuclease polypeptide comprises a nuclear localization signal sequence.

17. The composition of claim 1, wherein the non-LTR retrotransposon polypeptide comprises a nuclear localization signal.

18. The composition of claim 1, wherein the polynucleotide encoding a retrotransposon RNA comprises a poly-A tail.

19. An engineered composition for non-native, targeted transposition of donor sequence into targeted nucleic acids, comprising:

a. a fusion protein comprising a site-specific nuclease fused to the N-terminus of a non-LTR retrotransposon polypeptide, or a polynucleotide comprising a coding sequence thereof; and

b. a donor construct comprising a donor polynucleotide sequence located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide.

20. The composition of claim 19, wherein the donor polynucleotide further comprises a polymerase processing element to facilitate 3' end processing of the donor polynucleotide sequence.

21. The composition of claim 19, wherein the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both.

22. The composition of claim 21, wherein the homology region is between 8 and 25 base pairs.

23. The composition of claim 21, wherein the homology region is on the 3' end of the donor polynucleotide only.

24. The composition of claim 19, wherein the donor polynucleotide sequence is between 5bp and 50 kb in length.

25. The composition of claim 19, wherein non-LTR retrotransposon polypeptide is a wild-type non-LTR retrotransposon polypeptide.

26. The composition of claim 19, wherein the non-LTR retrotransposon polypeptide comprises one or more modification or truncations.

27. The composition of claim 26, wherein the one or more modifications or one or more truncations are in an endonuclease domain or reverse transcriptase domain.

28. The composition of claim 26, wherein the one or more modifications or truncations are truncations are in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain.

29. The composition of claim 19, wherein the fusion protein comprises a nuclear localization signal.

30. The composition of claim 19, wherein the site-specific nuclease polypeptide is a Cas polypeptide.

31. The composition of claim 30, further comprising a guide molecule capable of forming a CRISPR-Cas complex with the Cas polypeptide and directing site-specific binding to a target sequence of a target polynucleotide.

32. The composition of claim 30, wherein the guide directs the fusion protein to a target sequence 5' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

33. The composition of claim 30, wherein the guide directs the fusion protein to a target sequence 3' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

34. The composition of claim 30, wherein the Cas polypeptide is a Class 2, Type II Cas or a Type V Cas.

35. The composition of claim 34, wherein the Cas polypeptide is a Class 2, Type II Cas.

36. The composition of claim 35, wherein the Type II Cas is a Cas9.

37. The composition of claim 36, wherein the Cas9 has an HNH domain that is inactivated.

38. The composition of claim 36, wherein the Cas9 is $Cas9^{D10A}$ or $Cas9^{H840A}$.

39. The composition of claim 34, wherein the Cas polypeptide is a Class 2, Type V Cas.

40. The composition of claim 39, wherein the Type V Cas is Cas12a or Cas12b.

41. The composition of claim 1, where the site specific nuclease polypeptide is a IscB or a TnpB.

42. The composition of claim 1, wherein the polynucleotide encoding a retrotransposon RNA comprises a pol2 promoter, a pol3 promoter, or a T7 promoter.

43. The composition of claim 1, wherein a 3' end of the retrotransposon RNA is complementary to the target sequence, specifically to a portion of a nicked target sequence.

44. The composition of claim 1, further comprising an RNaseH.

45. The composition claim 1, wherein the two site-specific nuclease polypeptides bind to two target sites on the target polynucleotide, and the donor polynucleotide is inserted to a position between the two target sites.

46. The composition of claim 1, wherein the retrotransposon RNA comprises a region capable of hybridizing with an overhang of the target polynucleotide.

47. The composition of claim 1, wherein the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA.

48. The composition of claim 1, wherein the polynucleotide comprising the coding sequence of the site-specific nuclease polypeptide is an mRNA.

49. The composition of claim 48, wherein the mRNA comprises a poly-A tail.

50. The composition of claim 1, wherein the polynucleotide comprising the coding sequence of non-LTR retrotransposon polypeptide is an mRNA.

51. The composition of claim 50, wherein the mRNA comprises a poly-A tail.

52. The composition of claim 1, wherein the donor polynucleotide comprises a homology sequence of the target sequence.

53. The composition of claim 52, wherein the homology sequence is of a region on a strand of the target sequence that contains a PAM of the site-specific nuclease polypeptide.

54. The composition of claim 53, wherein the region comprises the PAM sequence.

55. The composition of claim 53, wherein the region is at 3' side of a cleavage site of the site-specific nuclease polypeptide.

56. The composition of claim 52, wherein the homology sequence comprises from 1 to 30, from 4 to 10, or from 10 to 25 nucleotides in length.

57. The composition of claim 52, wherein the homology sequence is of a region on a strand that binds to the guide.

58. The composition of claim 57, wherein the region comprises at least a portion of a guide-binding sequence.

59. The composition of claim 57, where the region comprises a sequence at 3' side of the guide-binding sequence.

60. The composition of claim 59, wherein the guide molecule forms a RNA-DNA duplex with the target sequence, and the region comprises a sequence of 5 to 15 nucleotides from 3' of the RNA-DNA duplex.

61. The composition of claim 60, wherein the region comprises a sequence of 10 nucleotides from 3' side of the RNA-DNA duplex.

62. The composition of claim 1, wherein the donor polynucleotide is an RNA comprising a poly-A tail.

63. A vector system comprising one or more vectors, the one or more vectors comprising one or more polynucleotides encoding the polypeptides and/or polynucleotides of claim 1, or a combination thereof.

64. The system according to claim 63, wherein the one or more polynucleotides comprise one or more regulatory elements operably configures to express the polypeptide(s) and/or the nucleic acid component(s), optionally wherein the one or more regulatory elements comprise inducible promoters.

65. The system of claim 63, wherein the polynucleotide molecule encoding the Cas polypeptide is codon optimized for expression in a eukaryotic cell.

66. A cell or progeny thereof transiently or non-transiently transfected with the vector system of claim 63.

67. An organism comprising the cell of claim 66.

68. A method of inserting a donor polynucleotide sequence into a target polynucleotide comprising: introducing the engineered or non-naturally occurring composition of claim 1 to a cell or population of cells, wherein the complex of the site-specific nuclease polypeptide and the guide directs the non-LTR retrotransposon polypeptide to the target sequence, and wherein the non-LTR retrotransposon polypeptide inserts the donor polynucleotide encoded by the retrotransposon RNA at or adjacent to the target sequence.

69. The method of claim 68, wherein the donor polynucleotide:

a. introduces one or more mutations to the target polynucleotide,

b. inserts a functional gene or gene fragment at the target polynucleotide,

c. corrects or introduces a premature stop codon in the target polynucleotide,

d. disrupts or restores a splice cite in the target polynucleotide,

e. causes a shift in the open reading frame of the target polynucleotide, or

f. a combination thereof.

70. The method of claim 69, wherein the one or more mutations include substitutions, deletions, and insertions.

71. The method of claim 68, wherein the donor polynucleotide is between 100 bases and 30 kb in length.

72. The method of claim 68, wherein the polypeptide and/or nucleic acid components are provided via one or more polynucleotide encoding the polypeptides and/or nucleic acid component(s), and wherein the one or more polynucleotides are operably configured to express the polypeptides and/or nucleic acid component(s).

73. The method of claim 68, wherein the composition is delivered via liposomes, nanoparticles, exosomes, microvesicles, microinjection, a gene-gun, or one or more viral vectors.

74. The method of claim 68, wherein the donor polynucleotide is inserted to a region on the target sequence that is 3' of a PAM-containing strand.

75. The method of claim 68, wherein the donor polynucleotide is inserted to a region on the target sequence that is 3' of a sequence complementary to the guide molecule.

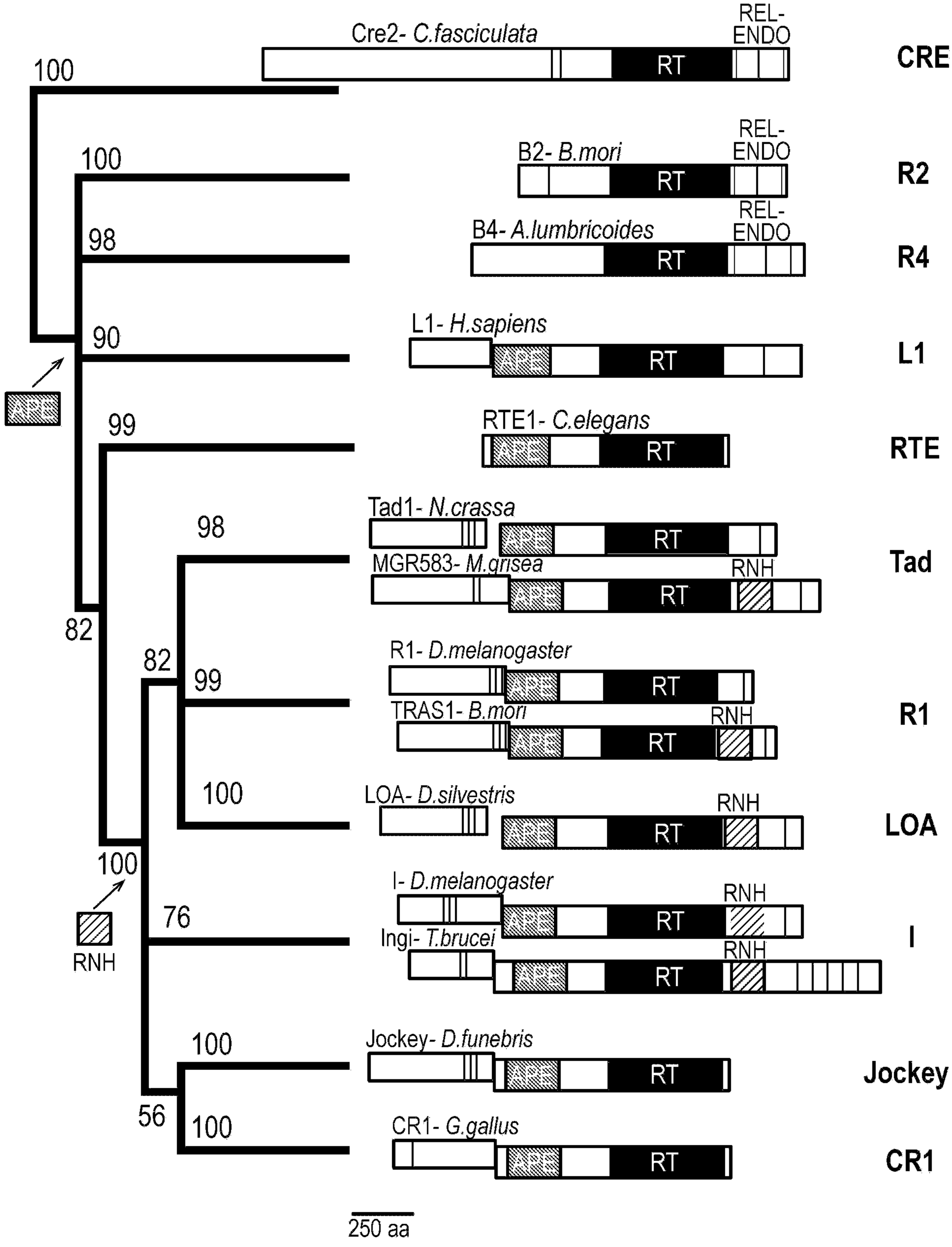

clade
Cre2- C.fasciculata
REL-ENDO
RT
CRE
100
B2- B.mori
R2
100
B4- A.lumbricoides
R4
98
L1- H.sapiens
L1
90
APE
RTE1- C.elegans
RTE
99
Tad1- N.crassa
APE
98
MGR583- M.grisea
RNH
Tad
82
R1- D.melanogaster
82
99
TRAS1- B.mori
R1
LOA- D.silvestris
100
LOA
100
I- D.melanogaster
76
Ingi- T.brucei
I
RNH
Jockey- D.funebris
100
Jockey
56
CR1- G.gallus
100
CR1
250 aa

FIG. 1

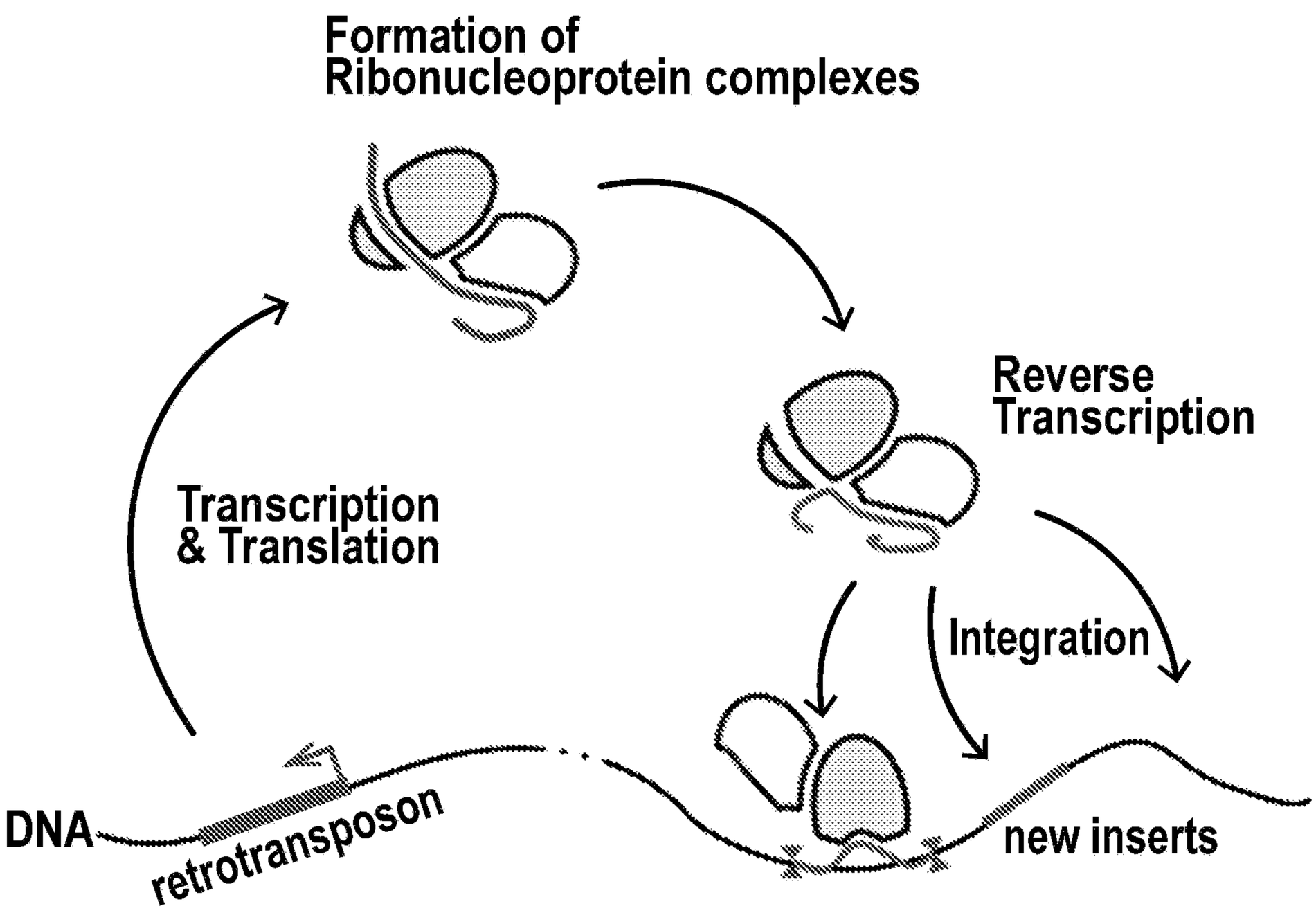

Formation of
Ribonucleoprotein complexes
Reverse
Transcription
Integration
Transcription
& Translation
DNA
retrotransposon
new inserts

FIG. 2

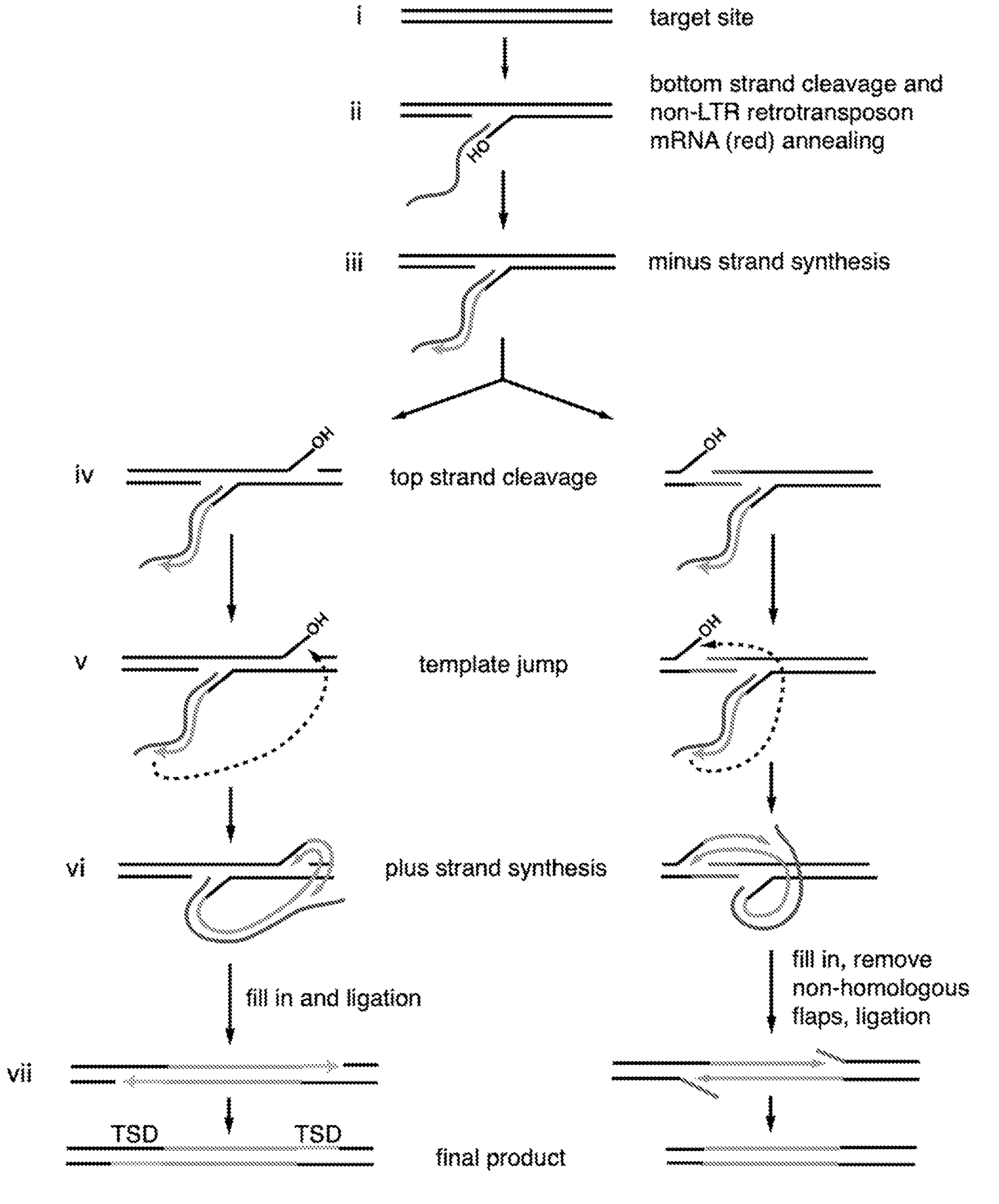

i
target site
bottom strand cleavage and
non-LTR retrotransposon
mRNA (red) annealing
ii
HO
iii
minus strand synthesis
iv
OH
top strand cleavage
OH
v
OH
template jump
OH
vi
plus strand synthesis
fill in and ligation
fill in, remove
non-homologous
flaps, ligation
vii
TSD
TSD
final product

FIG. 3

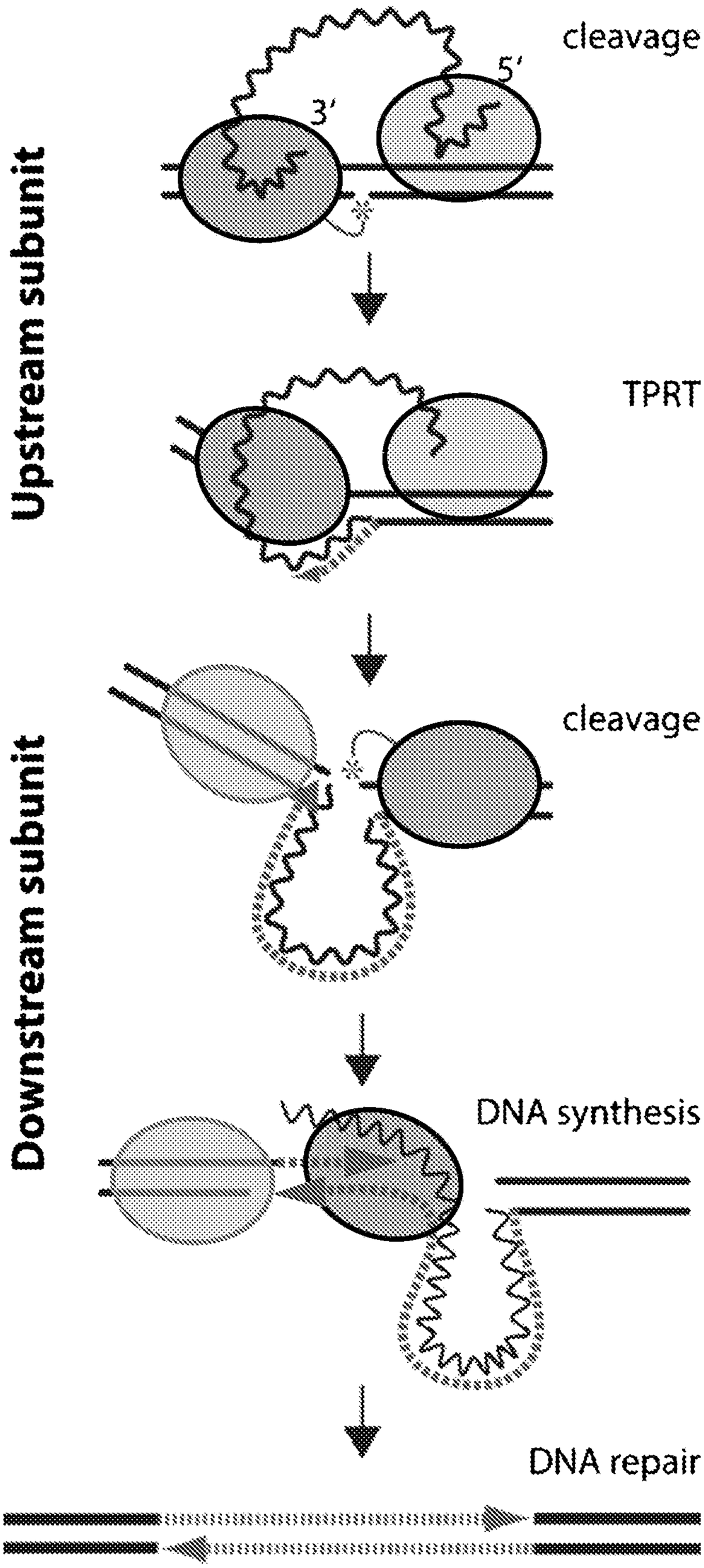

Upstream subunit
cleavage
3'
5'
TPRT
Downstream subunit
cleavage
DNA synthesis
DNA repair

FIG. 4

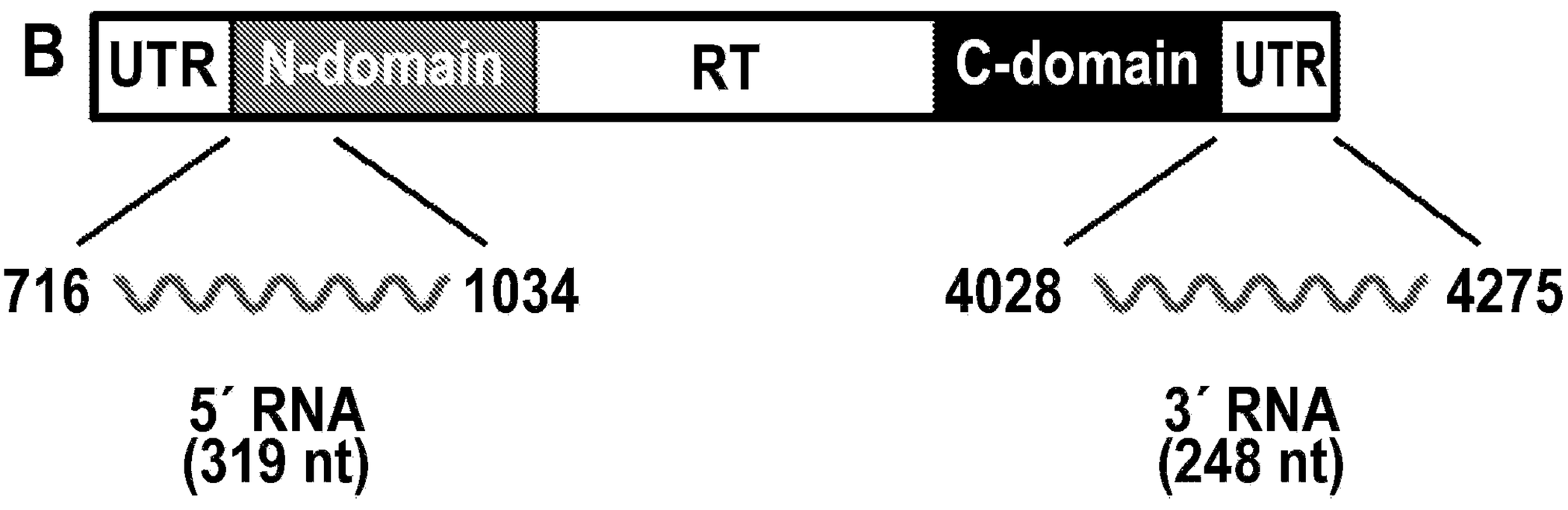

B
UTR
N-domain
RT
C-domain
UTR
716
1034
4028
4275
5´ RNA
(319 nt)
3´ RNA
(248 nt)

FIG. 5

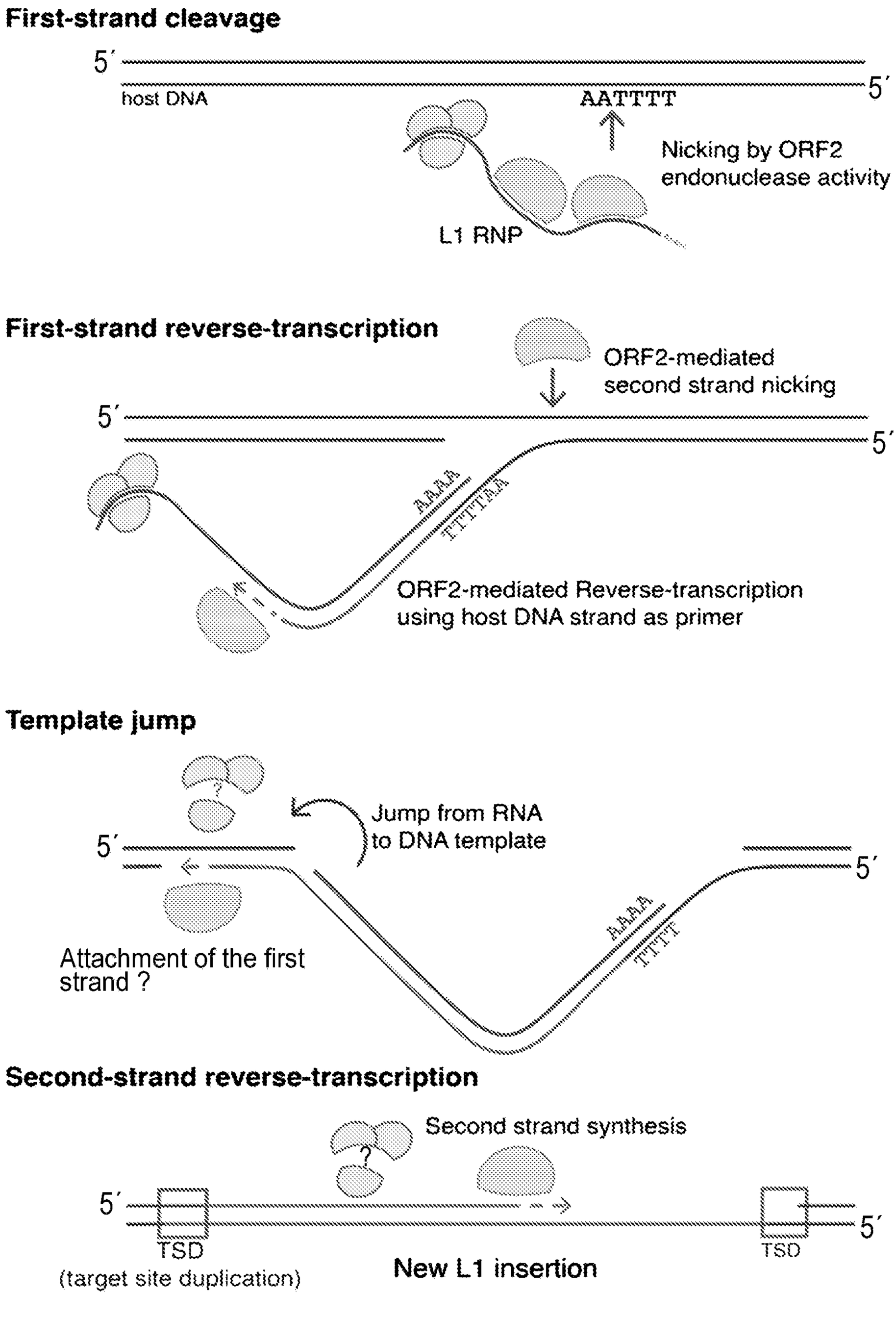

First-strand cleavage
5′
host DNA
AATTTT
5′
Nicking by ORF2
endonuclease activity
L1 RNP
First-strand reverse-transcription
ORF2-mediated
second strand nicking
5′
5′
AAAA
TTTTTTA
ORF2-mediated Reverse-transcription
using host DNA strand as primer
Template jump
Jump from RNA
to DNA template
5′
5′
AAAA
TTTT
Attachment of the first
strand ?
Second-strand reverse-transcription
Second strand synthesis
?
5′
5′
TSD
TSD
(target site duplication)
New L1 insertion

FIG. 6

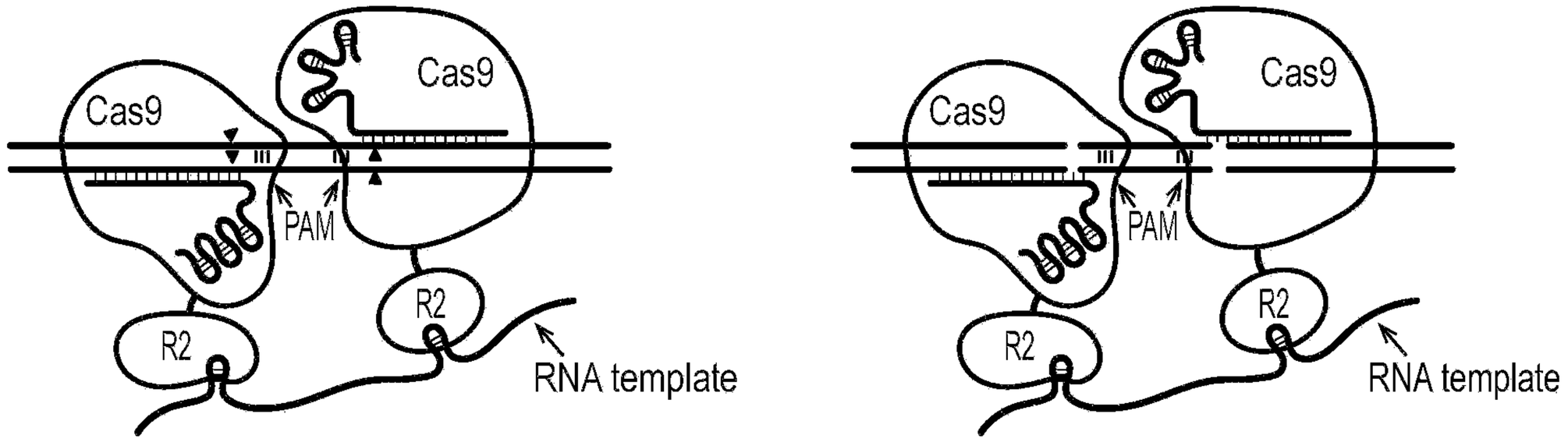

Cas9
Cas9
PAM
R2
R2
RNA template
Cas9
Cas9
PAM
R2
R2
RNA template
Cas9
Cas9
PAM
R2
R2
RNA template
Cas9
Cas9
PAM
R2
R2

FIG. 7

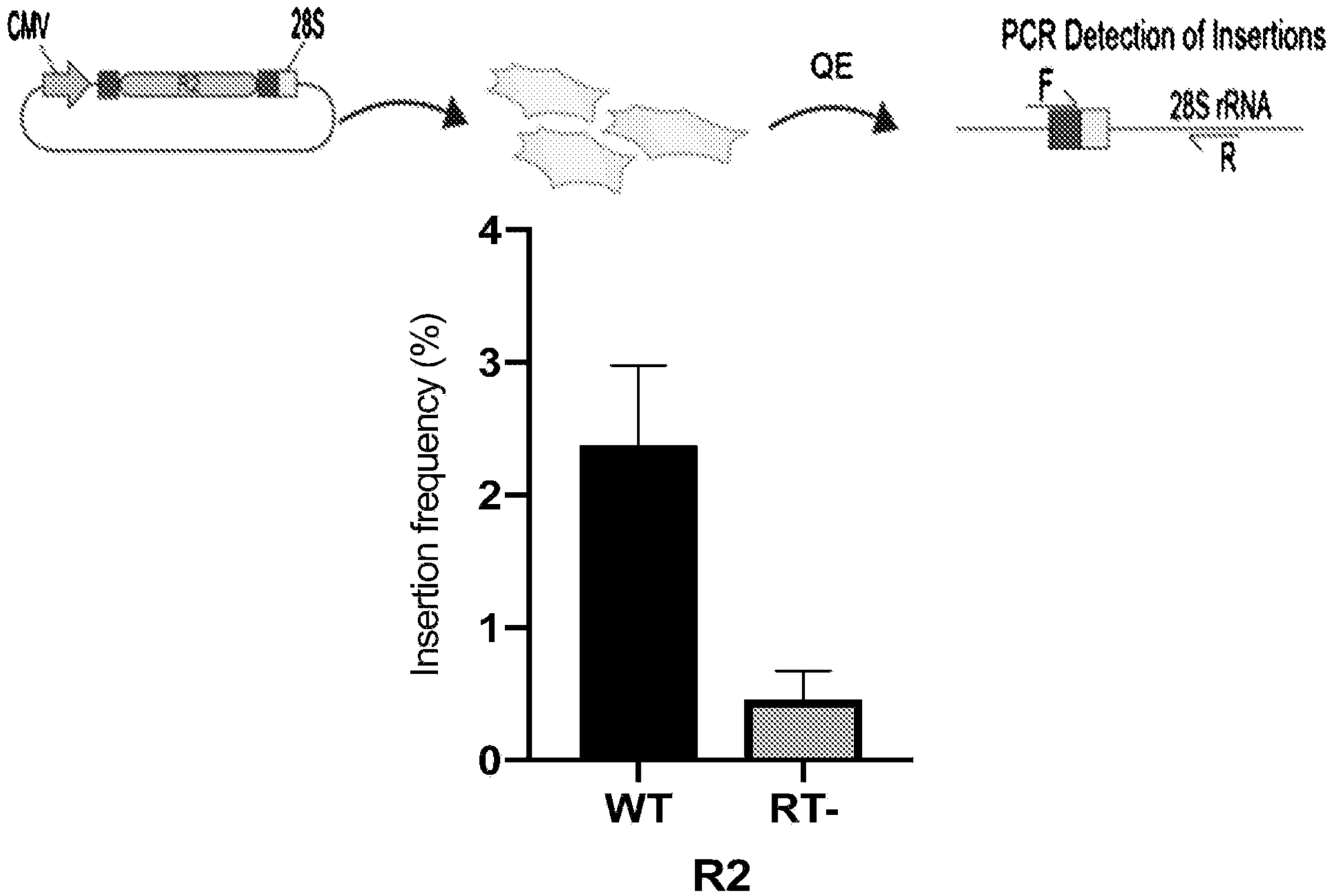

CMV
28S
QE
PCR Detection of Insertions
F
28S rRNA
R
Insertion frequency (%)
4
3
2
1
0
WT
RT-
R2

FIG. 8

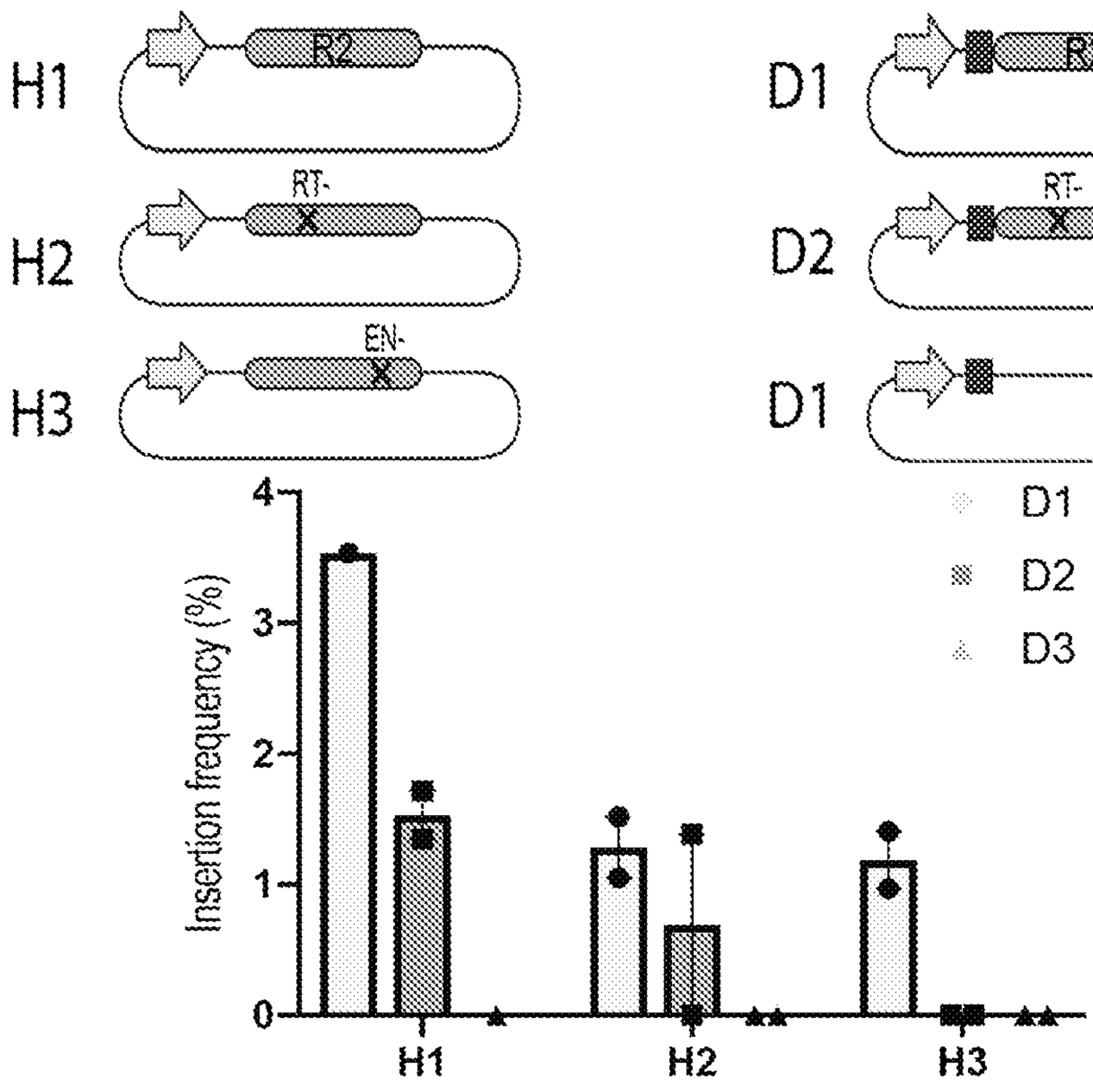

H1
R2
RT-
H2
EN-
H3
D1
R2
RT-
D2
D1
D1
D2
D3
Insertion frequency (%)
4
3
2
1
0
H1
H2
H3

FIG. 9

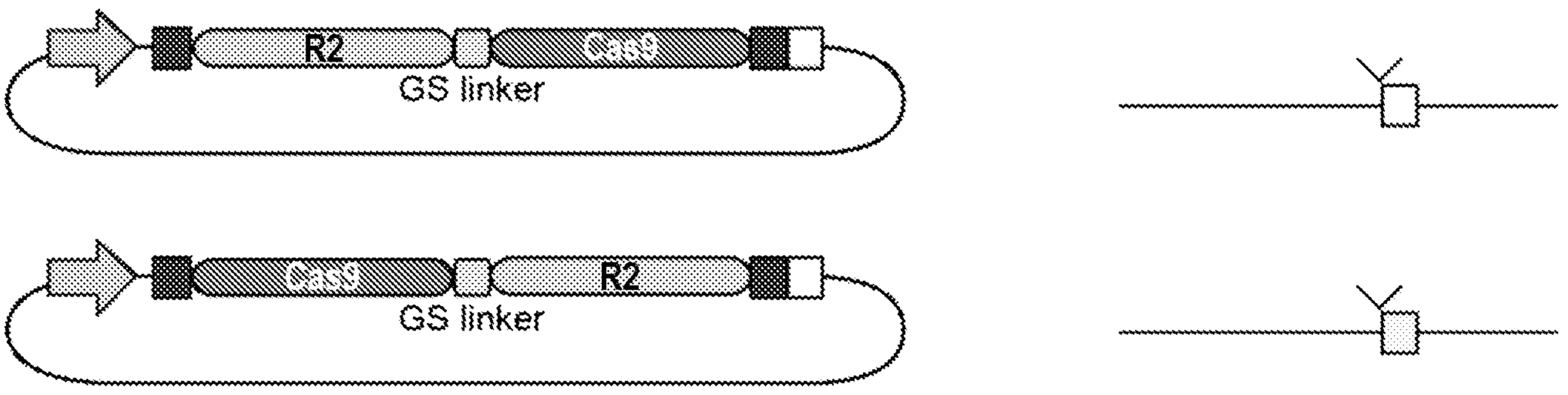


FIG. 10

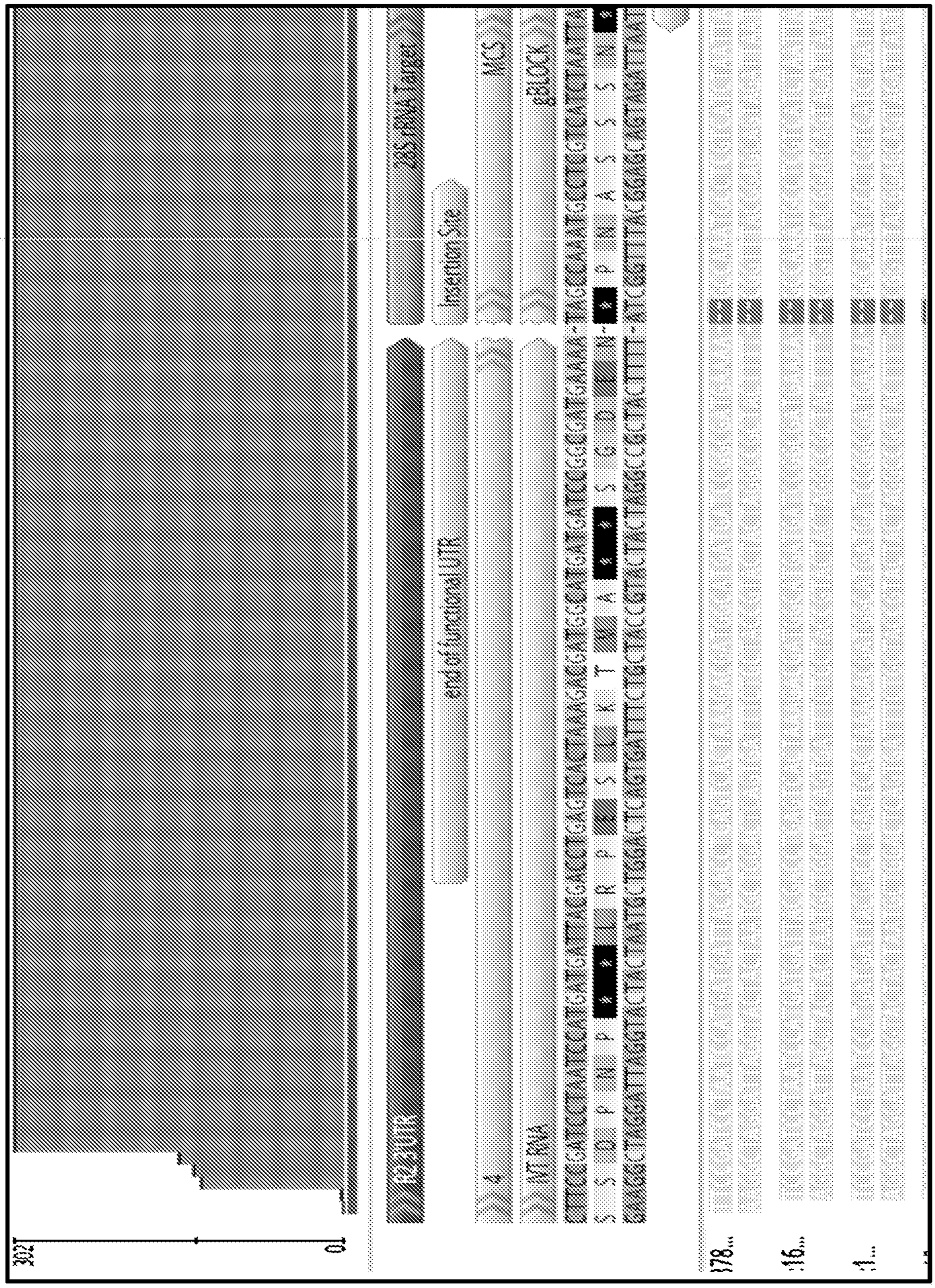

302
0
Exon
Intron
Exon
Insertion Site
5' UTR
3' UTR
Exon
Intron
Exon
Insertion Site
5' UTR
3' UTR
Exon
Intron
Exon
Insertion Site
5' UTR
3' UTR
178
116
1

FIG. 11A

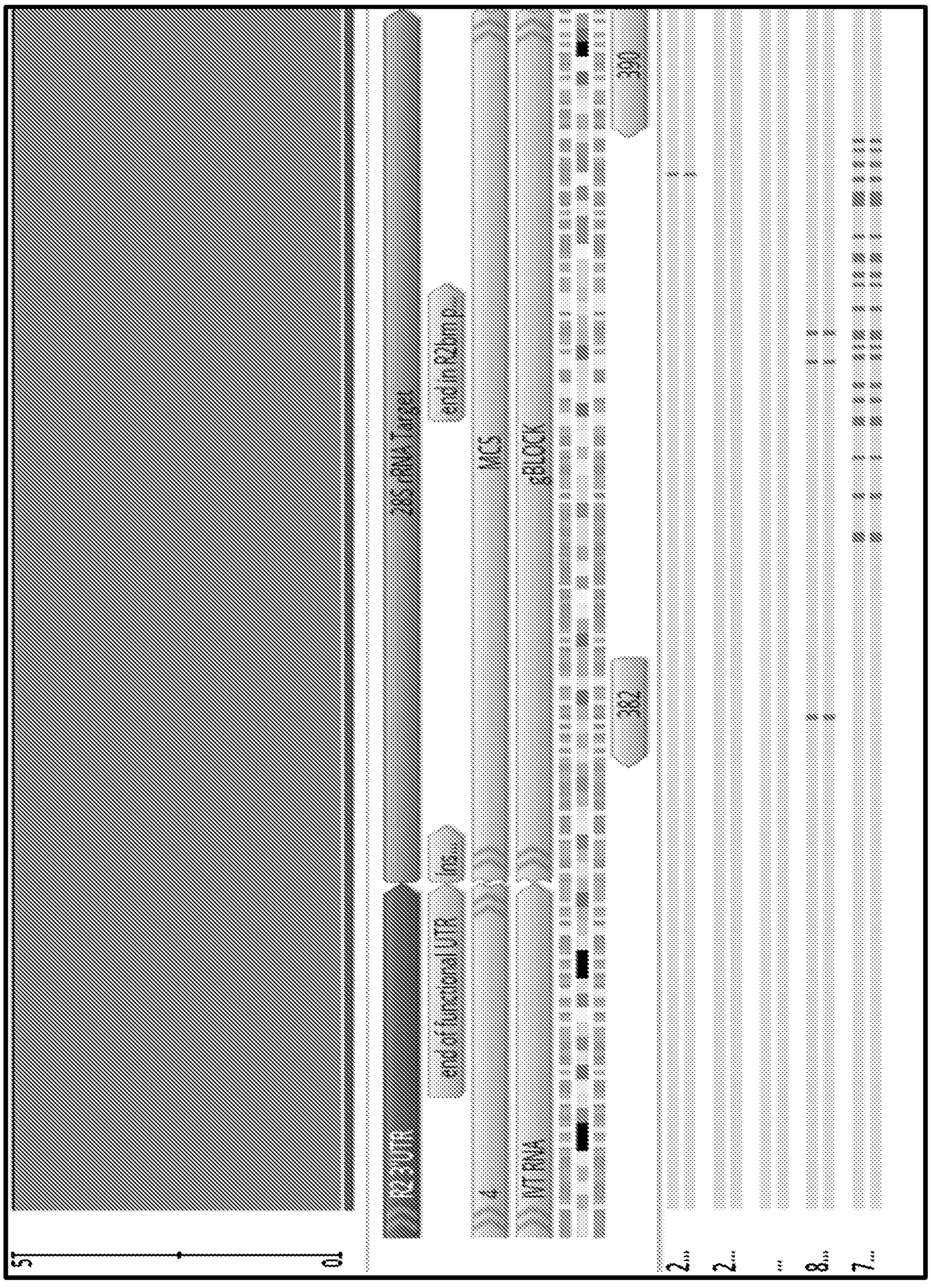

5'
0'
RC-DUH
end of hrs...
hrs...
end of 23m...
23S rRNA Target
MCS
BLOCK
3'
2'''
2''''
8'''
7''''

FIG. 11B

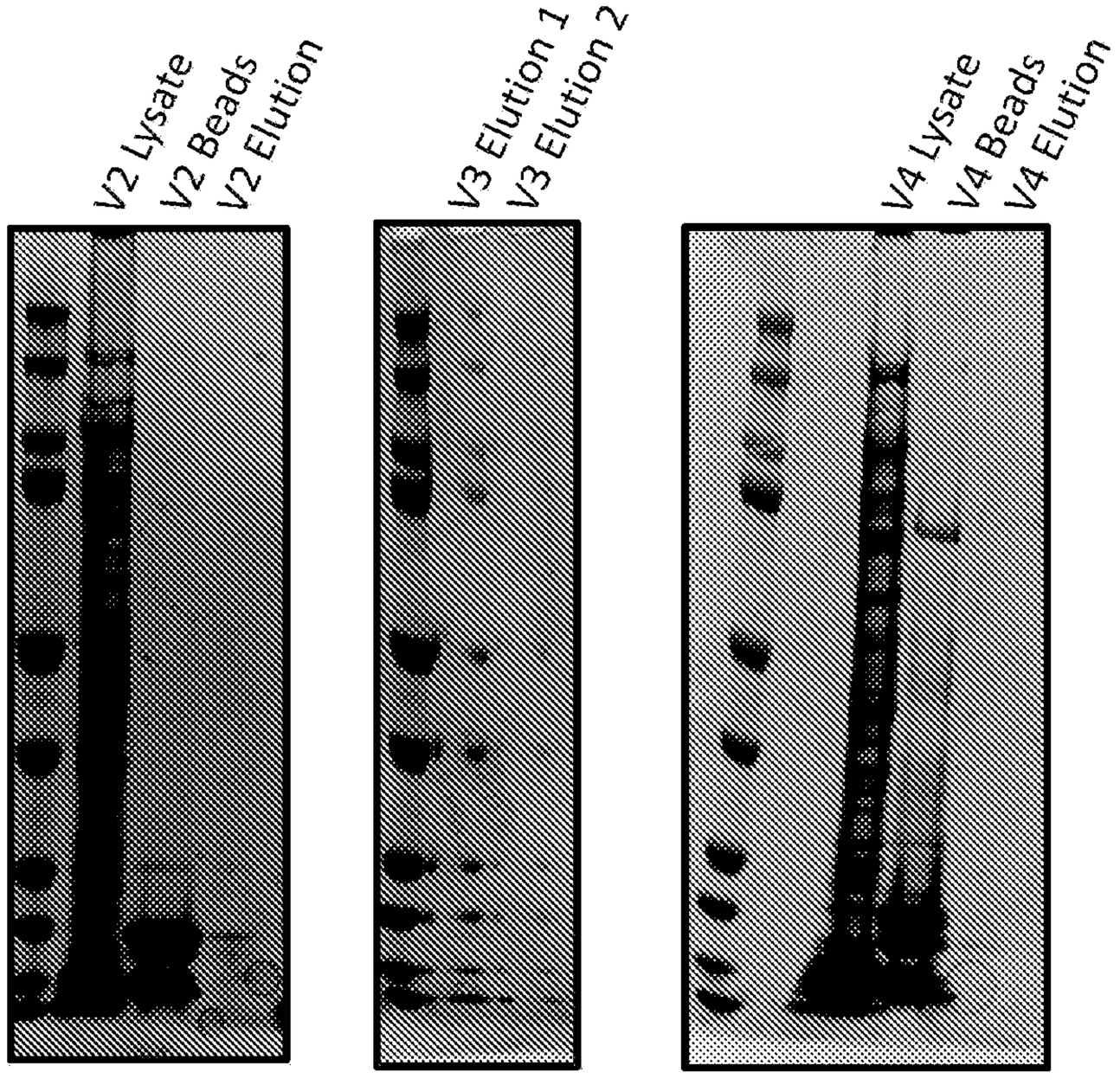

V2 Lysate
V2 Beads
V2 Elution
V3 Elution 1
V3 Elution 2
V4 Lysate
V4 Beads
V4 Elution

FIG. 12

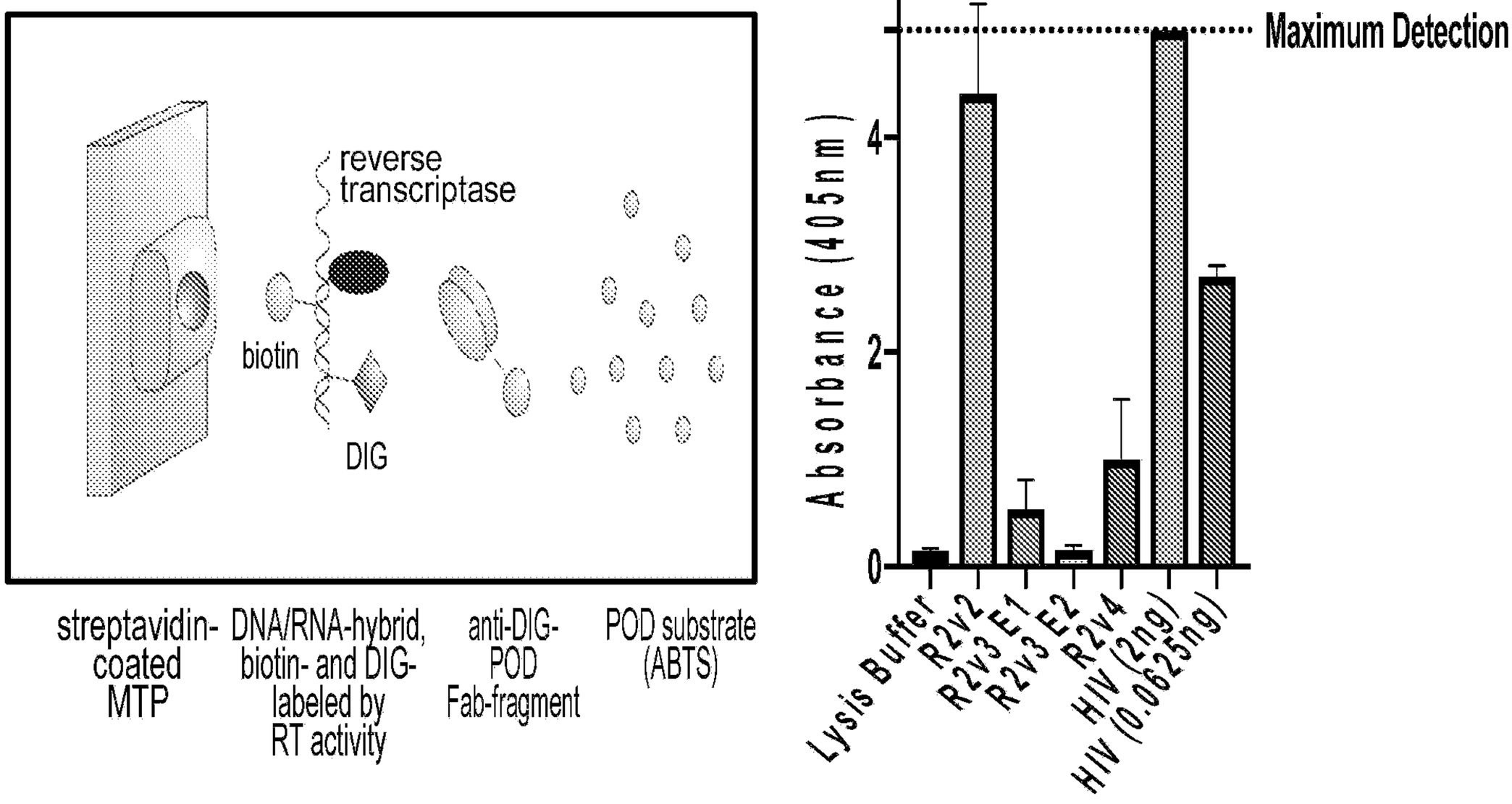

reverse transcriptase
biotin
DIG
streptavidin-coated MTP
DNA/RNA-hybrid, biotin- and DIG-labeled by RT activity
anti-DIG-POD Fab-fragment
POD substrate (ABTS)
Absorbance (405nm)
6
4
2
0
Maximum Detection
Lysis Buffer
R2v2
R2v3 E1
R2v3 E2
R2v4
HIV (2ng)
HIV (0.0625ng)

FIG. 13

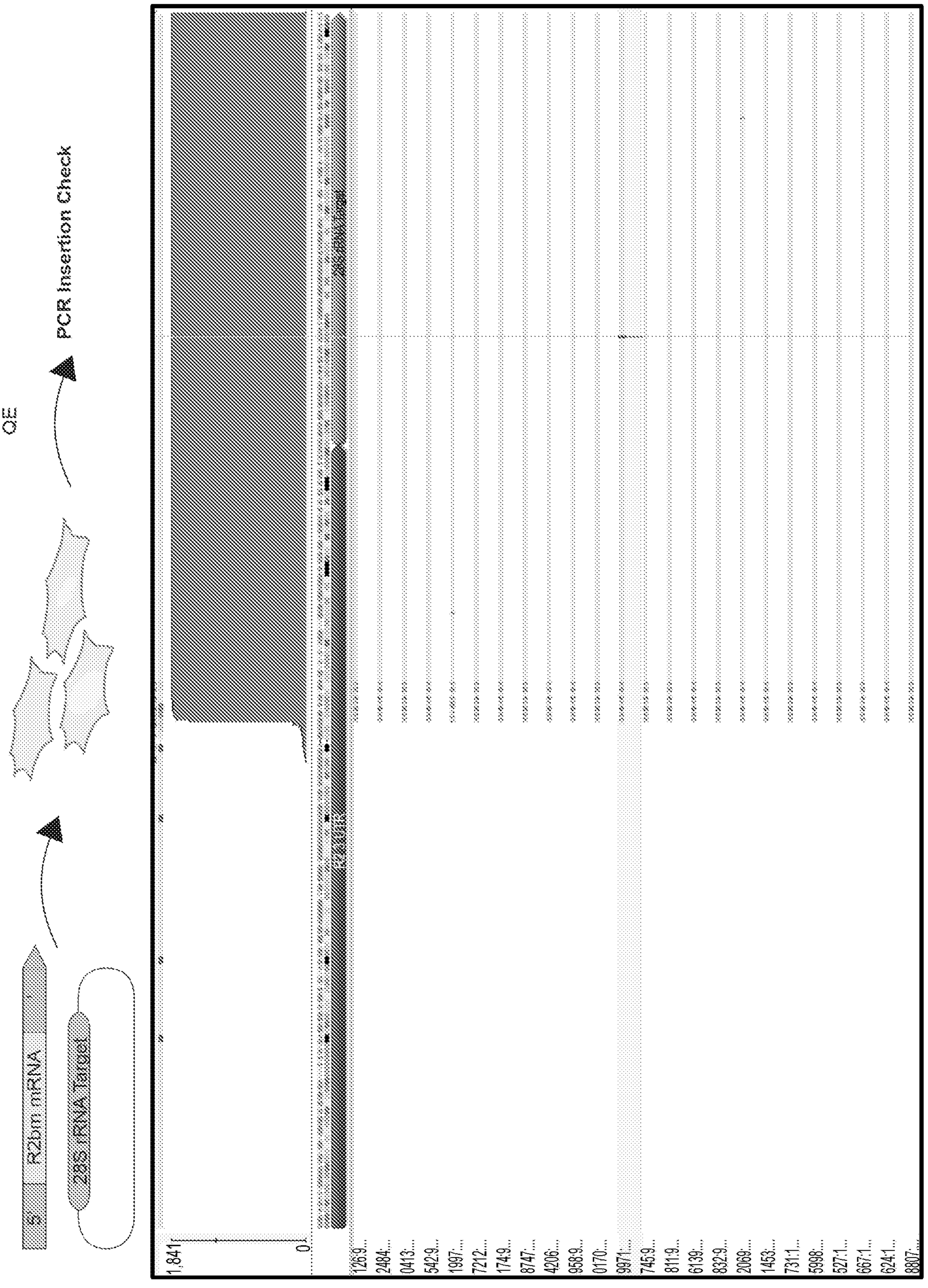

PCR insertion Check
OE
R2bm mRNA
28S rRNA Target
1,841
0
126:9...
2484:...
0413:...
542:9...
1997:...
7212:...
174:9...
8747:...
4206:...
958:9...
0170:...
9971:...
745:9...
811:9...
6139:...
832:9...
2069:...
1453:...
731:1...
5998:...
527:1...
667:1...
624:1...
8807:...

FIG. 14

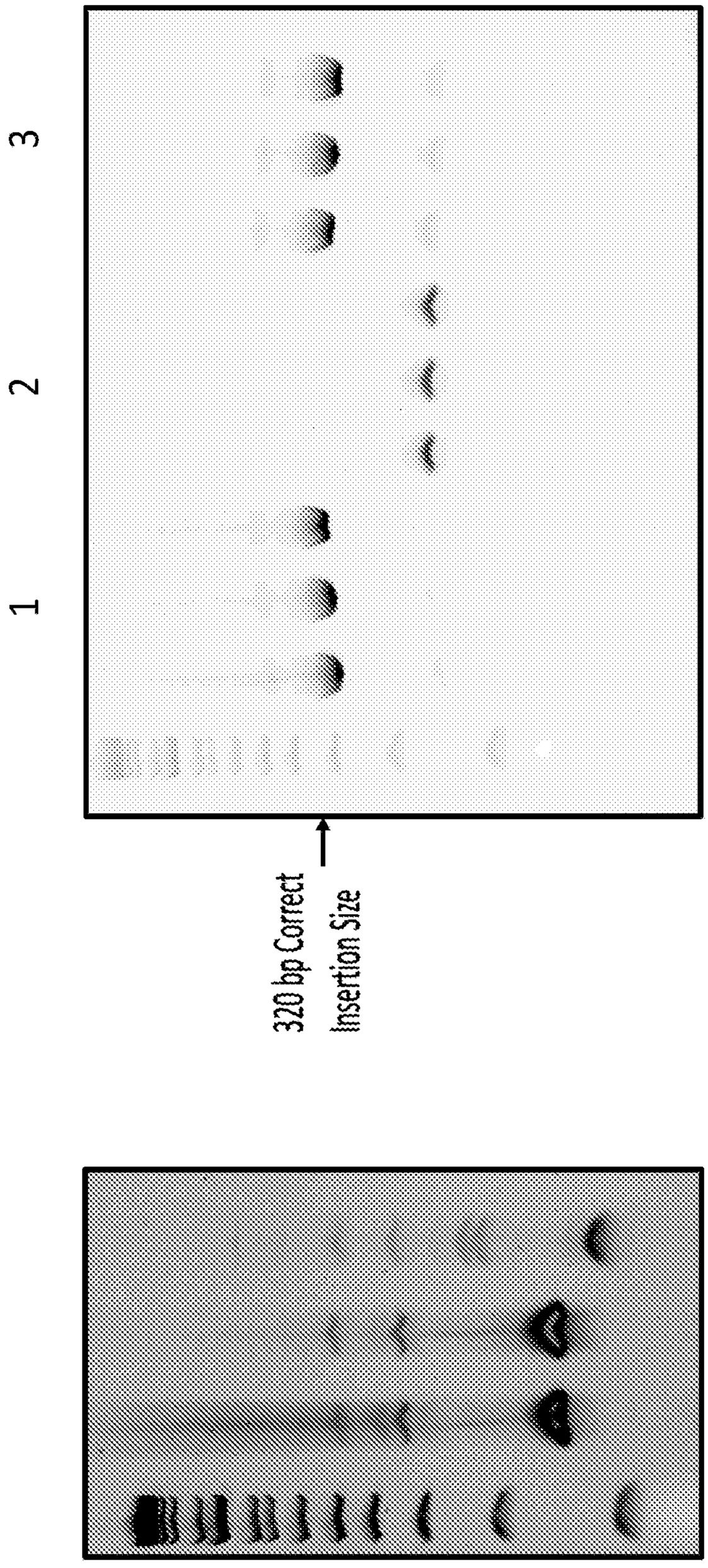

1. R2bm DNA + Target Plasmid lysate
2. R2bm ORF (no 3' UTR) + Target lysate
3. R2bm DNA
320 bp Correct Insertion Size
1. R2bm mRNA + Target Plasmid lysate
2. R2bm mRNA lysate + Target lysate
3. R2bm mRNA lysate
320 bp Correct Insertion Size

FIG. 15

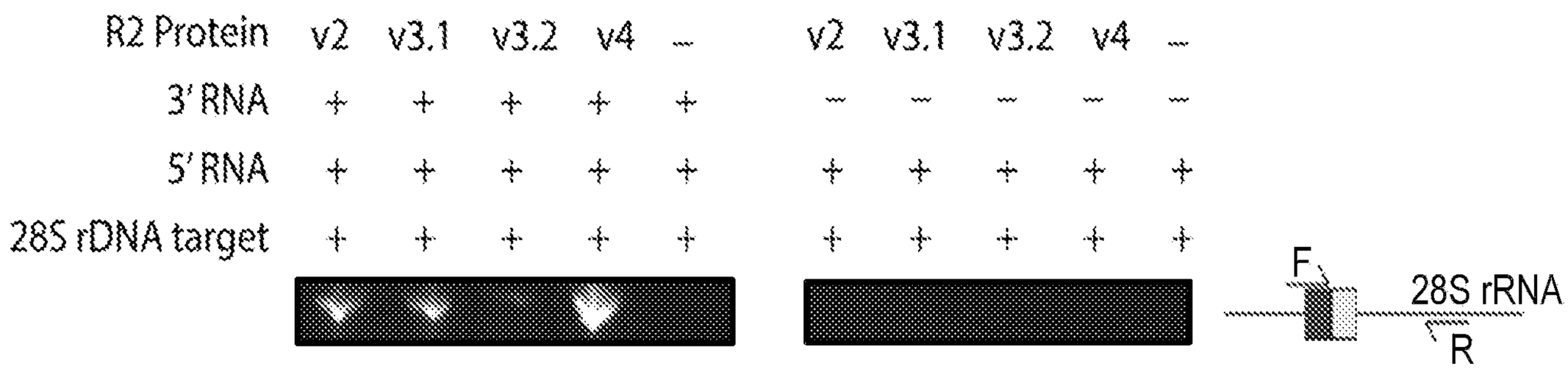


FIG. 16

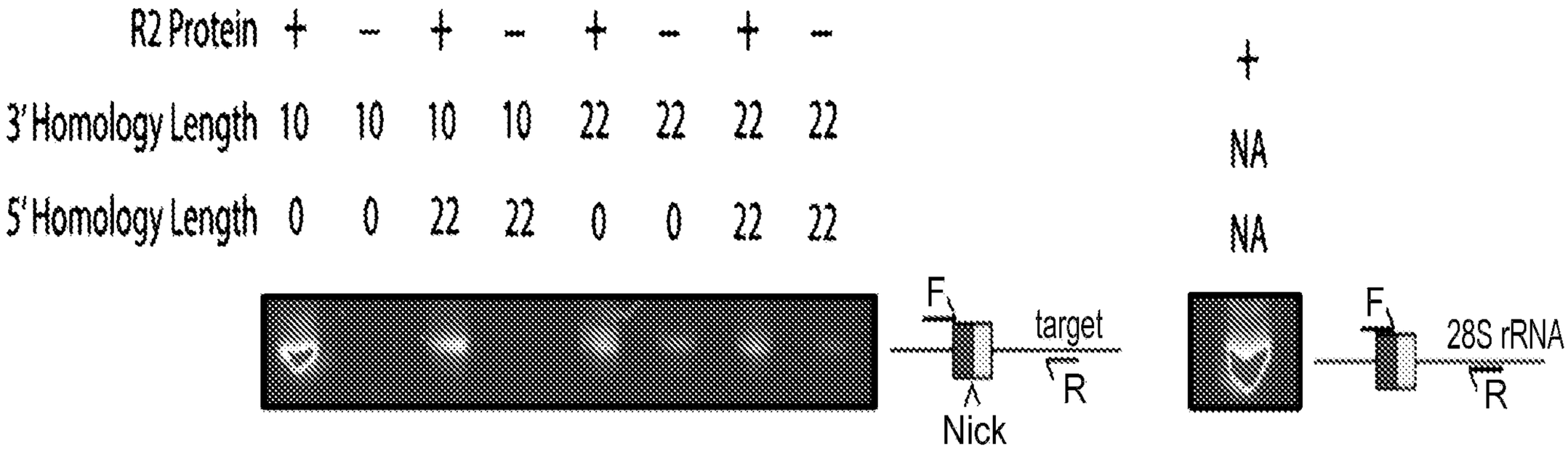


FIG. 17

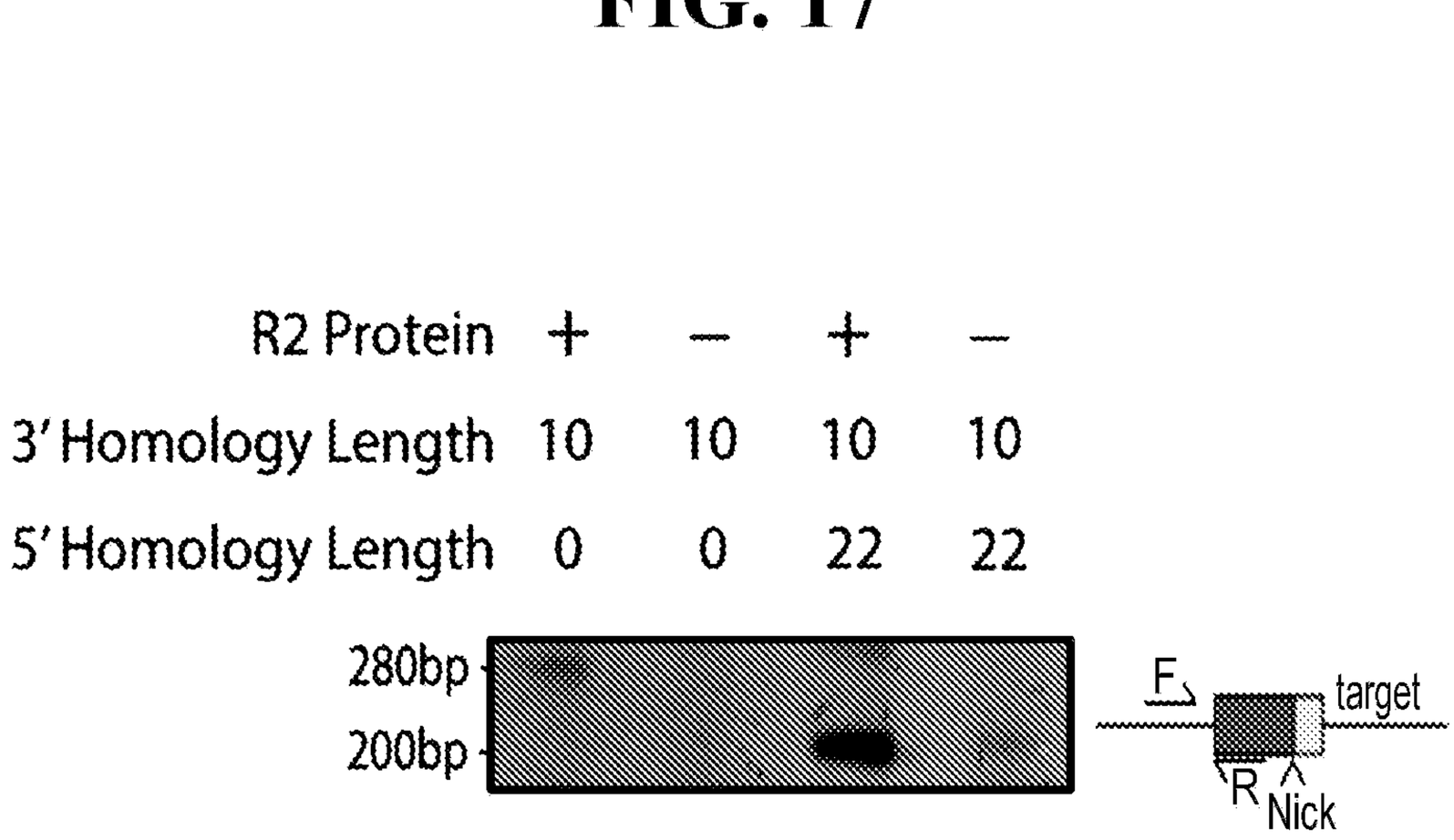


FIG. 18

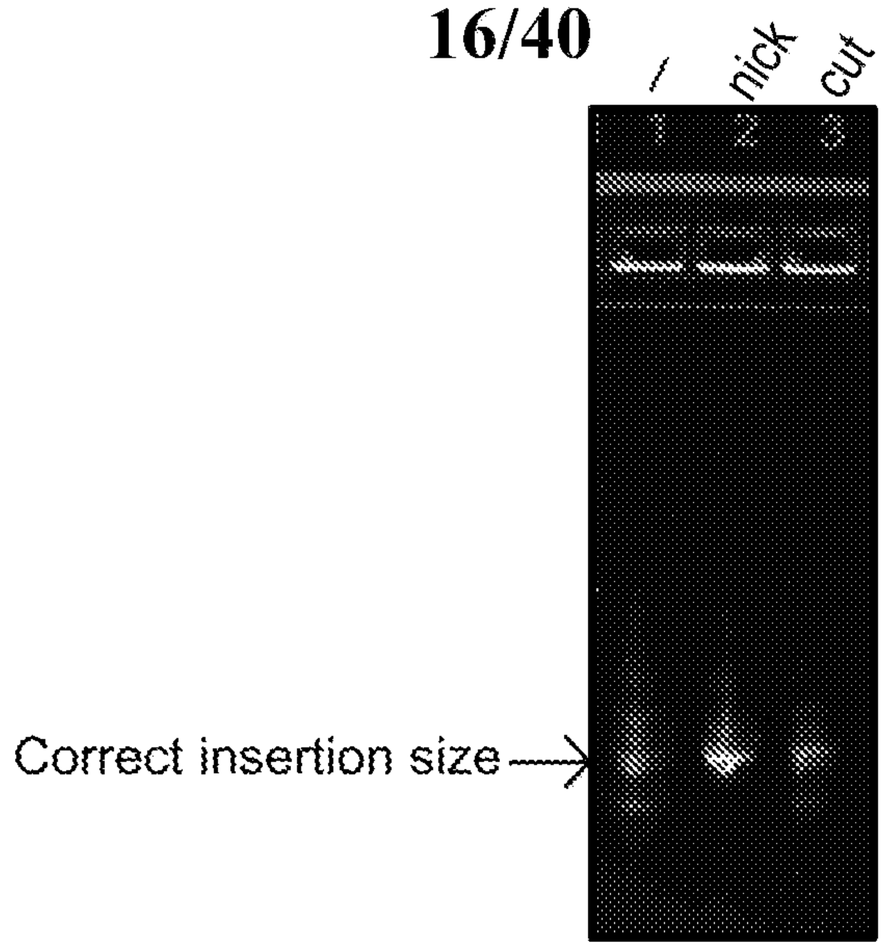


FIG. 19

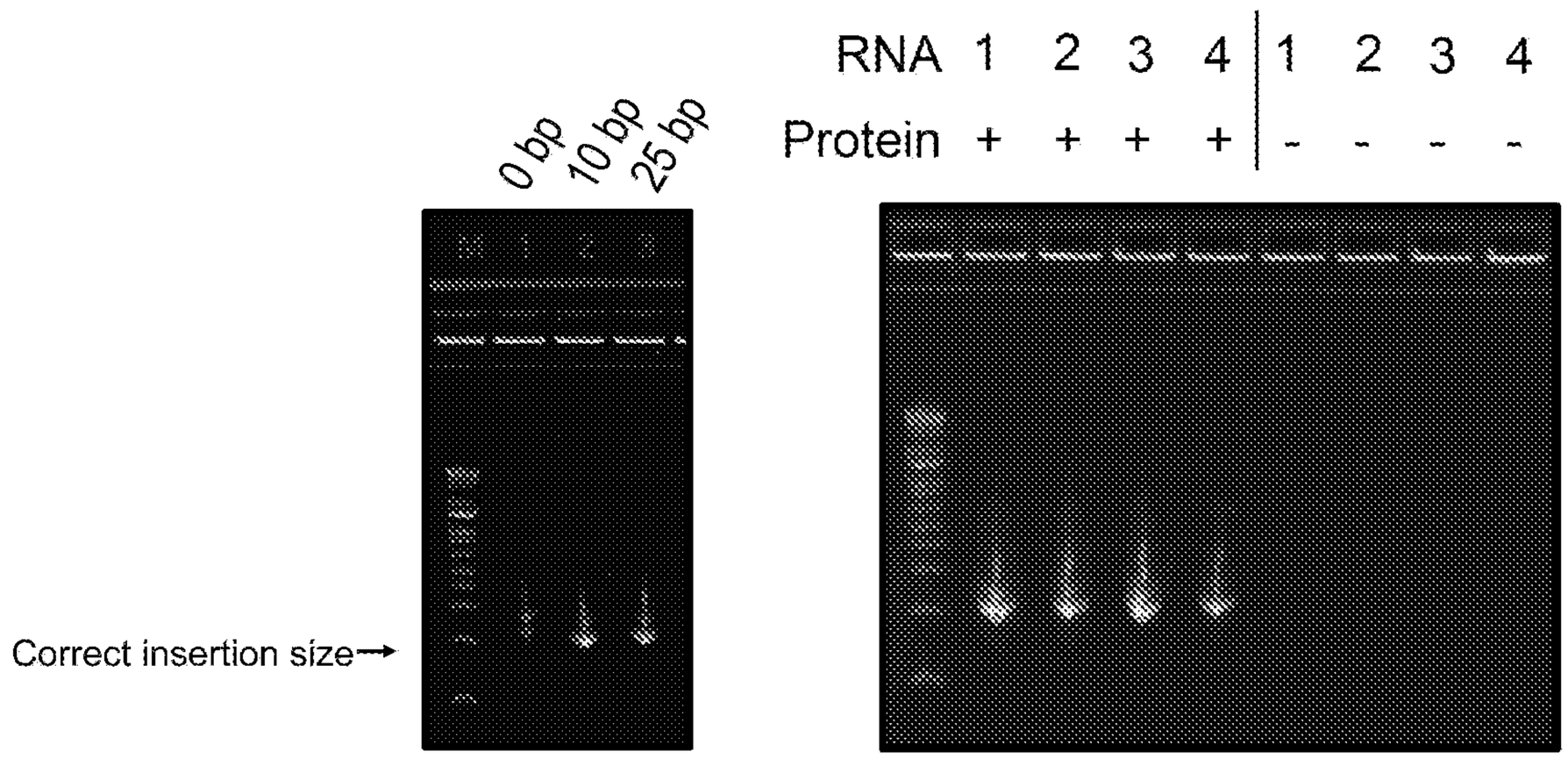


FIG. 20A

FIG. 20B

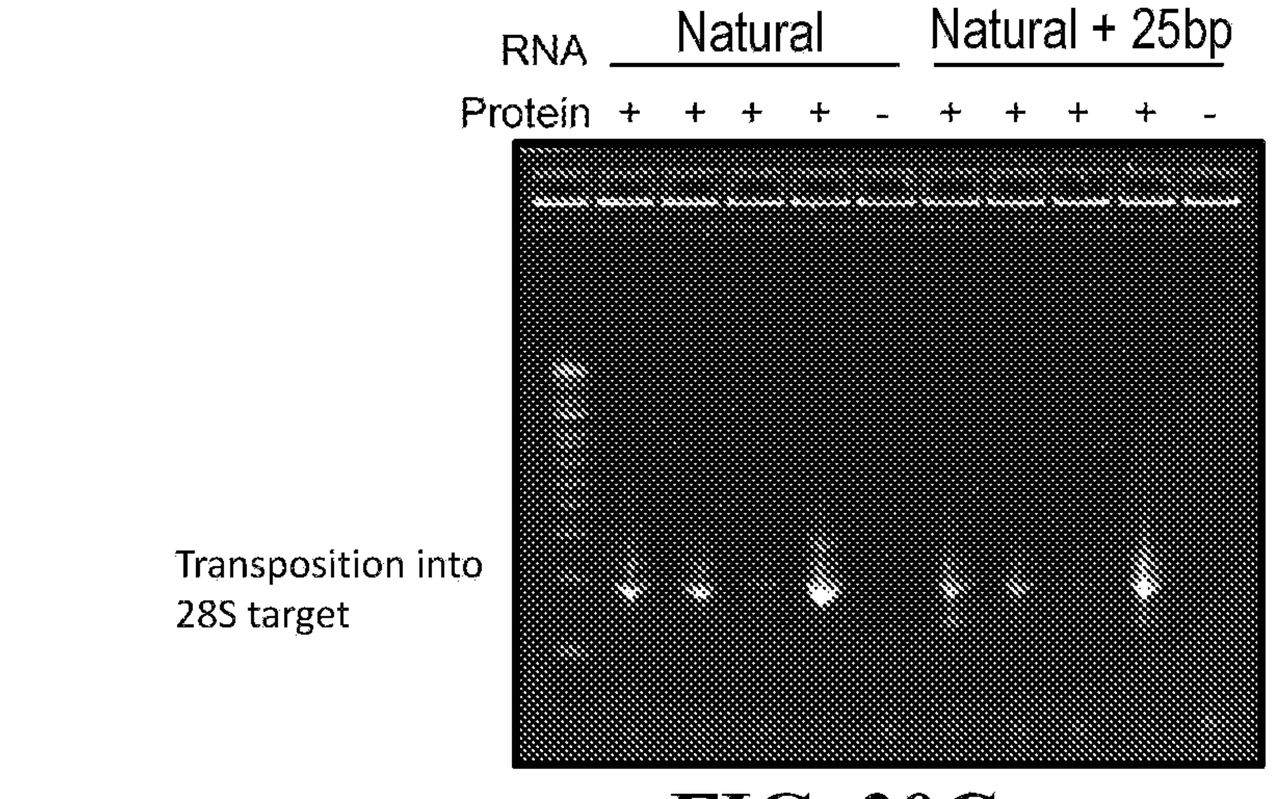


FIG. 20C

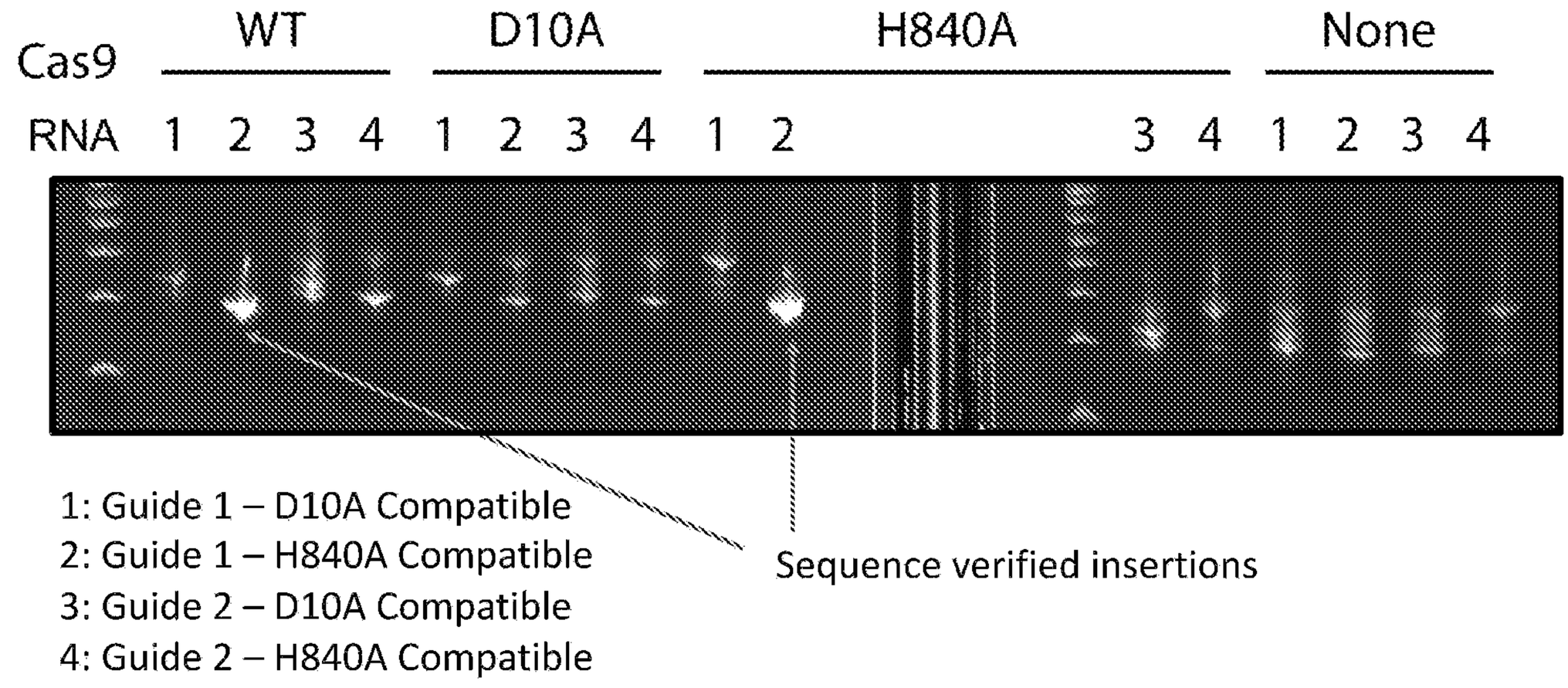

Cas9
WT
D10A
H840A
None
RNA 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
1: Guide 1 – D10A Compatible
2: Guide 1 – H840A Compatible
3: Guide 2 – D10A Compatible
4: Guide 2 – H840A Compatible
Sequence verified insertions

FIG. 21

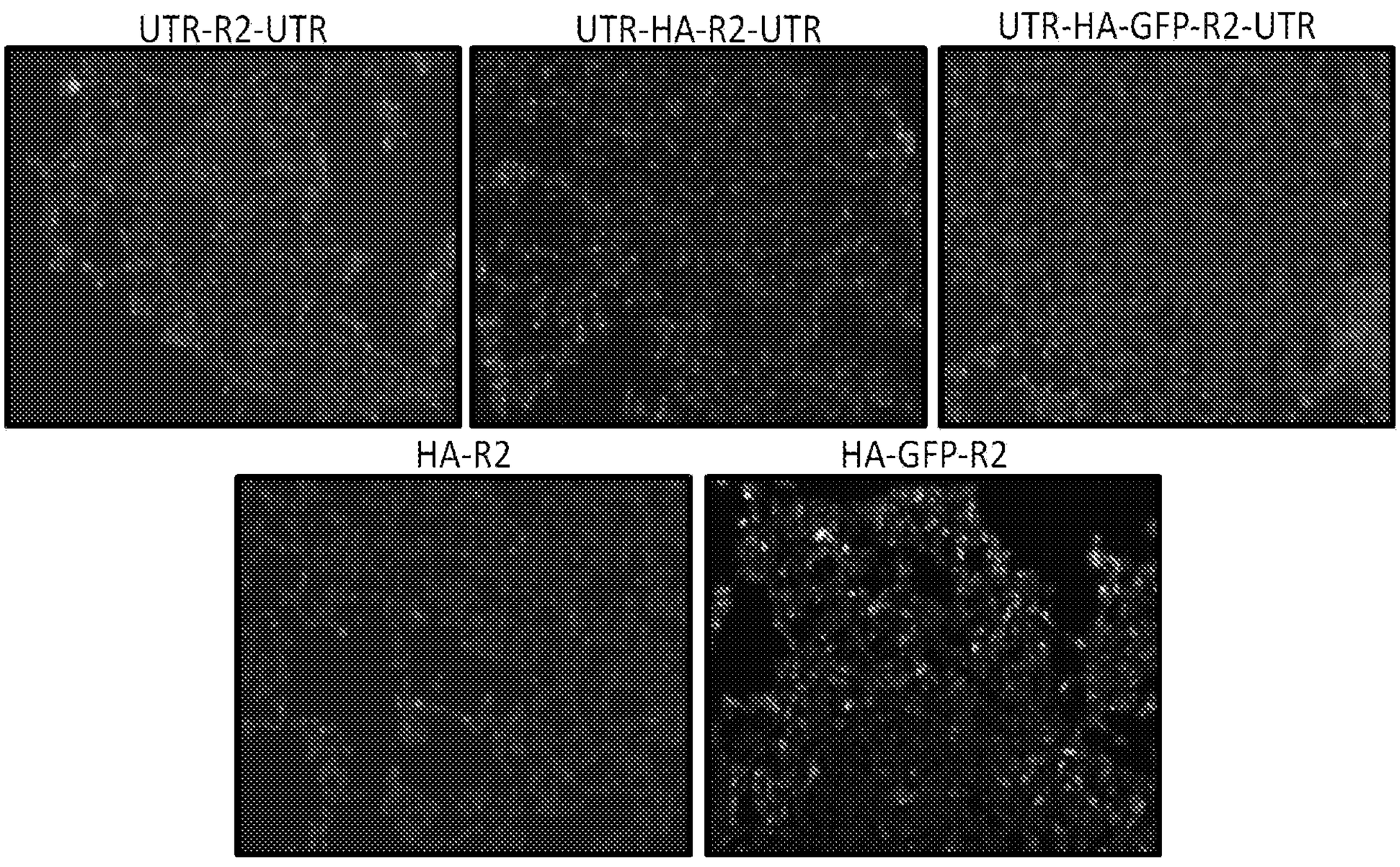

UTR-R2-UTR
UTR-HA-R2-UTR
UTR-HA-GFP-R2-UTR
HA-R2
HA-GFP-R2

FIG. 22

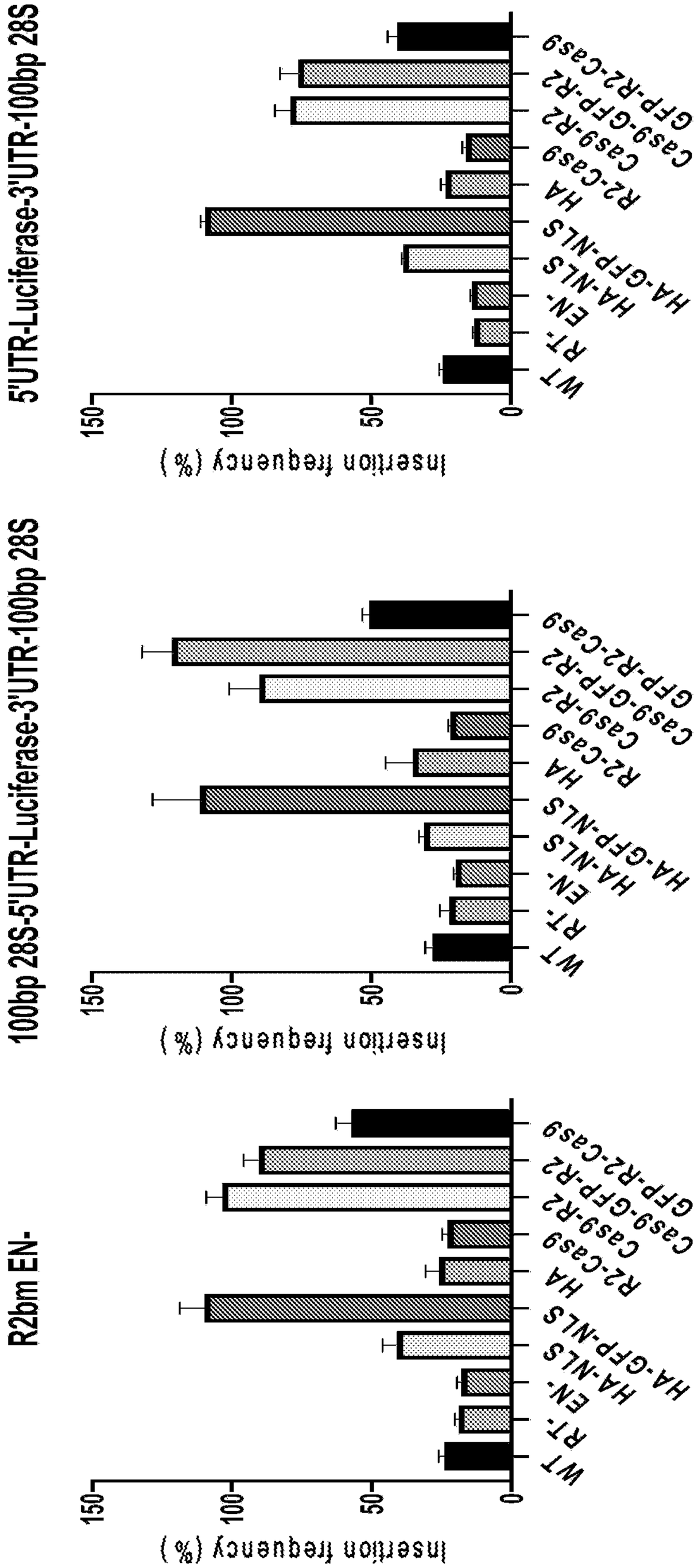

5'UTR-Luciferase-3'UTR-100bp 28S
100bp 28S-5'UTR-Luciferase-3'UTR-100bp 28S
R2bm EN-
insertion frequency (%)
150
100
50
0
WT
RT-
EN-
HA-NLS
HA-GFP-NLS
HA
R2-Cas9
Cas9-R2
Cas9-GFP-R2
GFP-R2-Cas9

FIG. 23

FIG. 24

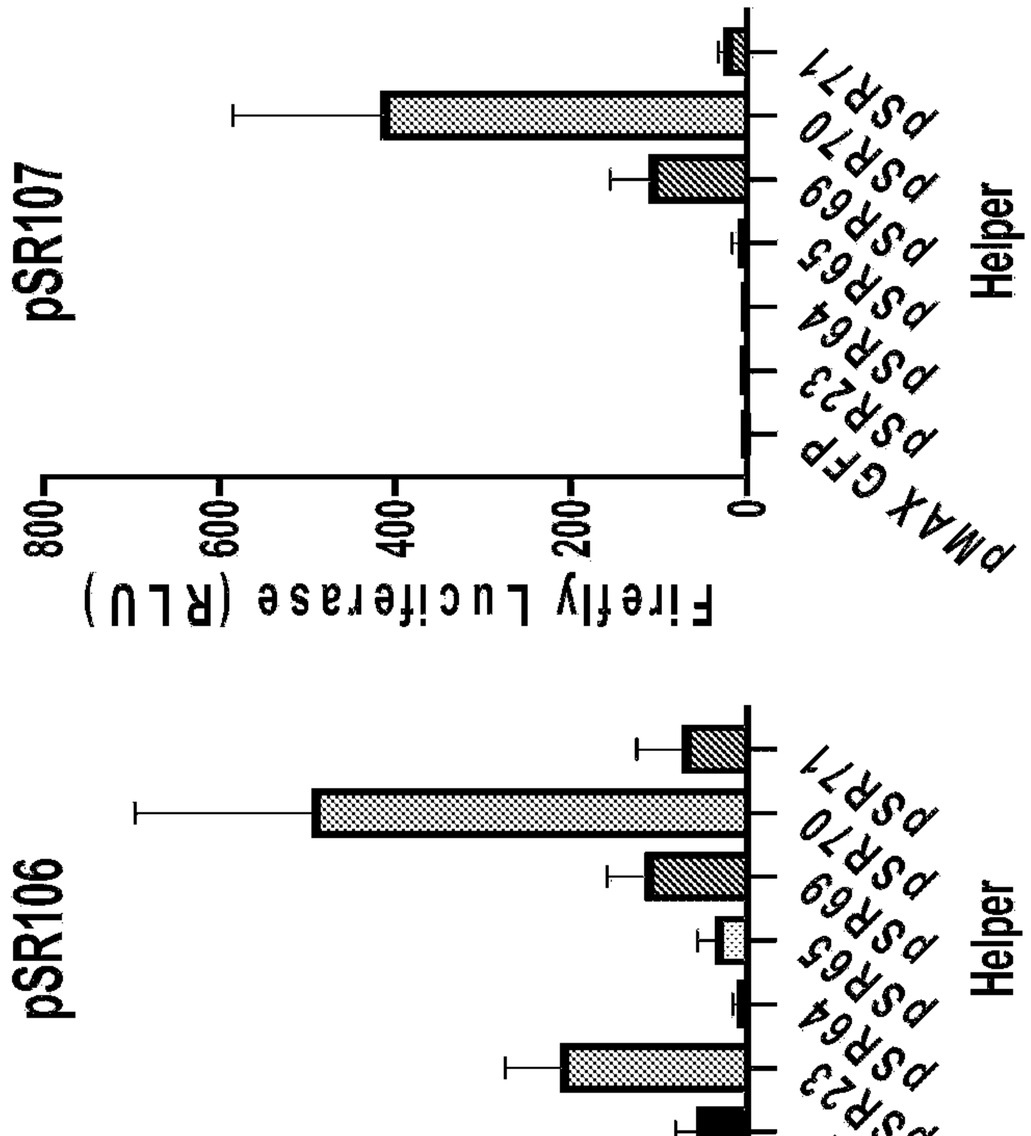


FIG. 25

FIG. 26

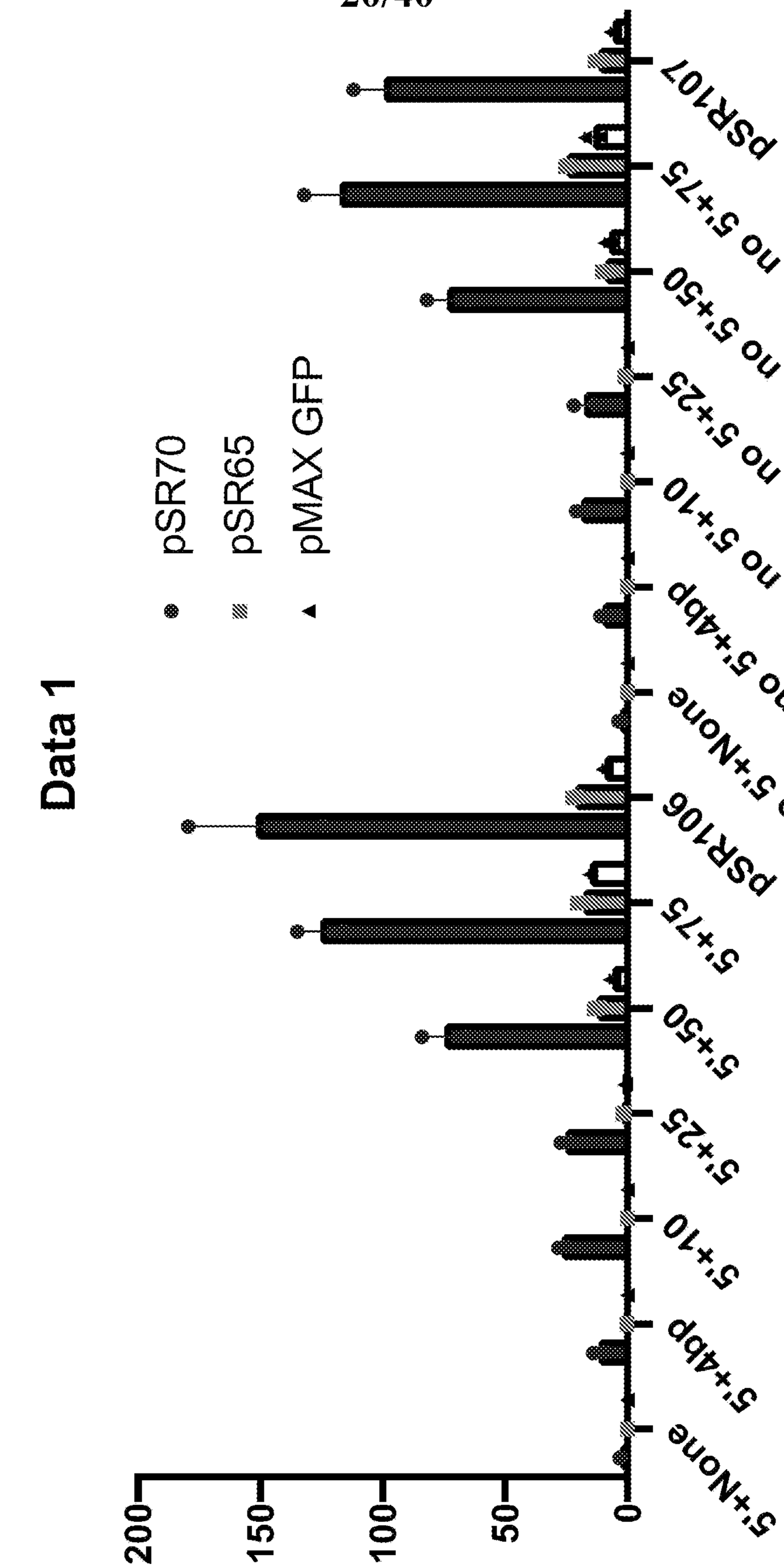

Data 1
pSR70
pSR65
pMAX GFP
pSR107
5'+75 nt
5'+50 nt
5'+25 nt
5'+10 nt
5'+4bp
5'+None
pSR106
5'+75
5'+50
5'+25
5'+10
5'+4bp
5'+None
200
150
100
50
0
Insertion frequency (%)

FIG. 26A

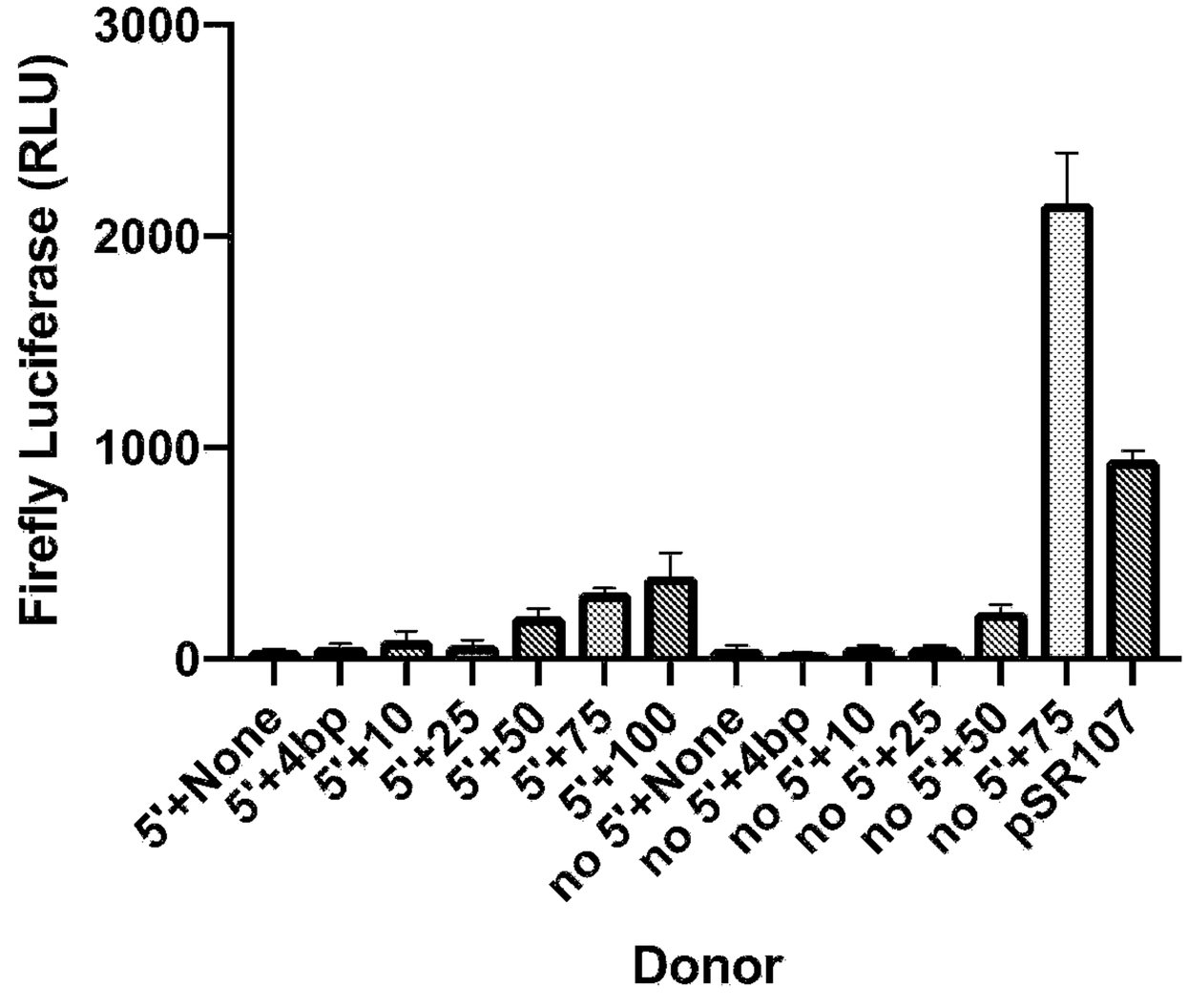


FIG. 26B

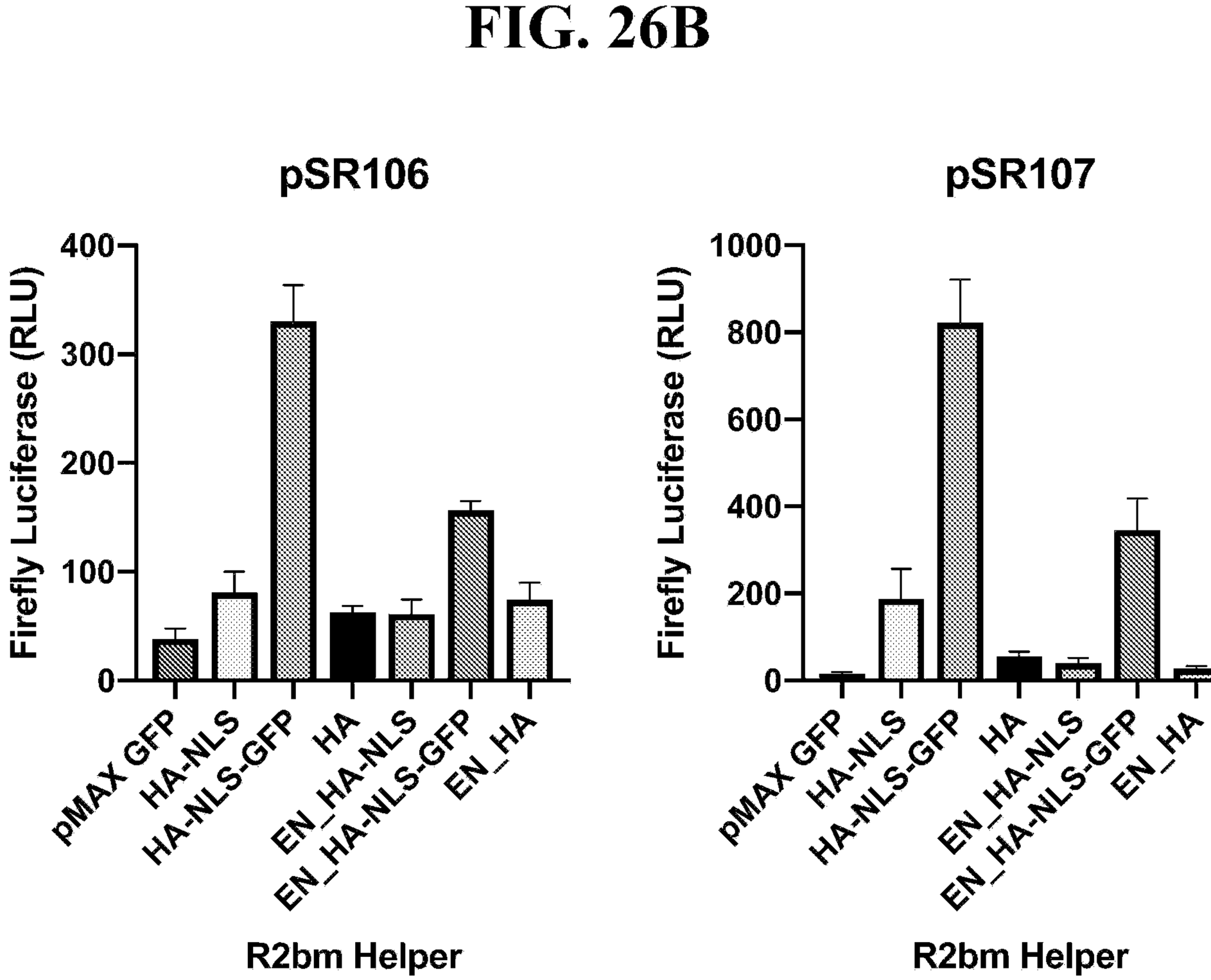


FIG. 27

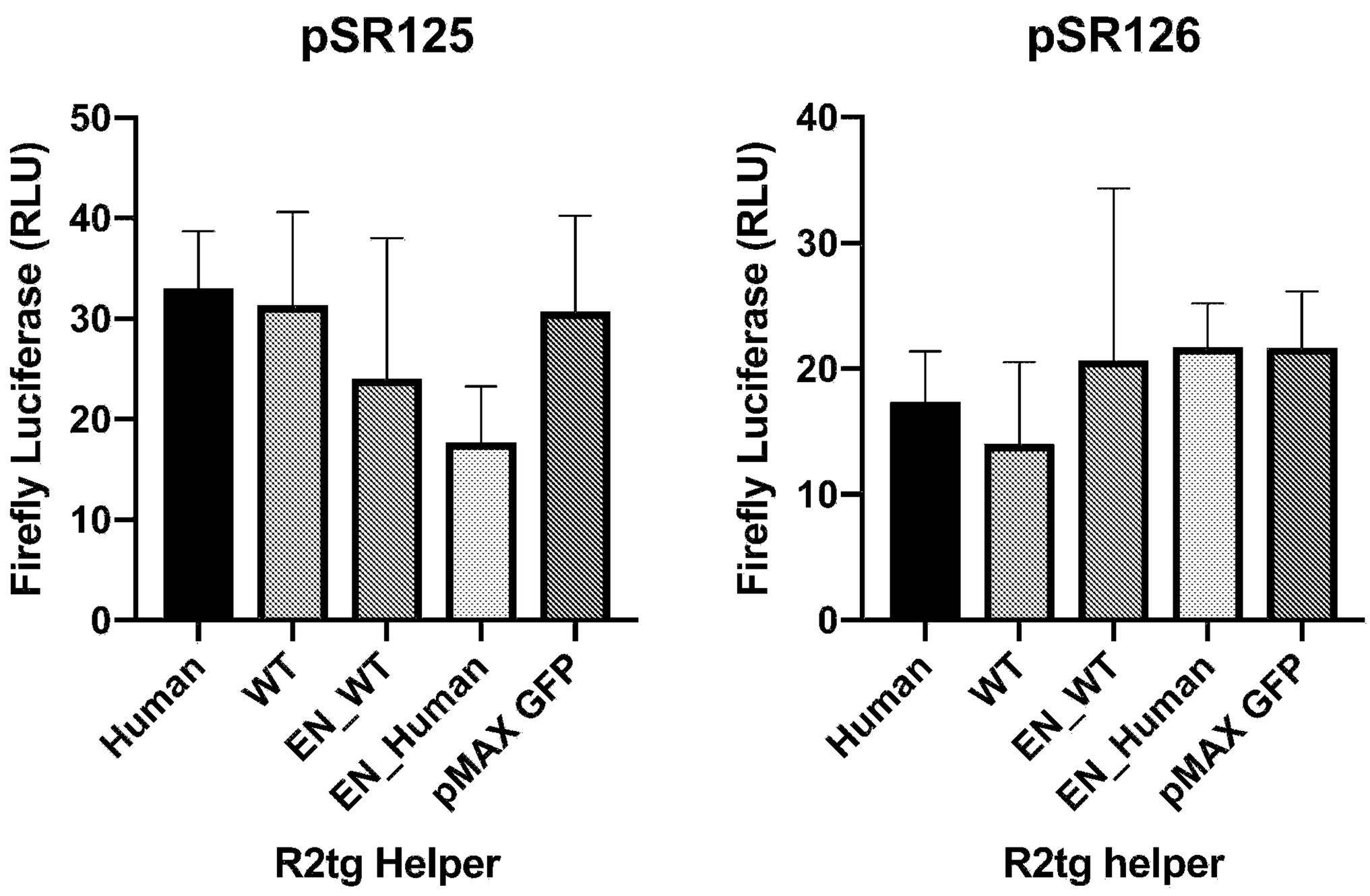

pSR125
Firefly Luciferase (RLU)
50
40
30
20
10
0
Human
WT
EN_WT
EN_Human
pMAX GFP
R2tg Helper
pSR126
Firefly Luciferase (RLU)
40
30
20
10
0
Human
WT
EN_WT
EN_Human
pMAX GFP
R2tg helper

FIG. 28

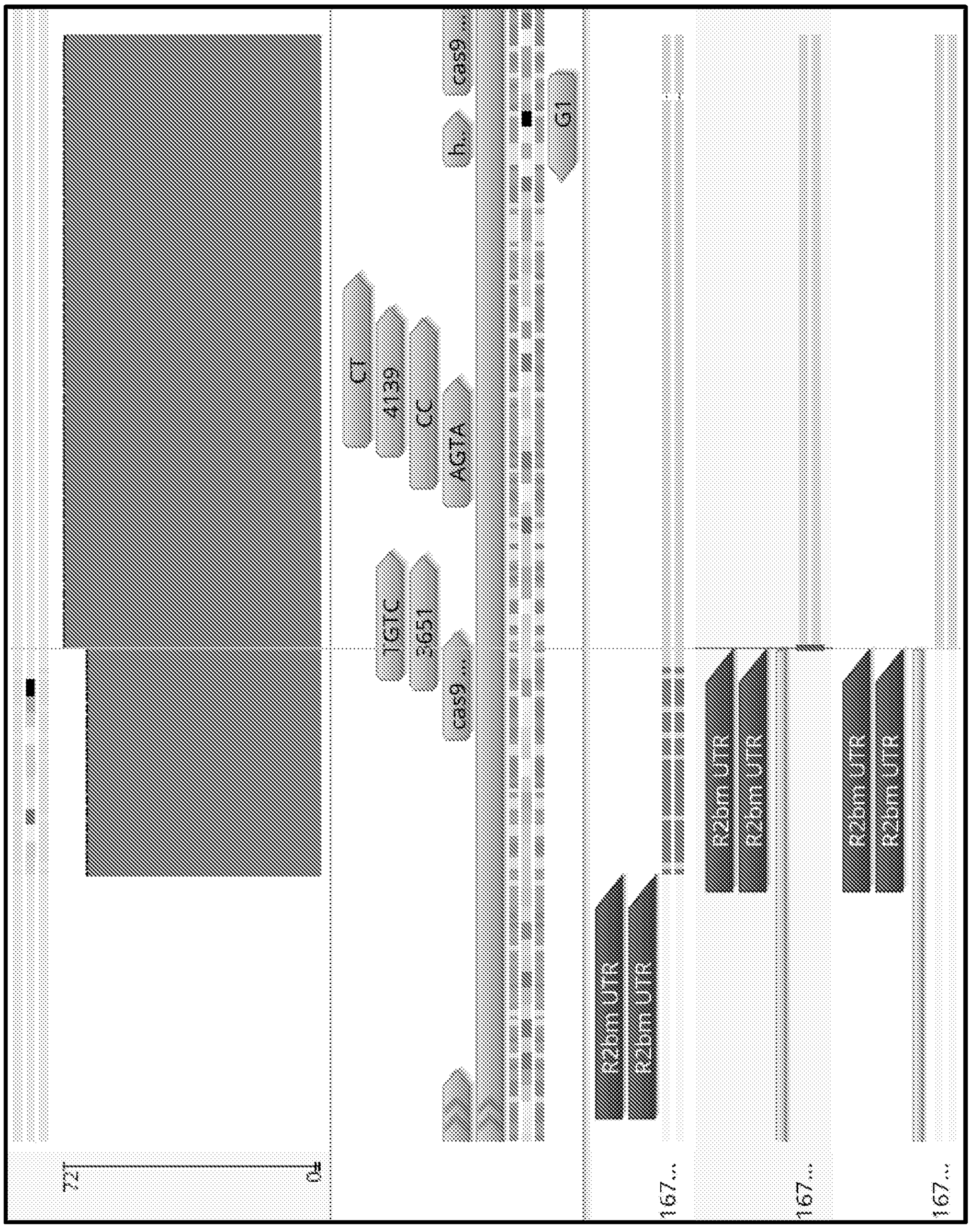

cas9...
h...
G1
CT
4139
CC
AGTA
TGTC
2651
cas9...
R2bm UTR
R2bm UTR
R2bm UTR
R2bm UTR
R2bm UTR
R2bm UTR
167...
167...
167...
1274
80

FIG. 29

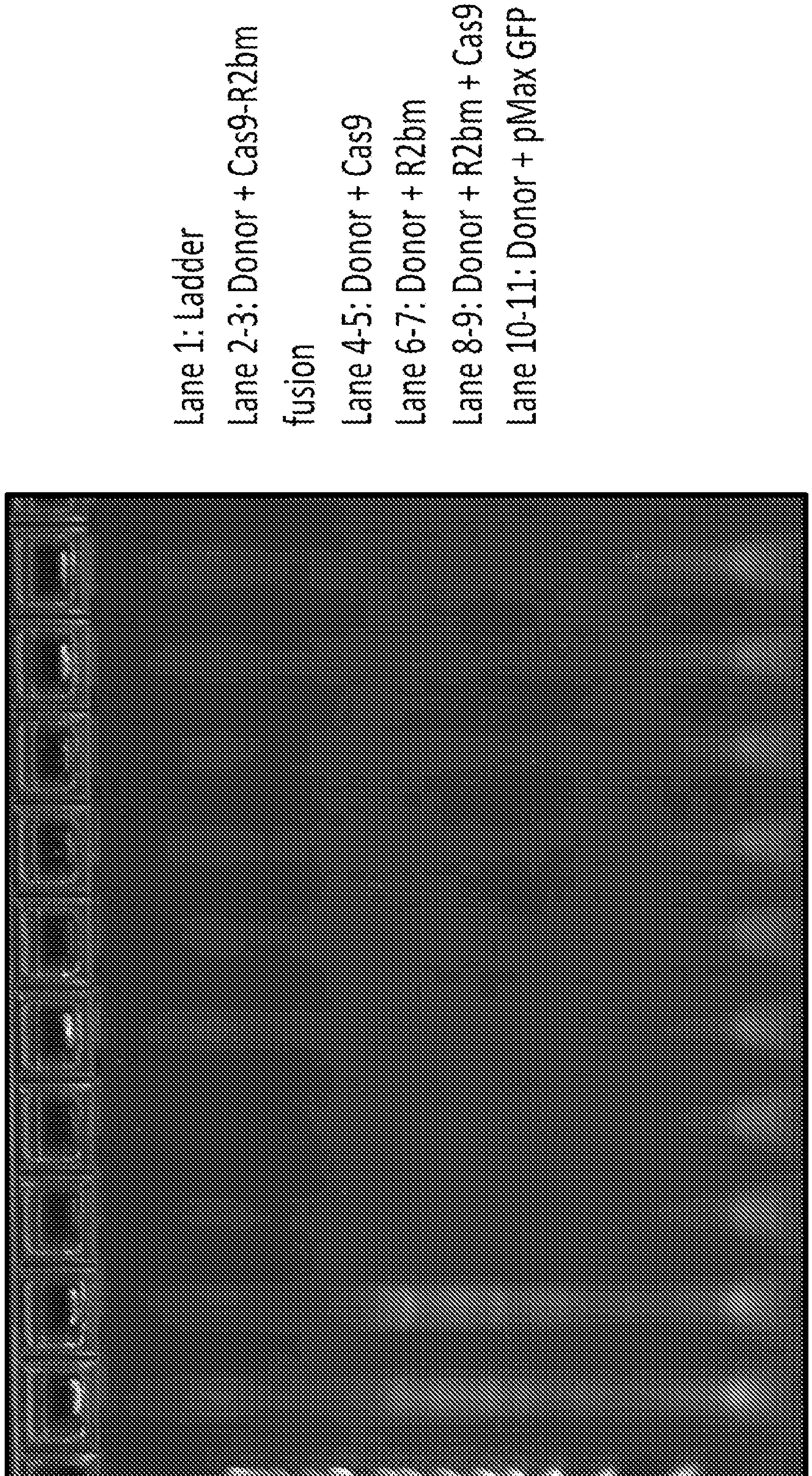

Lane 1: Ladder
Lane 2-3: Donor + Cas9-R2bm fusion
Lane 4-5: Donor + Cas9
Lane 6-7: Donor + R2bm
Lane 8-9: Donor + R2bm + Cas9
Lane 10-11: Donor + pMax GFP

FIG. 30

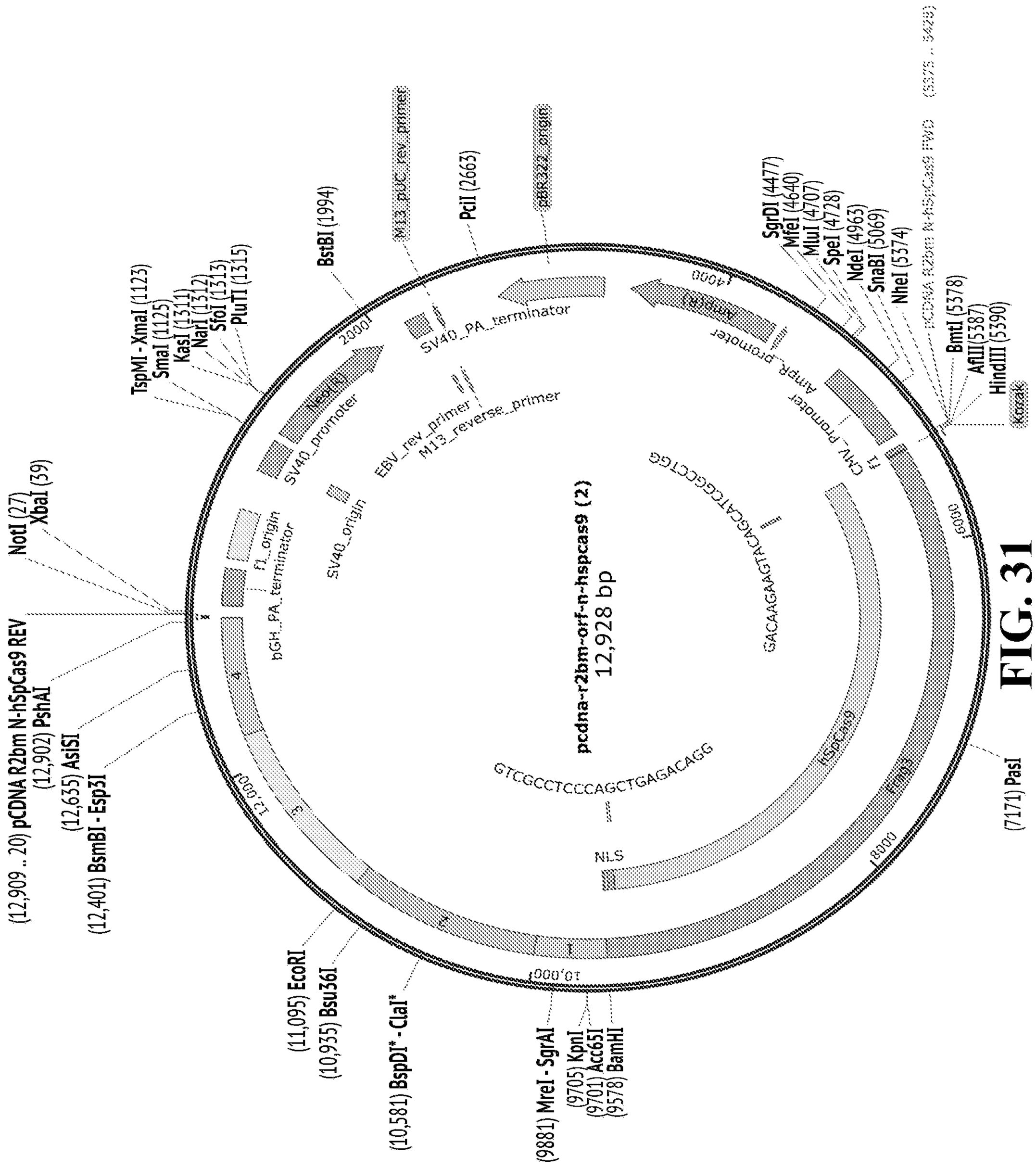

pcdna-r2bm-orf-n-hspcas9 (2)
12,928 bp
(12,909 .. 20) pCDNA R2bm N-hSpCas9 REV
(12,902) PshAI
(12,635) AsiSI
(12,401) BsmBI - Esp3I
(11,095) EcoRI
(10,935) Bsu36I
(10,581) BspDI* - ClaI*
(9881) MreI - SgrAI
(9705) KpnI
(9701) Acc65I
(9578) BamHI
(7171) PasI
(27) NotI
(39) XbaI
TspMI - XmaI (1123)
SmaI (1125)
KasI (1311)
NarI (1312)
SfoI (1313)
PluTI (1315)
BstBI (1994)
M13_pUC_rev_primer
PciI (2663)
pBR322_origin
SgrDI (4477)
MfeI (4640)
MluI (4707)
SpeI (4728)
NdeI (4963)
SnaBI (5069)
NheI (5374)
BmtI (5378)
AflII (5387)
HindIII (5390)
KOZAK
SV40_PA_terminator
NEOMYCIN
SV40_promoter
EBV_rev_primer
M13_reverse_primer
AmpR_promoter
CMV_promoter
f1_origin
bGH_PA_terminator
SV40_origin
GTCGCCTCCCAGCTGAGACAGG
GACAAGAAGTACAGCATCGGCCTGG
NLS
hSPCas9
ECMV
12,000
10,000
8000
6000
4000
2000

FIG. 31

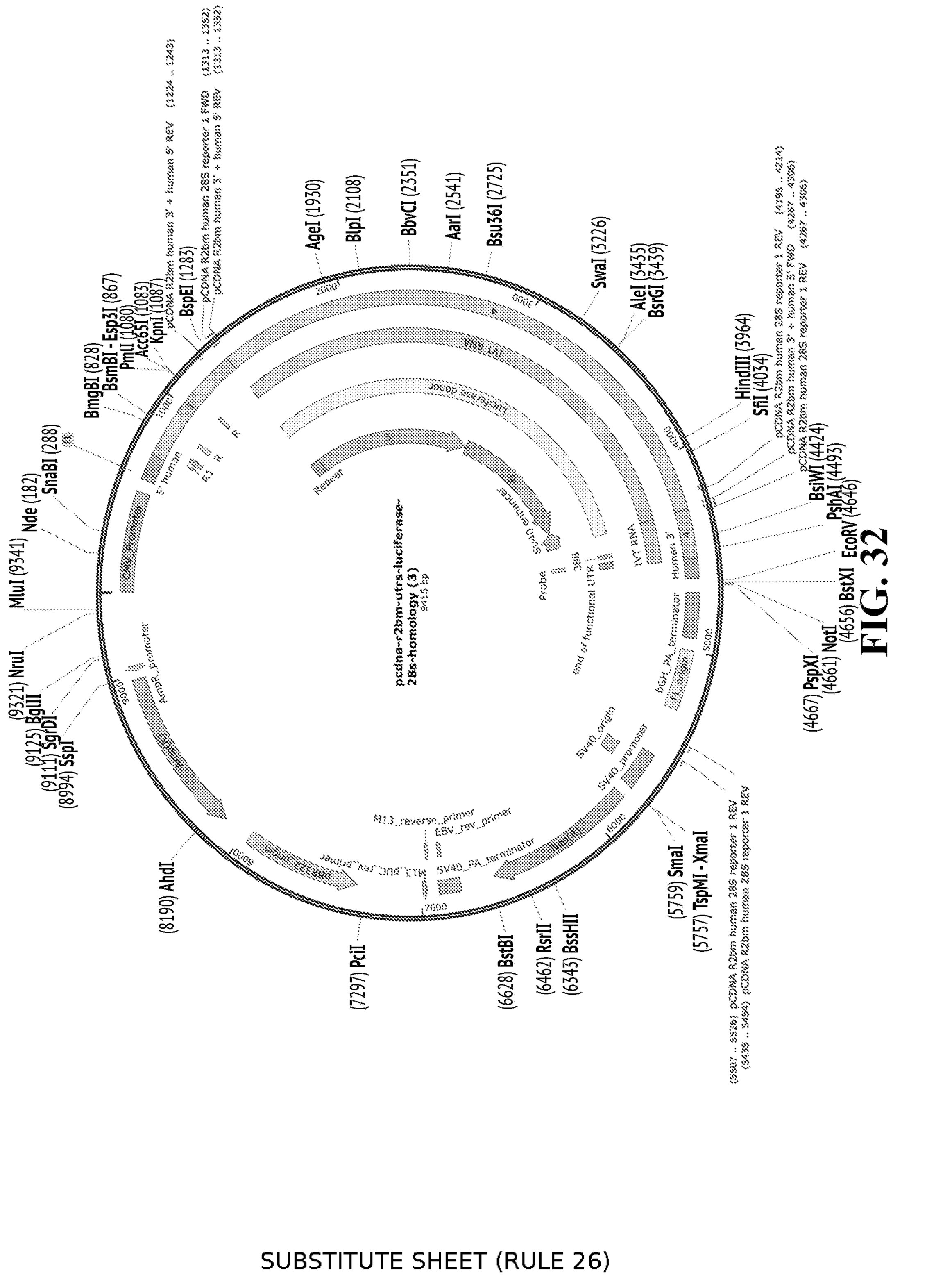
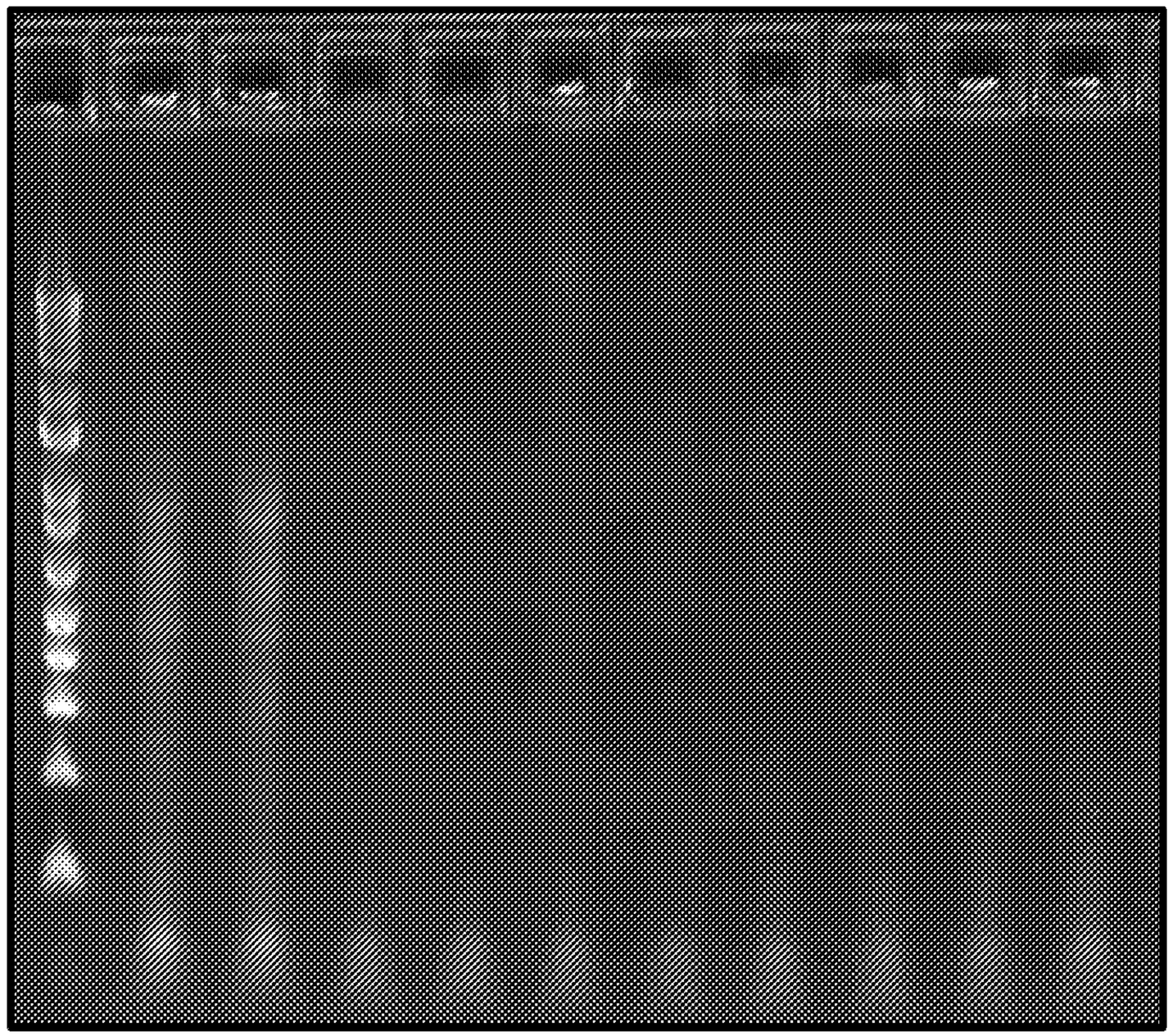

pcdna-r2bm-utrs-luciferase-2Bs-homology (3)
9415 bp
MluI (9341)
NdeI (182)
SnaBI (288)
BmgBI (828)
BsmBI - Esp3I (867)
PmlI (1080)
Acc65I (1083)
KpnI (1087)
BspEI (1283)
AgeI (1930)
BlpI (2108)
BbvCI (2351)
AarI (2541)
Bsu36I (2725)
SwaI (3226)
AleI (3435)
BsrGI (3439)
HindIII (3964)
SfiI (4034)
BsiWI (4424)
PshAI (4493)
EcoRV (4646)
NotI (4656)
PspXI (4661)
BstXI (4667)
NruI (9321)
BglII (9125)
SgrDI (9111)
SspI (8994)
AhdI (8190)
PciI (7297)
BstBI (6628)
RsrII (6462)
BssHII (6343)
SmaI (5759)
TspMI - XmaI (5757)
Ampr_promoter
M13_reverse_primer
SV40_pro_primer
SV40_PA_terminator
SV40_origin
SV40_promoter
bGH_PA_terminator
f1_origin
Probe
end of functional UTR
Repeat
Human 3'
IVT RNA
pCDNA R2bm human 3' + human 5' REV (1224 .. 1243)
pCDNA R2bm human 3' + human 5' FWD (1313 .. 1332)
pCDNA R2bm human 3' + human 5' REV (1313 .. 1332)
pCDNA R2bm human 28S reporter 1 REV (4195 .. 4214)
pCDNA R2bm human 3' + human 5' FWD (4287 .. 4306)
pCDNA R2bm human 28S reporter 1 REV (4287 .. 4306)
(5807 .. 5826) pCDNA R2bm human 28S reporter 1 REV
(5439 .. 5458) pCDNA R2bm human 28S reporter 1 REV

FIG. 32



Lane 1: Ladder

Lane 2-3: Donor + Cas9-R2bm fusion

Lane 4-5: Donor + Cas9

Lane 6-7: Donor + R2bm

Lane 8-9: Donor + R2bm + Cas9

Lane 10-11: Donor + pMax GFP

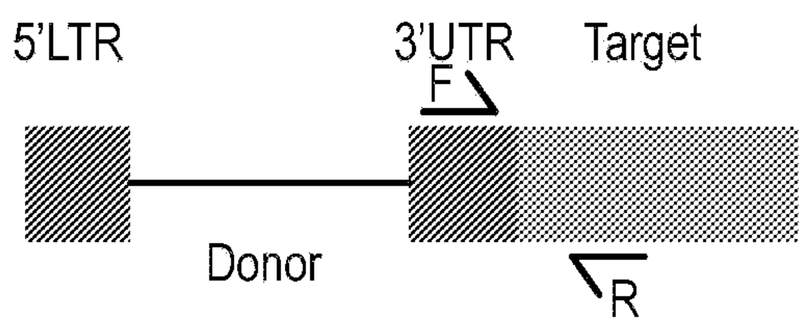


FIG. 33

SUBSTITUTE SHEET (RULE 26)

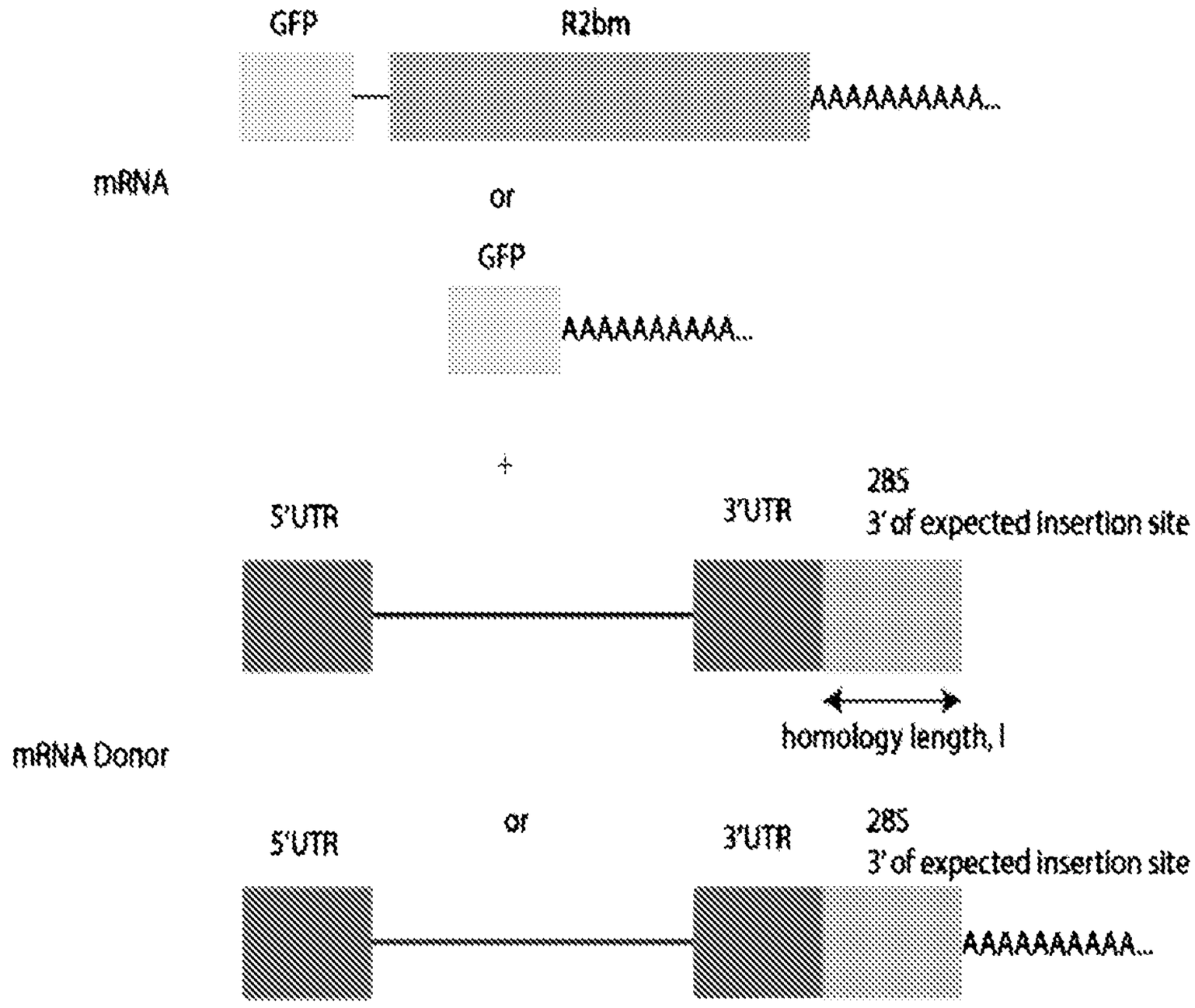

GFP
R2bm
AAAAAAAAAA...
mRNA
or
GFP
AAAAAAAAAA...
+
5'UTR
3'UTR
285
3' of expected insertion site
homology length, l
mRNA Donor
or
5'UTR
3'UTR
285
3' of expected insertion site
AAAAAAAAAA...
homology length, l

FIG. 34

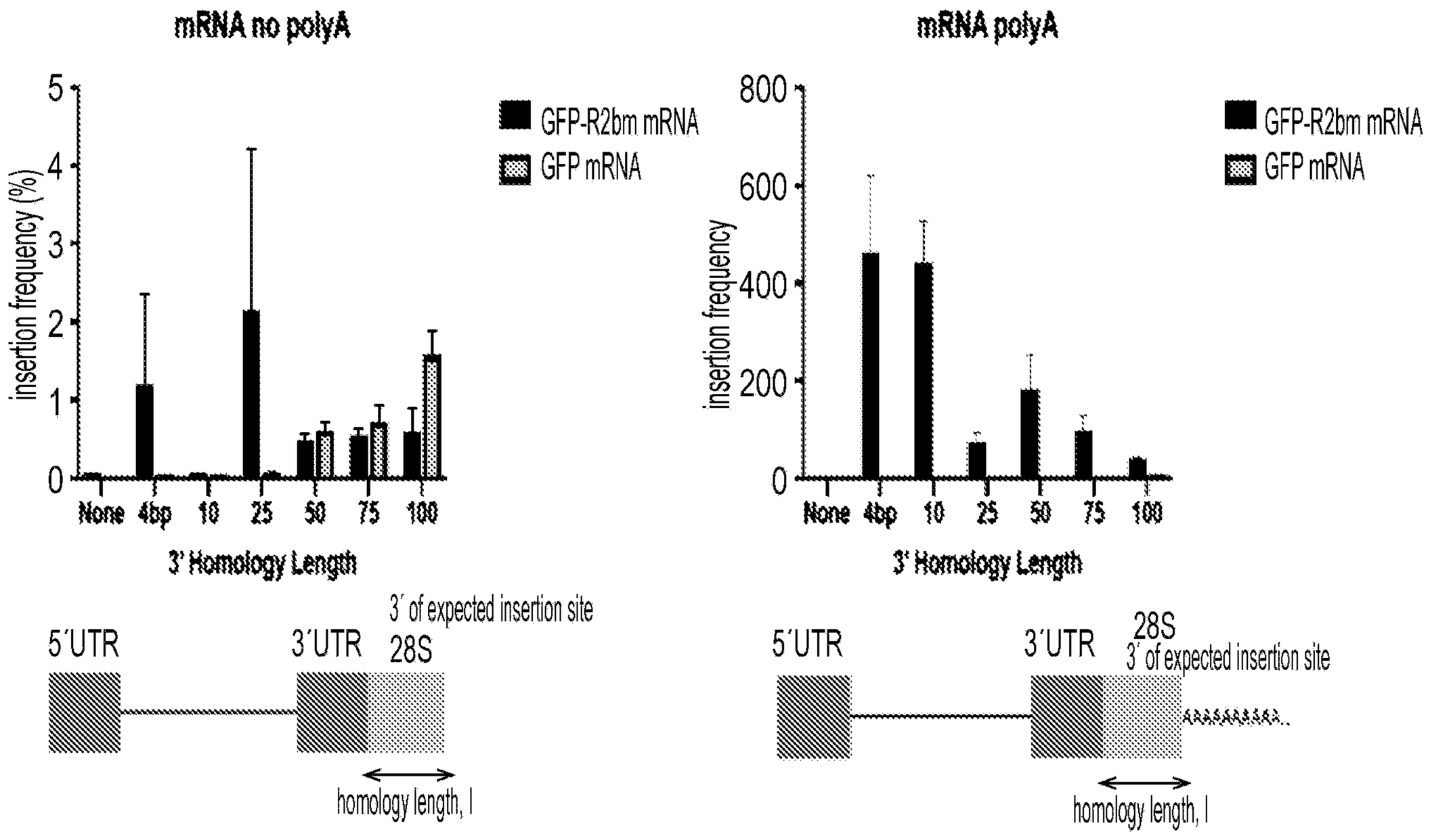

mRNA no polyA
insertion frequency (%)
GFP-R2bm mRNA
GFP mRNA
None 4bp 10 25 50 75 100
3' Homology Length
5'UTR
3'UTR
28S
3' of expected insertion site
homology length, l
mRNA polyA
insertion frequency
GFP-R2bm mRNA
GFP mRNA
None 4bp 10 25 50 75 100
3' Homology Length
5'UTR
3'UTR
28S
3' of expected insertion site
homology length, l

FIG. 35

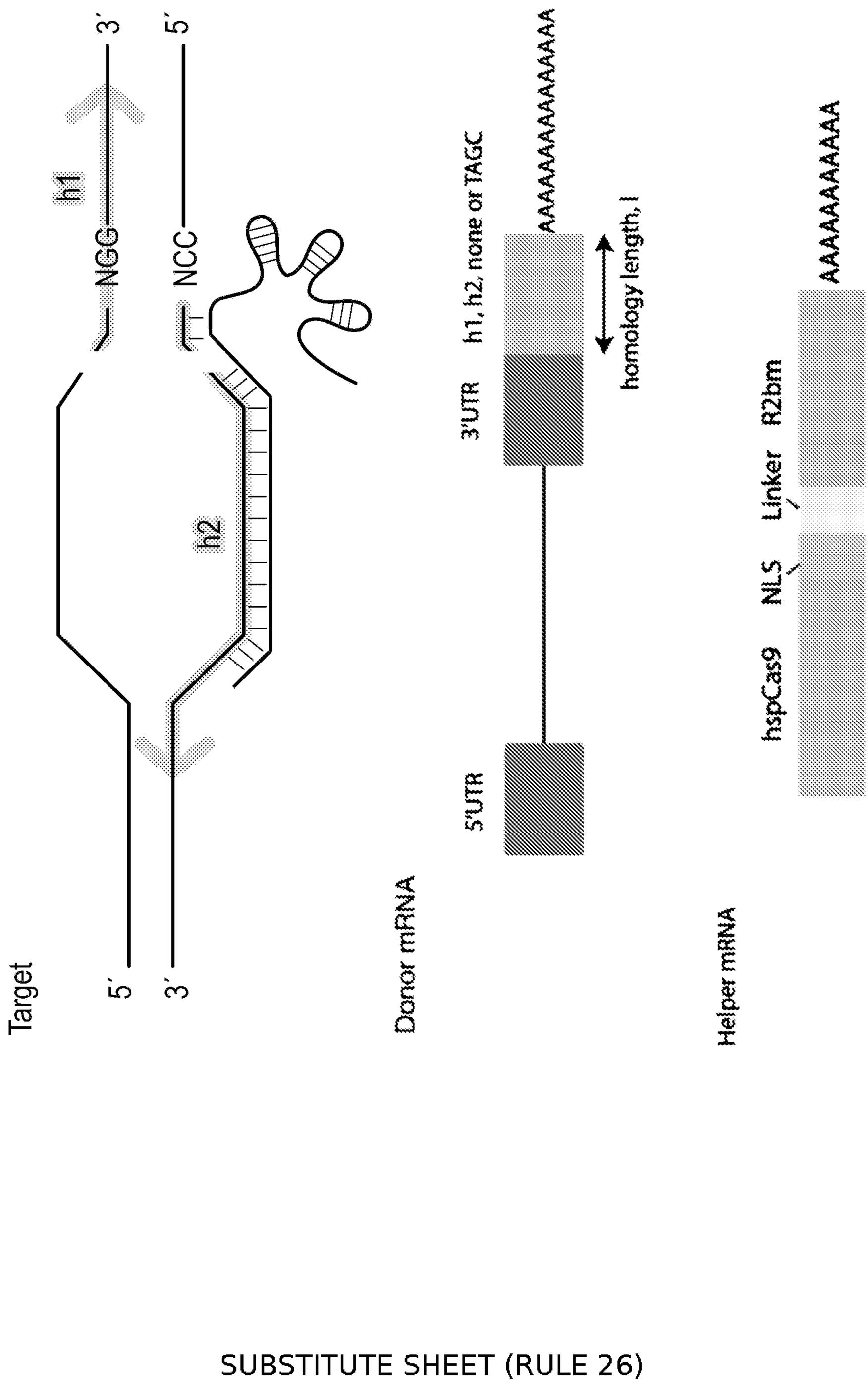

Target
5′
3′
h1
NGG
NCC
h2
Donor mRNA
3'UTR
h1, h2, none or TAGC
AAAAAAAAAAAA
homology length, l
5'UTR
Helper mRNA
Linker
R2bm
hspCas9
NLS
AAAAAAAAAA

FIG. 36

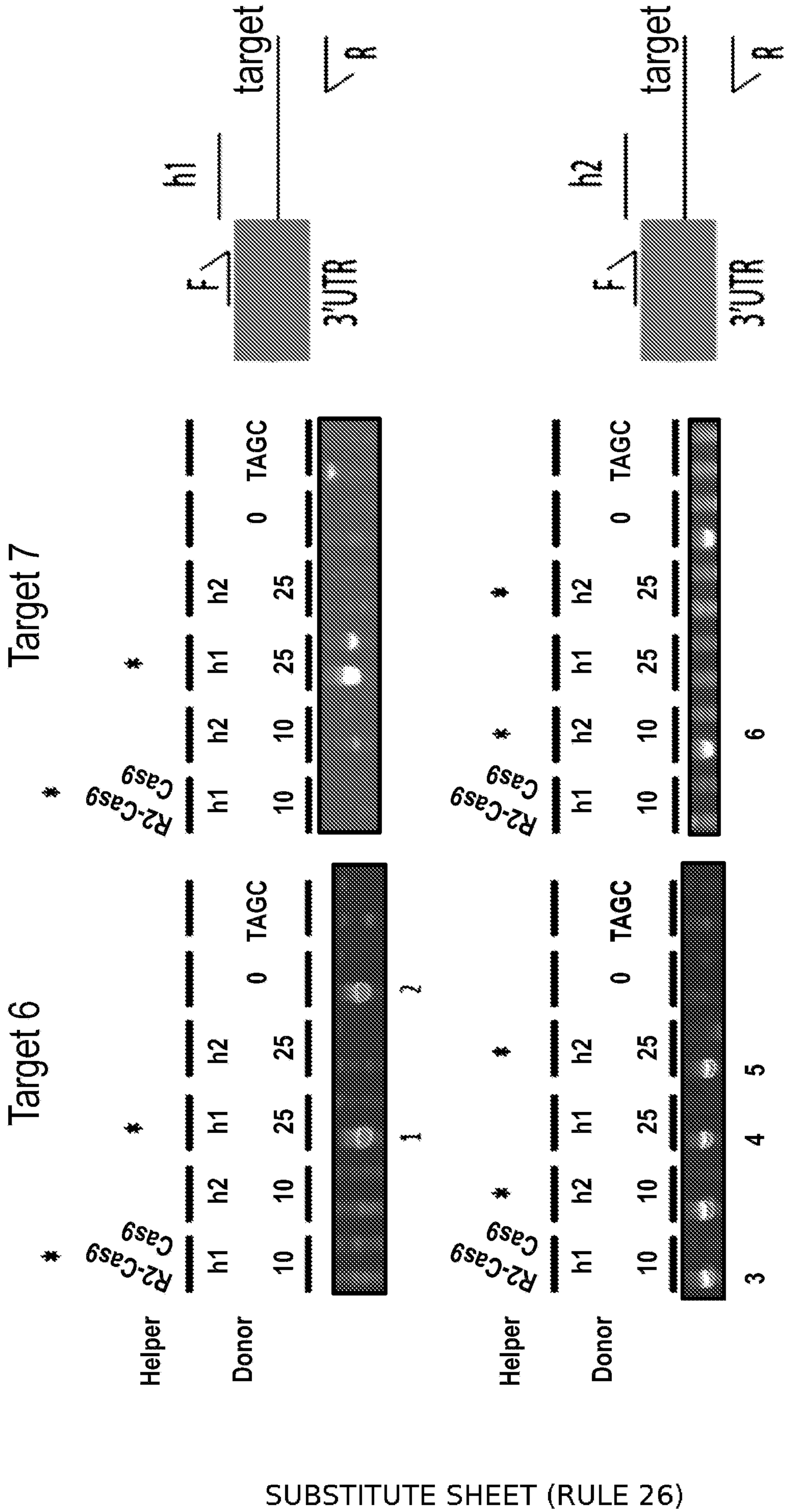
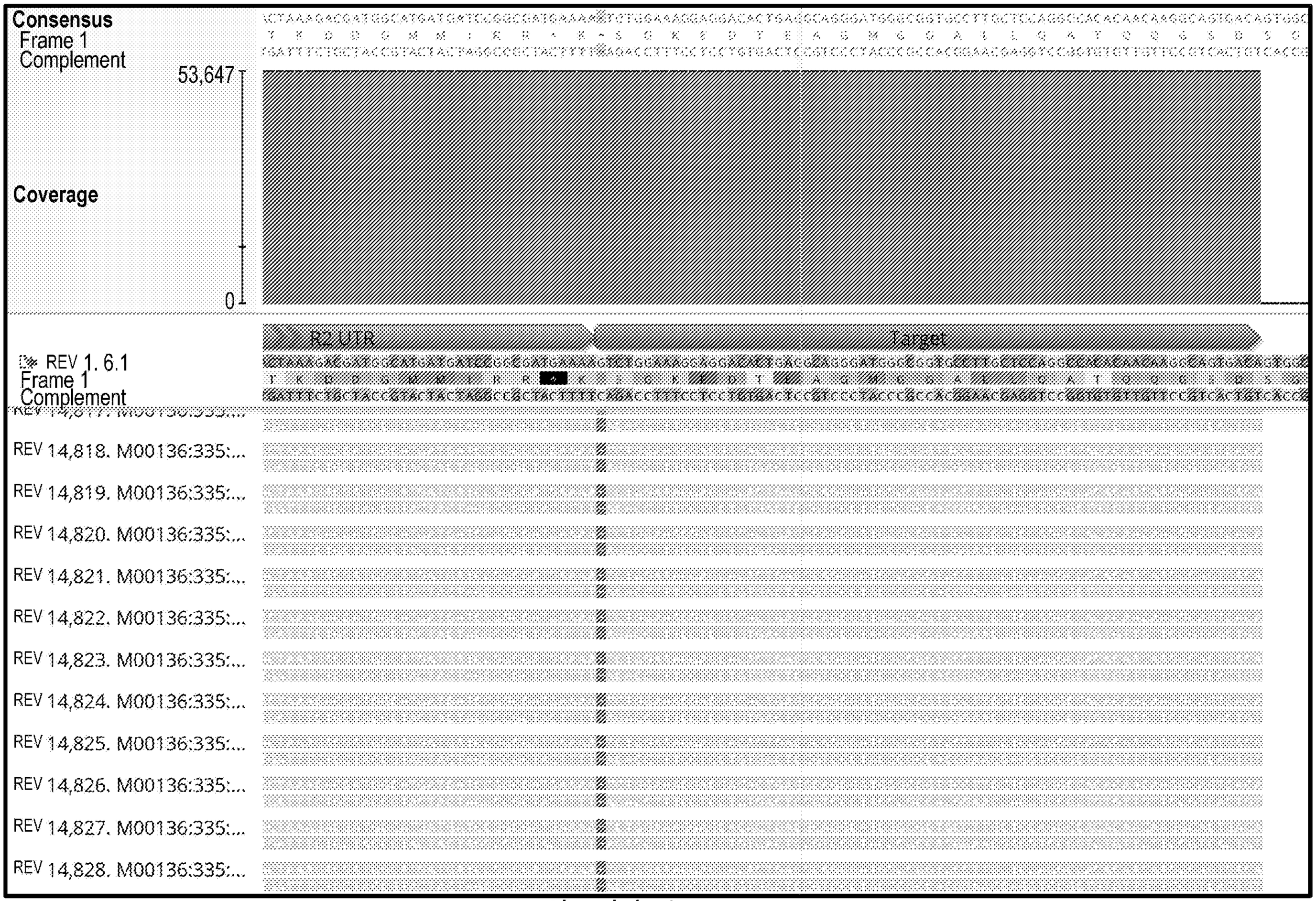

Target 7
Target 6
R2-Cas9
Cas9
Helper
Donor
h1
h2
F
R
3'UTR
TAGC
0
10
25
1
2
3
4
5
6
* Insertion expected

FIG. 37

SUBSTITUTE SHEET (RULE 26)

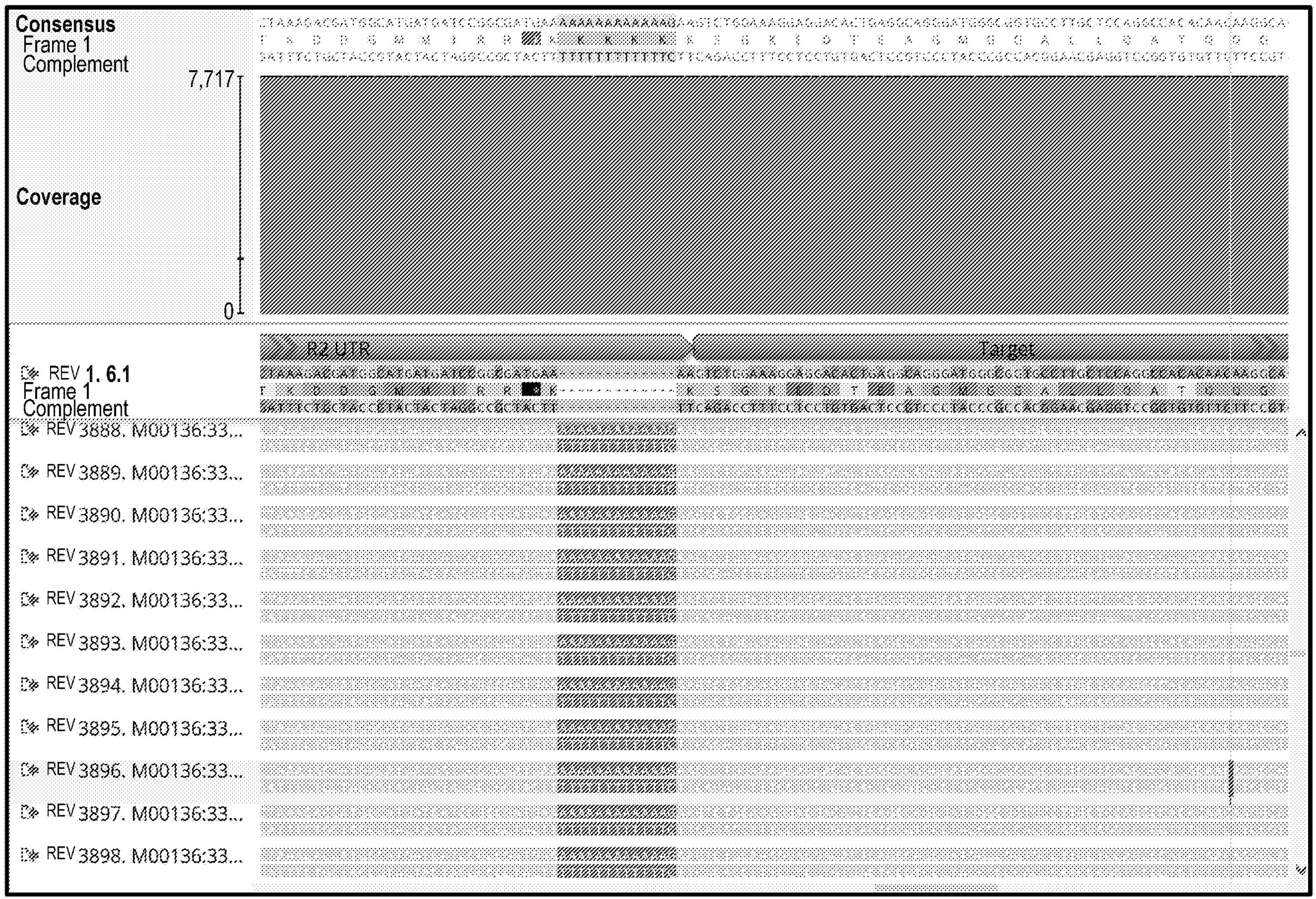
1 – h1 donor, h1 insertion orientation, 25bp homology arm



1bp deletion

FIG. 38A

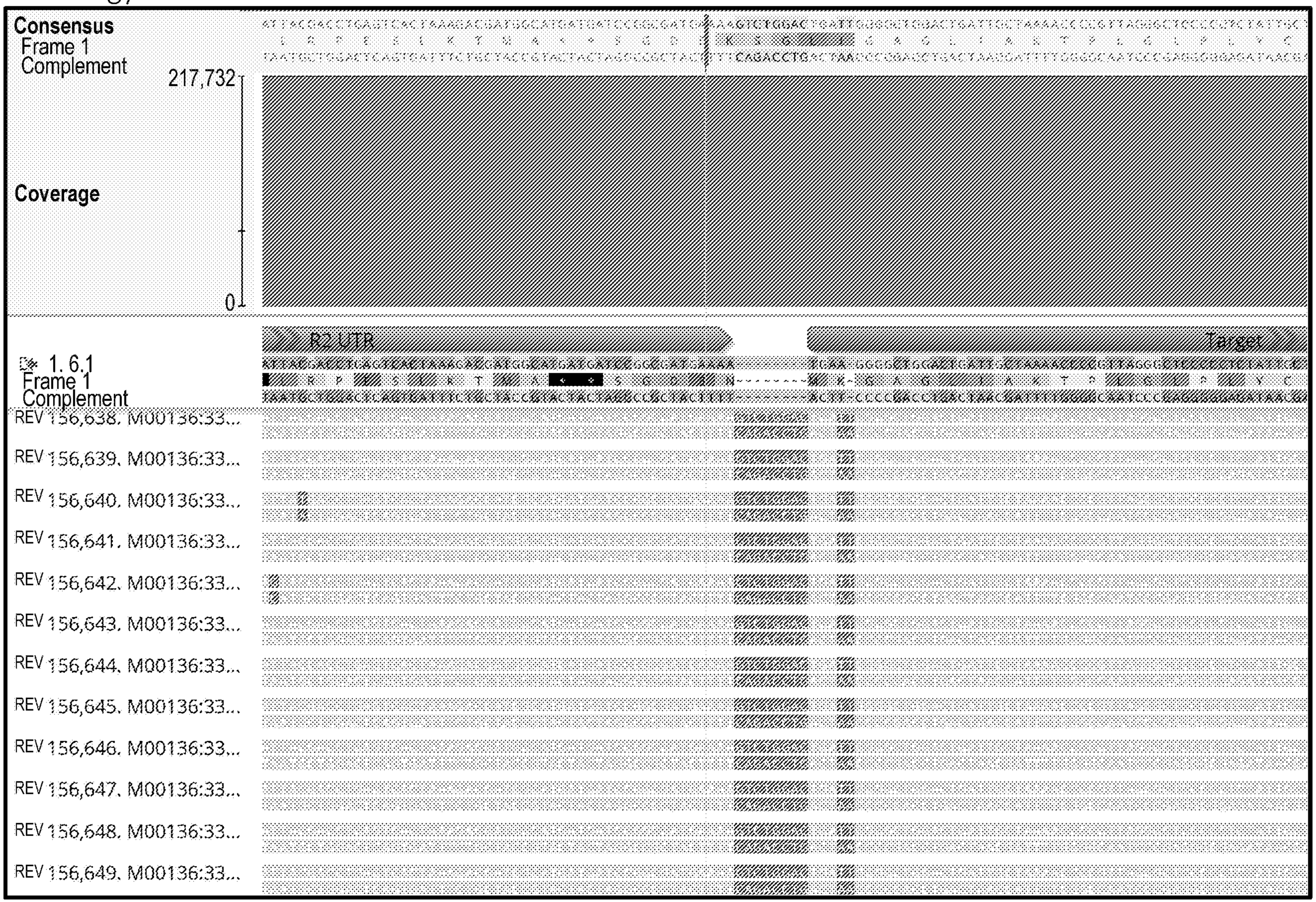
2 – no homology arm, h1 insertion orientation



Insertion of polyA tail

FIG. 38B

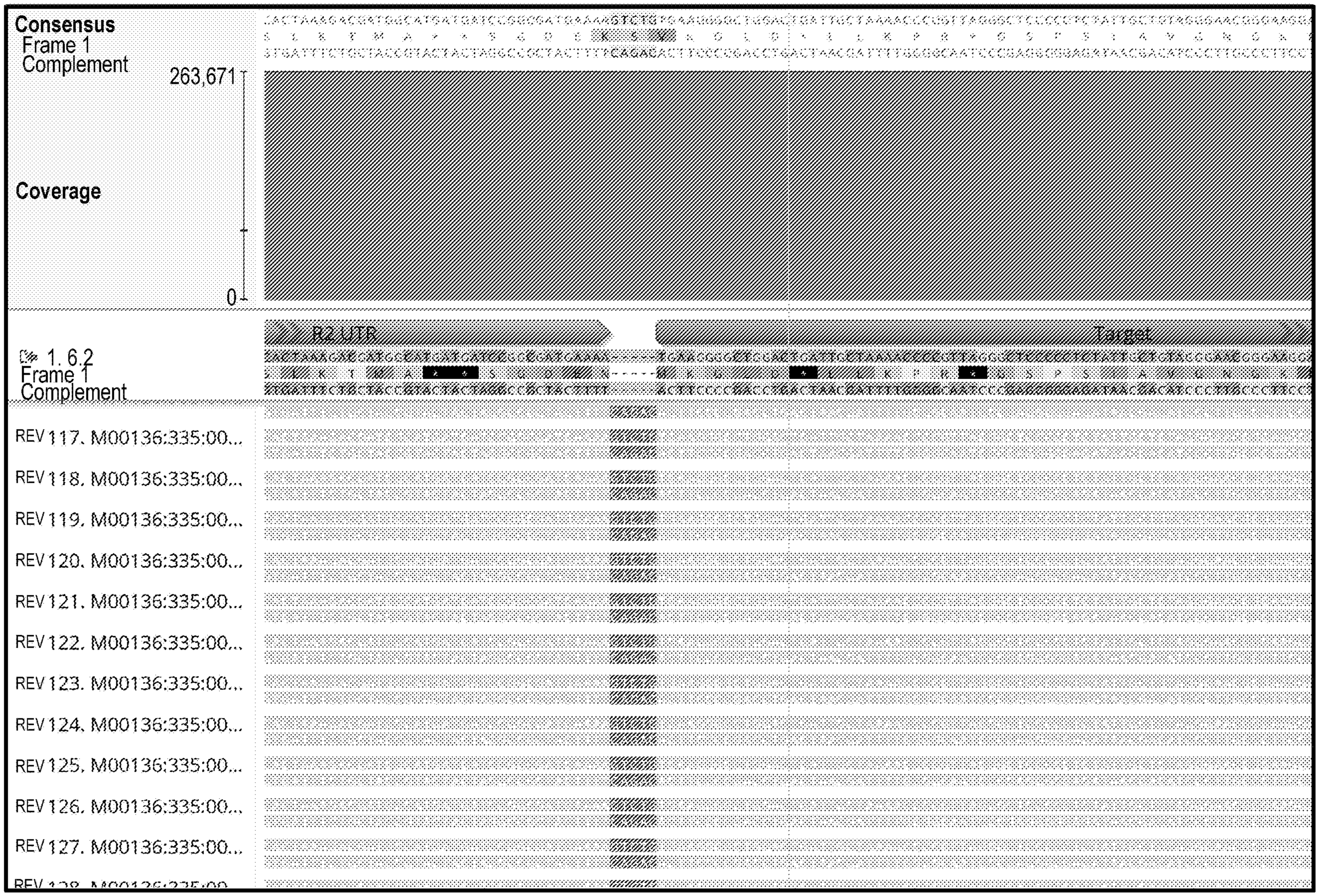
3 – h1 donor, h2 insertion orientation, 10 bp homology



Insertion of 7 bp of h1 donor homology + untemplated nucleotides in h2 orientation

FIG. 38C

4 – h1 donor, h2 insertion orientation, 25bp homology



Insertion of 5 bp of h1 donor in h2 orientation

FIG. 38D

5 – h2 donor, h2 insertion orientation, 25 bp homology arm

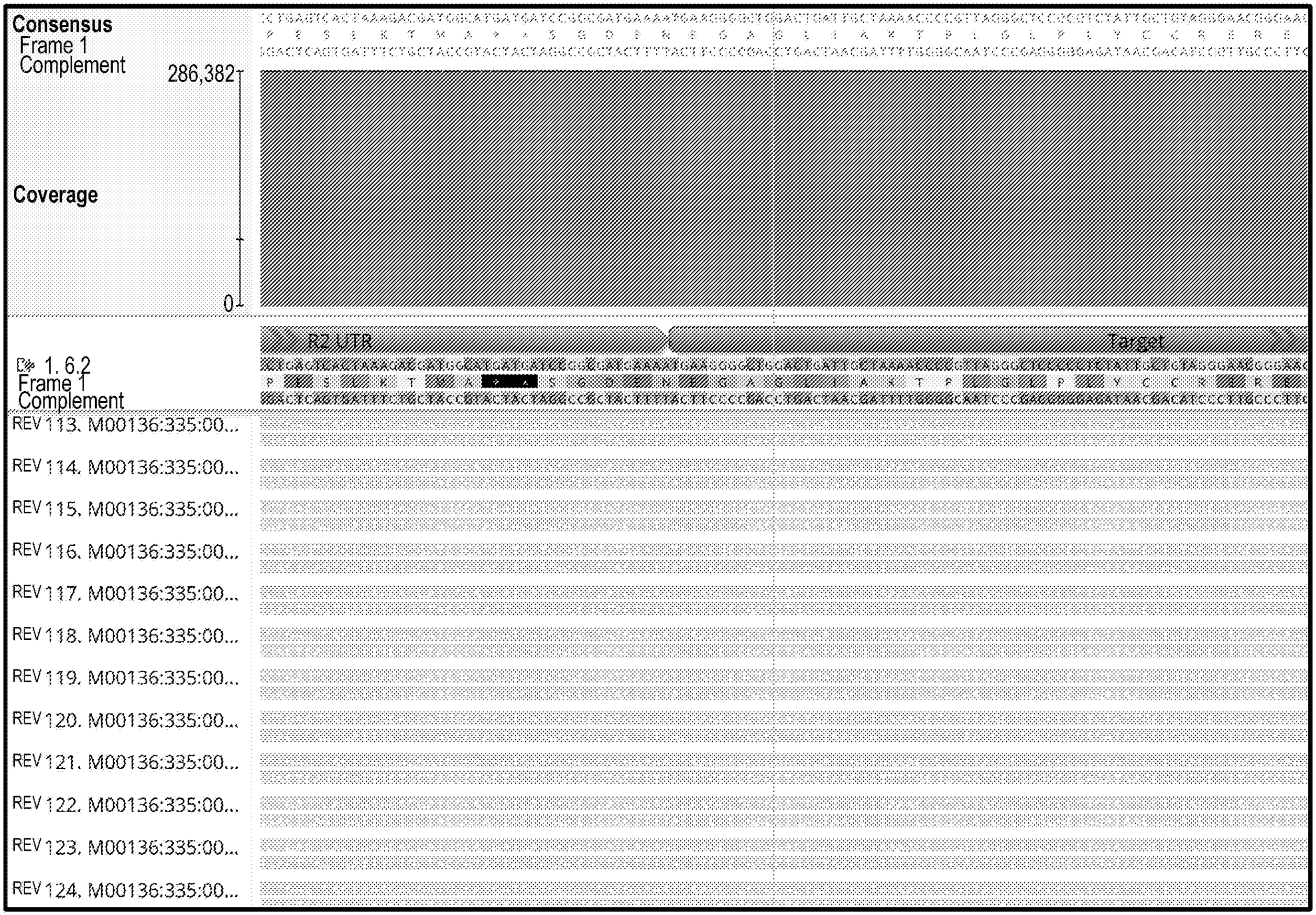


FIG. 38E

SUBSTITUTE SHEET (RULE 26)

6 – h2 donor, h2 insertion orientation, 10 bp homology arm

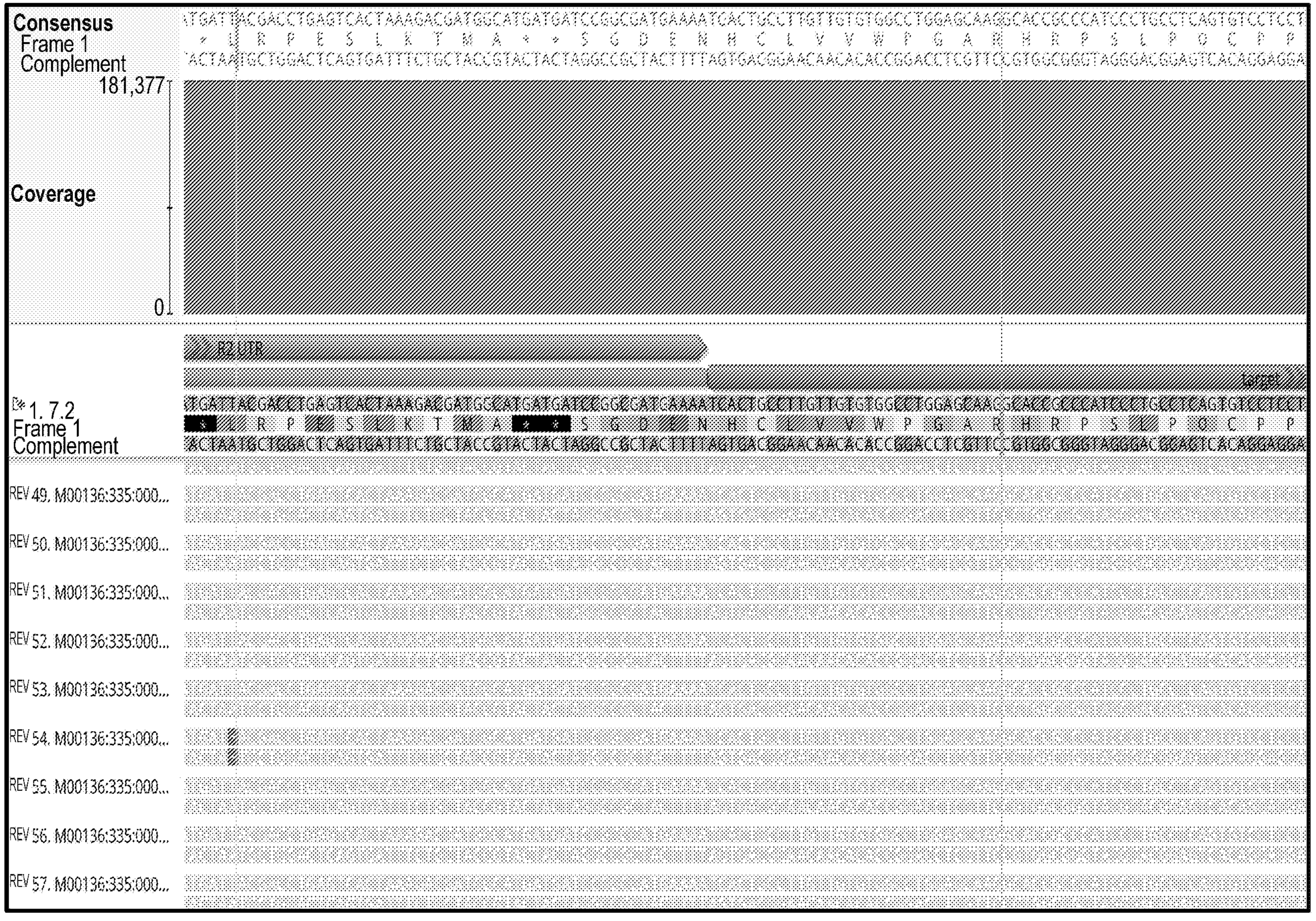


FIG. 38F

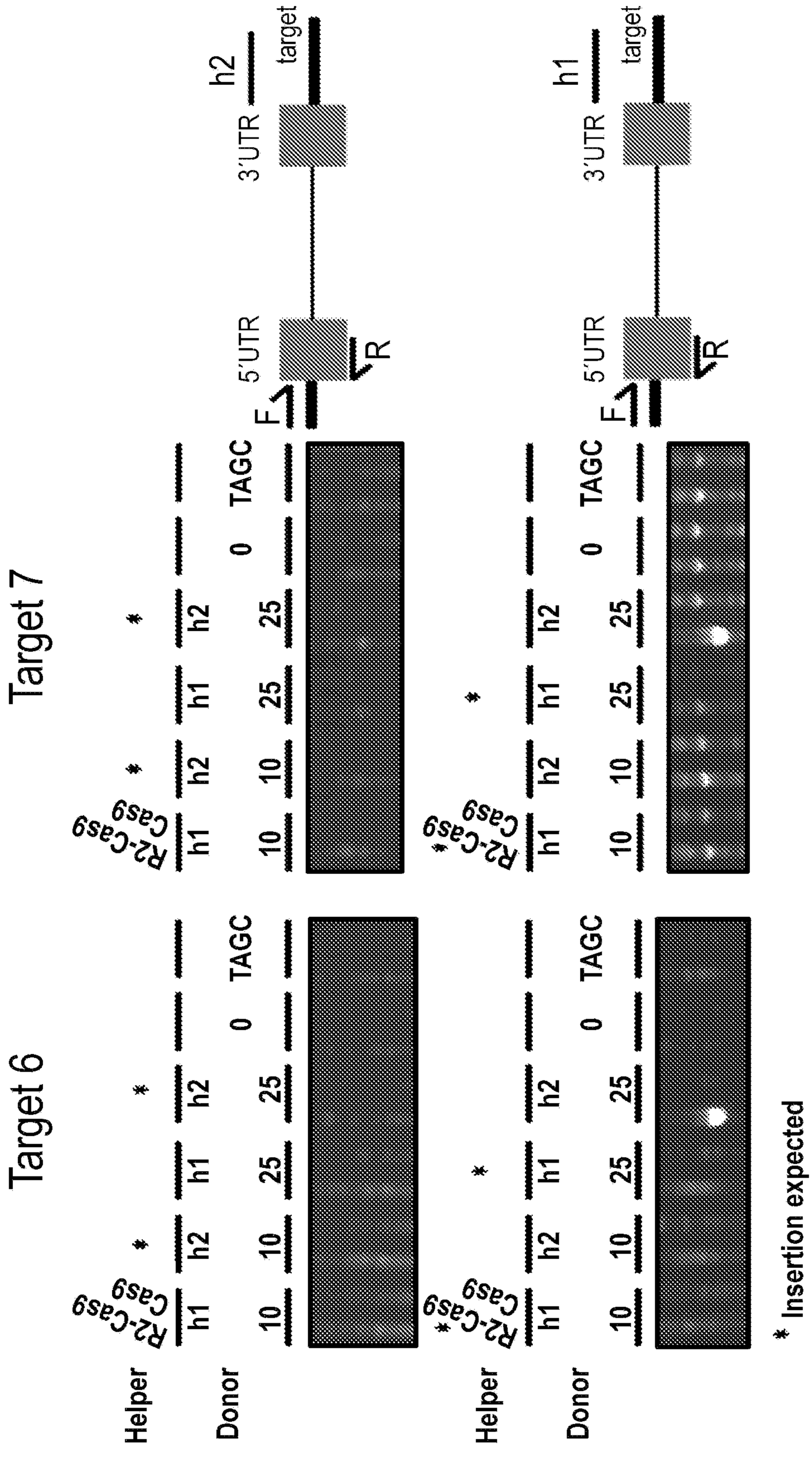

Target 7
h2
target
3'UTR
h1
target
3'UTR
5'UTR
F
R
5'UTR
F
R
0
TAGC
0
TAGC
*
*
*
*
R2-Cas9
Cas9
R2-Cas9
Cas9
h1
h2
h1
h2
h1
h2
h1
h2
10
25
25
10
10
25
25
10
Target 6
Helper
Donor
Helper
Donor
* Insertion expected

FIG. 39

Bright 5’ insertions represent a single truncation product

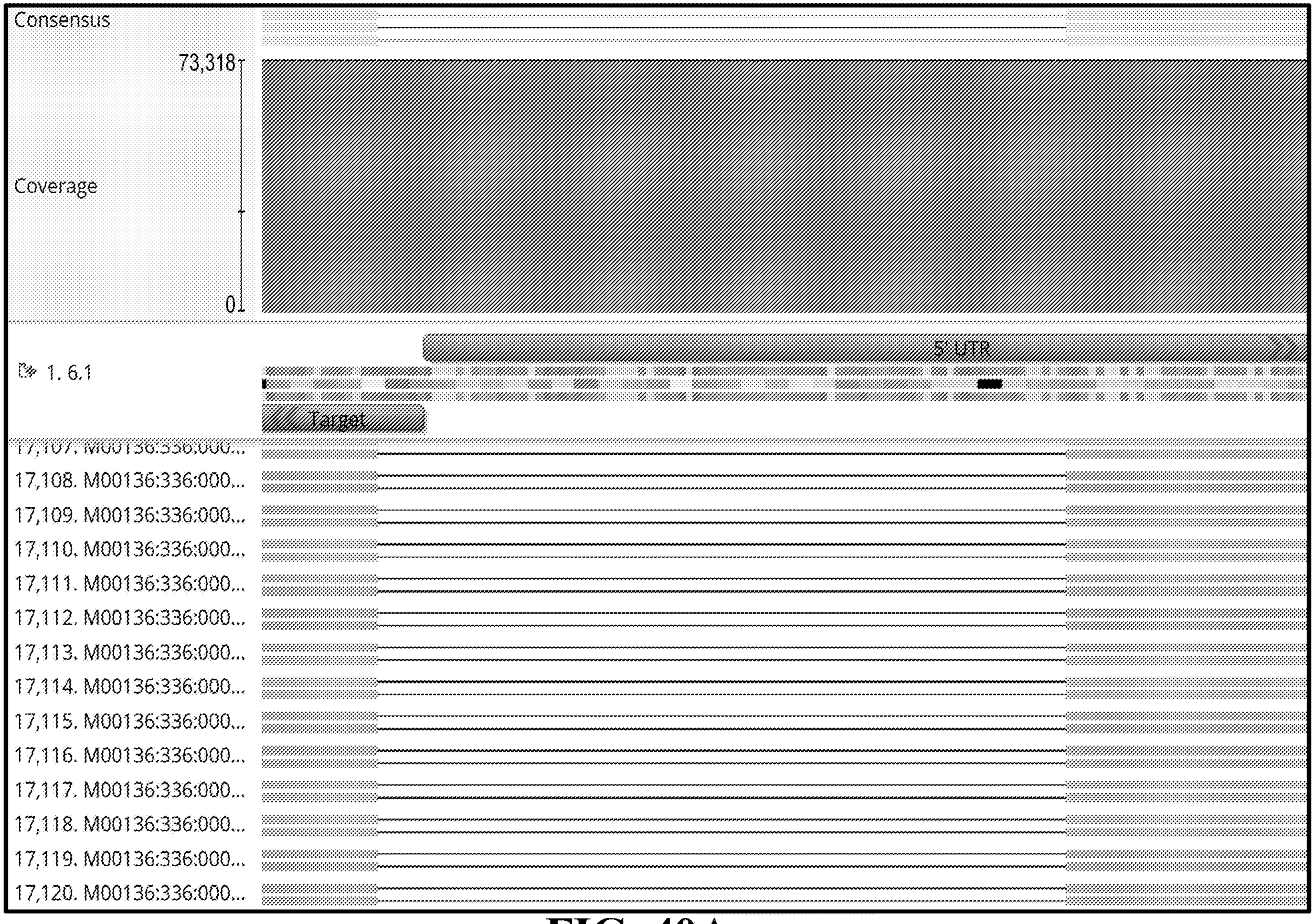


FIG. 40A

SUBSTITUTE SHEET (RULE 26)


Expected 5' insertions are detectable by NGS,
but 5' UTR is often truncated
5' UTR
Target
Consensus
14,705
Coverage
0
1
1, 6, 1
10,369. M00136:336:000...
10,370. M00136:336:000...
10,371. M00136:336:000...
10,372. M00136:336:000...
10,373. M00136:336:000...
10,374. M00136:336:000...
10,375. M00136:336:000...
10,376. M00136:336:000...
10,377. M00136:336:000...
10,378. M00136:336:000...
10,379. M00136:336:000...
10,380. M00136:336:000...
10,381. M00136:336:000...
10,382. M00136:336:000...

FIG. 40B

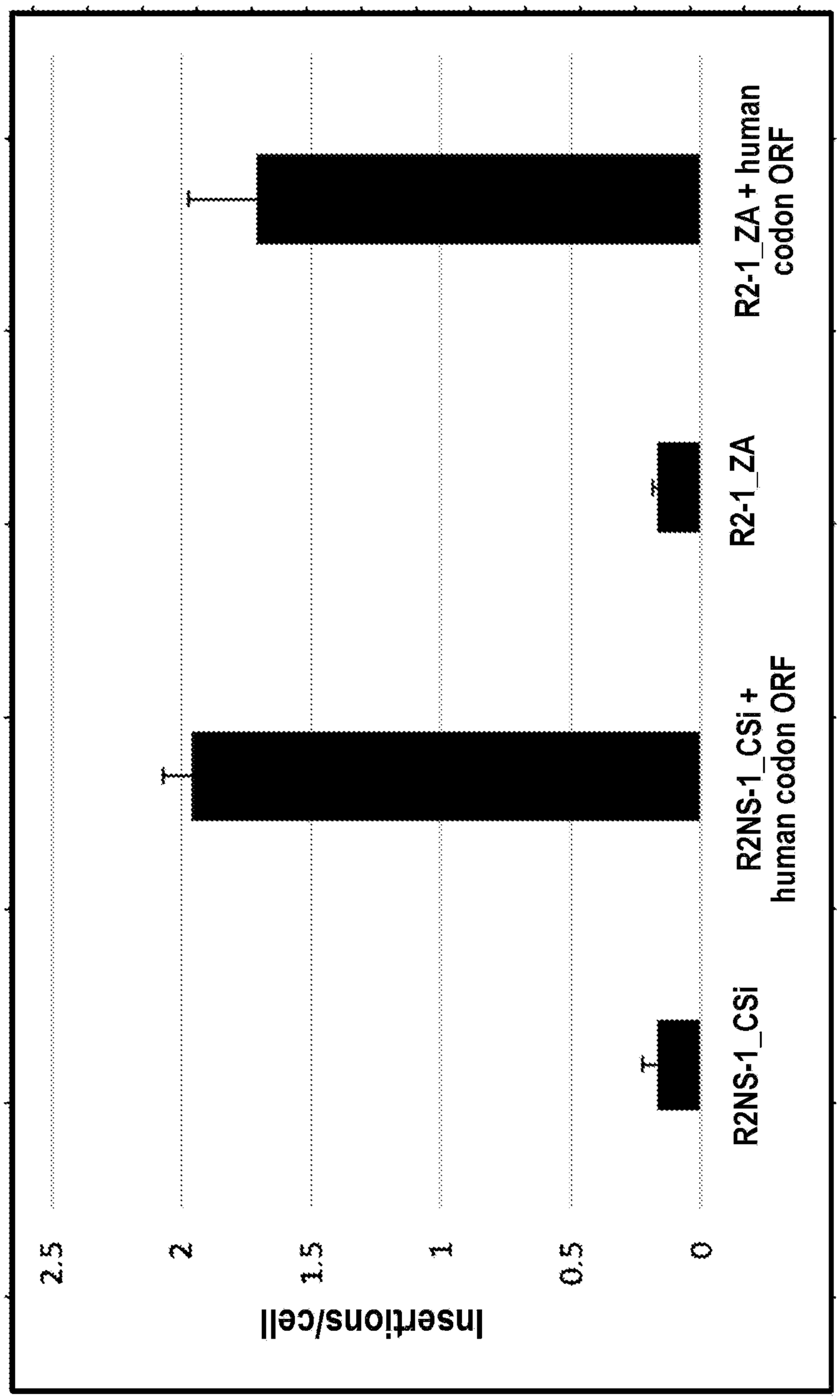

Insertions/cell
2.5
2
1.5
1
0.5
0
R2-1_ZA
R2-1_ZA + human codon ORF
R2NS-1_CSi
R2NS-1_CSi + human codon ORF

FIG. 41

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/61114

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/11, C12N 9/22, C07K 19/00 (2021.01)

CPC - C12N 15/11, C12N 2310/20, C12N 9/22, C12N 15/62, C07K 2319/01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2019/0071711 A1 (2D GENOMICS INC.) 07 March 2019 (07.03.2019) para [0037], [0093], [0145], [0158], [0174], [0178], [0186], [0220], [0214], [0356], [0550]	1, 3-18, 41-62
Y	US 2012/0042411 A1 (MALCUIT et al.) 16 February 2012 (16.02.2012) para [0001], [0008], [0038], [0064], [0068], [0081], [0086], [0140]	1, 3-18, 41-62
Y	US 2006/0281079 A1 (EICKBUSH et al.) 14 December 2006 (14.12.2006) para [0028]. [0129], Abstract	5
Y	US 2018/0334685 A1 (YEO et al.) 22 November 2018 (22.11.2018) para [0013], [0018], [0056], [0087], [0093], Claim 41, Abstract	7-8, 12, 53-55
Y	KOJIMA et al. The Wide Distribution and Change of Target Specificity of R2 Non-LTR Retrotransposons in Animals, PloS One, 23 September 2016, Vol 11, No 9, Article e0163496, Pages 1-16. Especially pg 2 para 2, Abstract	14
Y	US 2017/0321198 A1 (THE BROAD INSTITUTE INC.) 09 November 2017 (09.11.2017) para [0018], [0494], [0953], [0968], Abstract	41

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 February 2021 (04.02.2021)	APR 09 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/61114

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/61114

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

----------Please see continuation in first extra sheet----------

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 3-18, 41-62

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 20/61114

Continuation of Box No. III. Observations where unity of invention is lacking

Group I+, Claims 1-62, directed to an engineered composition for non-native, targeted transposition of donor sequence into targeted nucleic acids. The composition will be searched to the extent that the composition encompasses
a. a single site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;
b. a single non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;
c. a single guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide; (note, these are the first claimed components for the first method embodiment). It is believed that claims 1, 3-18, 41-62 encompass this first named invention, and thus these claims will be searched without fee to the extent that the composition encompasses a. a single site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;
b. a single non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof,
c. a single guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide. Additional composition(s) comprising additional component(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected composition(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a composition comprising
a. a fusion protein comprising a site-specific nuclease fused to the N terminus of a non-LTR retrotransposon polypeptide, or a polynucleotide comprising a coding sequence thereof; and
b. a donor construct comprising a donor polynucleotide sequence located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide (claims 19-40).

Group II, claims 63-67, directed to a vector system.

Group III, claims 68-75, directed to a method of inserting a donor polynucleotide sequence into a target polynucleotide.

The inventions listed as Groups I+, II and III do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special technical features
Group I+ has the special technical feature of an engineered composition for non-native, targeted transposition of donor sequence into targeted nucleic acids, that is not required by Groups II and III.

Group II has the special technical feature of vector system comprising one or more vectors, that is not required by Groups I+ and III.

Group III has the special technical feature of a method of inserting a donor polynucleotide sequence into a target polynucleotide, that is not required by Groups I+ and II.

Common technical features
The inventions of Groups I+, II and III share the common technical feature of:
an engineered or non-naturally occurring composition comprising:
a. a site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof.

Additionally, Group I+ inventions share the technical features of including:
an engineered or non-naturally occurring composition comprising:
a. a site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;
b. a non-LTR retrotransposon polypeptide connected to or otherwise capable of forming a complex with the site-specific nuclease polypeptide, or a polynucleotide comprising a coding sequence thereof;
c. a guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide; and
d. a polynucleotide encoding a retrotransposon RNA, wherein the retrotransposon RNA comprises or encodes a donor polynucleotide.

However, these shared technical features are made obvious by US 2019/0071711 A1 to 2D Genomics Inc. (hereinafter '2DGI') in view of US 2012/0042411 A1 to Malcuit et al. (hereinafter 'Malcuit').

-----please see continuation on next sheet ----------

--------continued from previous sheet ----------

2DGI discloses an engineered composition comprising:
a. a site-specific nuclease polypeptide (para [0145] - "a complex comprising a nuclease and/or a polynucleotide encoding a nuclease"; para [0158]), for targeted transposition of donor sequence into targeted nucleic acids (para [0158]; para [0220] - "the non-LTR retrotransposon or modified non-LTR retrotransposon is fused to a protein, a domain, a fusion partner, a target sequence...or any combination thereof."). 2DGI further discloses c. a guide molecule capable of forming a complex with the site-specific nuclease polypeptide and directing site-specific binding to a target sequence of a target polynucleotide (para [0145] - "a complex comprising a nuclease and/or a polynucleotide encoding a nuclease and at least one specifically designed guide oligonucleotide...at least one guide oligonucleotide comprises at least one sequence complementary to at least one dsDNA."; para [0371] - "Cas9 nuclease is complexed with strategically designed guiding RNA oligonucleotides"), but fails to specifically disclose (d.) a polynucleotide encoding a retrotransposon RNA, wherein the retrotransposon RNA comprises or encodes a donor polynucleotide.
Malcuit discloses a method of targeted gene modification (para [0001] - "comprising genetically altered nuclear DNA, nucleic acid sequences and polypeptides generated as a result of introduction of genetic changes...the introduction of novel nucleic acids into chromosomal sites of interest and the generation of novel domains within genes of interest in a targeted manner"), that comprises retrotransposon RNA and donor polynucleotide (para [0064] - "(B) Retrotransposon RNA-based system of DNA amplification. Donor fragment is fused to a primer binding site (PBS)...The generated RNA is subsequently used for a reverse transcription reaction using retrotransposon specific reverse transcriptase"; para [0068] - "gene targeting using the retrotransposon RNA-based donor amplification system."; para [0081] - "Retrotransposon RNA-based system of DNA amplification is based on the generation of RNA from a transgene cassette containing a donor DNA fragment"). Since 2DGI discloses a method comprising a retrotransposon and donor complex with a retrotransposon (para [0214] - "the modified non-LTR retrotransposon...RNA encoding the modified enzymes"; para [0174] - "the donor complex is a donor R2 complex. In some embodiments, the donor R2 complex comprises an RNA R2 primer."; para [0186] - "the reverse transcriptase is an R2 non-LTR retrotransposon or a modified R2 non-LTR retrotransposon."), it would have been obvious to one of ordinary skill in the art to use the retrotransposon RNA-based system of Malcuit in the methods of 2DGI to improve targeted gene modification of a gene of interest.

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Group I+, II and III inventions lack unity under PCT Rule 13 because they do not share the same or corresponding special technical feature.